8 The history of mental models

P. N. Johnson-Laird

> Deduction is that mode of reasoning which examines the state of things asserted in the premisses, forms a diagram of that state of things, perceives in the parts of the diagram relations not explicitly mentioned in the premisses, satisfies itself by mental experiments upon the diagram that these relations would always subsist, or at least would do so in a certain proportion of cases, and concludes their necessary, or probable, truth.
>
> (C. S. Peirce, 1931–1958, 1.66)

What is the end result of perception? What is the output of linguistic comprehension? How do we anticipate the world, and make sensible decisions about what to do? What underlies thinking and reasoning? One answer to these questions is that we rely on mental models of the world. Perception yields a mental model, linguistic comprehension yields a mental model, and thinking and reasoning are the internal manipulations of mental models. The germ of this answer was first proposed during World War II by the remarkable Scottish psychologist and physiologist, Kenneth Craik. In a short but prescient book, *The nature of explanation*, he sketched such a theory. He wrote:

> If the organism carries a "small-scale model" of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.
>
> (Craik, 1943, Ch. 5, p. 61)

Craik would have developed his sketch into a thoroughgoing theory and tested its empirical consequences. But, on the eve of VE day in 1945, he was cycling in Cambridge when a car door opened in front of him and he was thrown into the path of a lorry. He was 31 years old. It was left to others to follow up his ideas.

Where did the notion of a mental model come from? And how have Craik's successors brought his ideas to fruition? This chapter aims to tell you. There are several historical precursors, although they probably had no direct influence on Craik. The present author confesses that for many years his knowledge of mental models went no further back than 1943. It was a shock to discover that there were important antecedents, particularly certain 19th-century physicists and the great American logician and philosopher, Charles Sanders Peirce. The chapter begins with these precursors, and then describes Craik's hypothesis and some similar ideas about "cognitive maps" proposed by Tolman. It outlines theories of mental representation in the 20th century, which presaged the revival of mental models. It then explains their role in perception, comprehension, and the representation of knowledge. It turns to the mental model theory of deductive reasoning, and describes the application of this theory to reasoning with quantifiers such as "all" and "some" and to reasoning with sentential connectives such as "if" and "or". It outlines the role of models in different strategies for reasoning. It concludes with an assessment of the theory.

THE PRECURSORS

Several 19th-century thinkers anticipated the model theory. The physicist Lord Kelvin stressed the importance to him of the construction of mechanical models of scientific theories. In his 1884 Baltimore lectures, he asserted:

> I never satisfy myself until I can make a mechanical model of a thing. If I can make a mechanical model I can understand it. As long as I cannot make a mechanical model all the way through I cannot understand; and that is why I cannot get the electro-magnetic theory.
>
> (cited by Smith & Wise, 1989, p. 464)

Indeed, he never quite accepted Maxwell's equations for electro-magnetism, because he could not construct a mechanical model of them. Ironically, Maxwell did have a mechanical model in mind in developing his theory (Wise, 1979). This use of models in scientific thinking is characteristic of 19th-century physics. Ludwig Boltzmann (1890) argued that all our ideas and concepts are only internal pictures. And he wrote:

> The task of theory consists in constructing an image of the external world that exists purely internally and must be our guiding star in thought and experiment; that is in completing, as it were, the thinking process and carrying out globally what on a small scale occurs within us whenever we form an idea.
>
> (Boltzmann, 1899)

These notions became obsolescent in the 20th century with the development of quantum theory. As the late Richard Feynman (1985) has remarked, no one can have a model of quantum electrodynamics. The equations make unbelievably accurate predictions, but they defy commonsense interpretation.

A principle of the modern theory of mental models is that a model has the same structure as the situation that it represents. Like an architect's model, or a molecular biologist's model, the parts of the model and their structural relations correspond to those of what it represents. Like these physical models, a mental model is also partial because it represents only certain aspects of the situation. There is accordingly a many-to-one mapping from possibilities in the world to their mental model. Maxwell (1911) in his article on diagrams in the *Encyclopaedia Britannica* stressed the structural aspect of diagrams. But the theory's intellectual grandfather is Charles Sanders Peirce.

Peirce formulated the major system of logic known as the predicate calculus and published its principles in 1883 (3.328; this notation, which is standard, refers to paragraph 328 of Volume 3 of Peirce, 1931–1958). Frege (1879) independently anticipated him. Peirce made many other logical discoveries, and he also devised two diagrammatic systems for logic, which were powerful enough to deal with negation, sentential connectives such as "if", "and", and "or", and quantifiers such as "all", "some", and "none", i.e., with the predicate calculus (see, e.g., Johnson-Laird, 2002). He anticipated semantic networks, which were proposed in the 20th century to represent the meanings of words and sentences (Sowa, 1984), the recent development of discourse representation theory in linguistics (Kamp, 1981), and the theory of mental models.

Peirce distinguished three properties of signs in general, in which he included thoughts (4.447). First, they can be iconic and represent entities in virtue of structural similarity to them. Visual images, for example, are iconic. Second, they can be indexical and represent entities in virtue of a direct physical connexion. The act of pointing to an object, for example, is indexical. Third, they can be symbolic and represent entities in virtue of a conventional rule or habit. A verbal description, for example, is symbolic. The properties can co-occur: a photograph with verbal labels for its parts is iconic, indexical, and symbolic. Diagrams, Peirce believed, ought to be iconic (4.433). He meant that there should be a visible 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). The London tube map is a wonderful iconic representation of the city's subway system. Its designer, Harry Beck, realized that underground travellers need to know only the order of stations on each tube line, and where they can change from one line to another. His map captures these relations in a pleasingly transparent way, but it makes no attempt to capture distances systematically. It is an old joke to suggest to out-of-towners that they go by tube from Bank to Mansion House. The map calls for a change from one line to another, and a journey through several stations. Bank and Mansion House are about a 5-minute walk apart.

In his early work, the philosopher Ludwig Wittgenstein (1922) defended a "picture" theory of meaning, which was inspired by the use of model cars in the reconstruction of an accident. It can be summarized in a handful of propositions from his Tractatus:

2.1 We make to ourselves pictures of facts.
2.12 The picture is a model of reality.
2.13 To the objects [in the world] correspond in the picture the elements of the picture.
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.
2.17 What the picture must have in common with reality in order to be able to represent it after its manner – rightly or falsely – is its form of representation.

The Gestalt notion that vision creates an isomorphism between brain fields and the world (Köhler, 1938) is yet another version of the same idea. Peirce, however, had anticipated Maxwell, Wittgenstein, and Köhler. His concept of an iconic representation contrasts, as he recognized, with the syntactical symbols of a language. The iconic nature of diagrams made possible Peirce's (1.66) thesis in the epigraph to the present chapter. He argued that the inspection of an iconic diagram reveals truths to be discerned over and above those of the propositions that were used in its construction (2.279, 4.530). This property of iconicity is fundamental to the modern theory of mental models (Johnson-Laird, 1983, pp. 125, 136).

THE FIRST MODEL THEORISTS

Craik's (1943) book, *The nature of explanation*, addresses philosophical problems, and argues against both scepticism and an a priori approach to the existence of the external world. The core of the book, however, is its fifth chapter, which is about thought. Craik argues that its fundamental property is its power to predict events. This power depends on three steps:

(1) The translation of an external process into words, numbers, or other symbols, which can function as a model of the world.
(2) A process of reasoning from these symbols leading to others.
(3) The retranslation back from the resulting symbols into external processes, or at least to a recognition that they correspond to external processes.

Stimulation of the sense organs translates into neural patterns, reasoning produces other neural patterns, and they are retranslated into the excitation of the motor organs. The process is akin, Craik writes, to one in which the final result was reached by causing actual physical processes to occur. Instead of building a bridge to see if it works, you envisage how to build it. The brain accordingly imitates or models the physical processes that it is trying to predict. Craik makes the prescient claim that the same process of imitation can be carried out by a mechanical device, such as a calculating machine, an anti-aircraft "predictor", or Kelvin's machine for predicting the tides. The programmable digital computer had yet to be invented, although its precursor at Bletchley Park was in use to crack the German Enigma cipher.

One difference between Craik's views and the modern theory of mental models concerns iconicity. Craik eschews it. He writes (pp. 51–2):

> the model need not resemble the real object pictorially; Kelvin's tide-predictor, which consists of pulleys on levers, does not resemble a tide in appearance, but it works in the same way in certain essential respects – it combines oscillations of various frequencies so as to produce an oscillation which closely resembles in amplitude at each moment the variation in tide level at any place.

So, for Craik, a model parallels or imitates reality, but its structure can differ from the structure of what it represents. In contrast, mental models are now usually considered to mirror the structure of what they represent (for a Craikian view, see Holland, 1998). A model of the world can have a three-dimensional structure for high-level processes such as spatial reasoning. But it does not necessarily call for a three-dimensional layout in the brain (or a computer). Its physical embodiment has merely to support a representation that functions as three dimensional for reasoning. Underlying the high level of representation, there might be – as Craik supposed – something as remote from it as Kelvin's tidal predictor is from the sea.

Craik was among the first to propose a philosophy of mind now known as "functionalism" (Putnam, 1960). This doctrine proposes that what is crucial about the mind is not its dependence on the brain, but its functional organization. Craik wrote (1943, p. 51):

> By a model we thus mean any physical or chemical system which has a similar relation-structure to that of the process it imitates. By "relation-structure" I do not mean some obscure non-physical entity which attends the model, but the fact that it is a physical working model which works in the same way as the process it parallels, in the aspects under consideration at any moment.

He added (p. 57):

> My hypothesis then is that thought models, or parallels, reality – that its essential feature is not "the mind", "the self", "sense-data", nor

> propositions but symbolism, and this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation.

Hence Craik likens the nervous system to a calculating machine capable of modelling external events, and he claims that this process of paralleling is the basic feature of thought and of explanation (pp. 120–121).

Craik has little to say about reasoning, the process that leads from the input to the output symbols. He implies, however, that it is a linguistic process: "[language] must evolve rules of implication governing the use of words, in such a way that a line of thought can run parallel to, and predict, causally determined events in the external world" (p. 81).

An American contemporary of Craik's, Edward C. Tolman, independently developed a similar idea. Tolman was a behaviourist who was influenced by Gestalt theory. His research concerned rats running mazes, and he and his colleagues showed that they tend to learn the spatial location of the reward rather than the sequence of required responses to get there. Tolman's hypotheses were couched, not in mentalistic terms, but in the language of neo-behaviourism, i.e, in terms of variables that intervened between stimuli and responses (Tolman, 1932, 1959). Some maze studies had shown that animals appeared to be able to reason, i.e., they could learn two paths on separate occasions and, if necessary, combine them in order to reach a goal (e.g., Maier, 1929; Tolman & Honzik, 1930). This performance could be explained in behaviourist terms by the mediation of "fractional anticipatory goal responses". But, in an influential paper, Tolman (1948) introduced the concept of a "cognitive map". He suggested that the rat's brain learns something akin to a map of the environment. This map governs the animal's behaviour. For instance, if it learns a complicated dog-legged route to food, then when this route is blocked, it chooses a path directly to a point close to the food box. Students of ethology will know von Frisch's (1966) similar findings about the dance of the honey bees. After they have flown a dog-leg to nectar, their dance signals the direct route.

How much of the environment does a cognitive map cover? Tolman's study showed that rats could acquire not merely a narrow strip, but "a wider comprehensive map to the effect that the food was located in such and such a direction in the room" (1948, p. 204). Tolman argued that what militates against comprehensive maps are inadequate cues, repetitive training, and too great a motivation. He speculated that regression to childhood, fixation, and hostility to an "out-group" are expressions of narrow cognitive maps in human beings. What is missing from his account is any extrapolation from cognitive maps to human spatial representations and navigation. That extrapolation was left to others (e.g., Kitchin, 1994; Thorndyke & Hayes-Roth, 1982). In short, neither of the original model theorists addressed the puzzle of how models underlie reasoning.

MENTAL REPRESENTATIONS

Mentalistic psychologists investigated visual imagery around the end of the 19th century (e.g., Binet, 1894; Perky, 1910). In a study of syllogistic reasoning, Störring (1908) reported that his participants used either visual images or verbal methods to reason. The study of imagery, however, fell into neglect during the era of behaviourism. With the revival of mentalism, cognitive psychologists again distinguished between visual and verbal representations (Bower, 1970; Paivio, 1971). Shepard and his colleagues demonstrated that individuals can transform objects mentally in a variety of ways. In one of their experiments (Shepard & Metzler, 1971), the participants saw two drawings of a "nonsense" figure assembled out of 10 blocks glued together to form a rigid object with right-angled joints. They had to decide whether the pictures depicted one and the same object. Their decision times increased linearly with the angular difference between the orientations of the object in the two pictures. This result held for rotations in the picture plane, but also held for rotations in depth. As Metzler and Shepard (1982, p. 45) wrote: "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." In other words, the participants were rotating mental models of the objects.

Kosslyn and his colleagues asked experimental participants to scan from one landmark to another in their image of a map that they had committed to memory (Kosslyn, Ball, & Reiser, 1978). In another study, participants had to form an image of, say, an elephant and then to imagine walking towards it until the image began to overflow their mind's eye. In this way, Kosslyn (1980) was able to estimate the size of the mental "screen" on which the participants project their images. It is about the same size for an image as for a visual percept. Other investigations of visual imagery – from its mnemonic value (Luria, 1969) to its need for a special short-term memory store (Baddeley & Hitch, 1974) – implied that it is a distinct medium of mental representation.

Sceptics such as Pylyshyn (1973) rejected this view. A distinct medium of representation would be part of the functional architecture of the mind and so its properties could not be affected by an individual's beliefs or attitudes. The case is comparable, Pylyshyn claimed, to the architecture of a computer: the design of its hardware cannot be modified by a program that the computer is running. Mental architecture is thus "cognitively impenetrable", whereas imagery is influenced by an individual's beliefs. And, Pylyshyn argued, the results of the rotation and scanning experiments might merely show that individuals can simulate how long it would take to rotate an actual object, or to scan an actual map. Such simulations therefore reveal nothing about the real nature of mental representations. In Pylyshyn's view, the mind makes no use of images. They occur as subjective experiences, but they play no causal role in mental processes. The mind carries out formal computations on a single medium of representation, so-called "propositional representations", that is, syntactically structured expressions in a mental language. This claim dovetailed with the then orthodox theory of reasoning, which postulated that formal rules of inference akin to those of logic are applied to representations of the logical form of assertions (see below).

There appear to be two ways to resolve the argument between the "imagists" and "propositionalists" (Johnson-Laird, 1983, ch. 7). In one sense of propositional representation, the propositionalists are right. The mind depends on the brain's "machine code", i.e., everything must be reduced to nerve impulses and synaptic events, just as the execution of any computer program reduces to the shifting of bits from one memory register to another. Yet, in another sense of propositional representation, the imagists are right. Images and propositional representations are both high-level representations within the same computational medium, just as arrays and lists are distinct data-structures in a high-level programming language.

Recent theorists have argued against propositional representations (Barsalou, 1999; Markman & Dietrich, 2000). These authors claim that both theory and evidence suggest that cognitive science should eschew such abstract representations in favour of representations rooted in perception. As these investigators show, perceptual representations such as visual or kinaesthetic images are powerful. Yet not everything can be represented iconically. No image can alone capture the content of a negation, such as:

The circle is not to the right of the triangle.

Even if you form an image of a cross superimposed on your image of the un-negated situation, you have to know that the cross denotes negation, and you have to know the meaning of negation, that it reverses the truth value of the corresponding, un-negated assertion. No image can capture this meaning (Wittgenstein, 1953). Defenders of imagery might argue that negation can be represented by a contrast class, e.g., the case of the circle to the left of the triangle. Individuals do indeed envisage contrast classes (Oaksford & Chater, 1994; Schroyens, Schaeken, & d'Ydewalle, 2001). But to capture the full meaning of negation, you need to envisage a disjunction of all the sorts of possibility in the class. An affirmative assertion that is true, such as:

The circle is to the right of the triangle

is compatible with one sort of possibility. Its falsity is compatible with many sorts of possibility. In contrast, a negative assertion that is true, such as:

The circle is not to the right of the triangle is compatible with many sorts of possibility. Its falsity is compatible with only one sort of possibility. (This pattern may explain why people are faster to evaluate true affirmatives than false affirmatives, but faster to evaluate false negatives than true negatives, see Clark & Chase, 1972; Wason, 1959). Negations could be spelt out in the form of disjunctions, but the representation of disjunctions cannot be iconic. You cannot perceive whether two signs denote a conjunction or a disjunction of alternatives (Johnson-Laird, 2001).

The moral is that the use of conventional symbols is necessary to represent negation and disjunction, and that in principle at least three distinct sorts of mental representation could exist:

- Propositional representations, which are strings of syntactically structured symbols in a mental language.
- Images, which are two-dimensional visualizable icons, typically of an object or scene from a particular point of view.
- Mental models, which are also iconic as far as possible, but which can contain elements, such as negation, that are not visualizable (Johnson-Laird & Byrne, 1991; Newell, 1990). They can also represent three-dimensional objects and scenes (as in Shepard's studies of mental rotation described earlier).

THE INHERITORS: THE REVIVAL OF MENTAL MODELS

The original model theorists' most immediate influence was on cybernetics (e.g., McCulloch, 1965). But in the 1970s a revival of mental models in psychology occurred in three research areas, which this section explores: vision, knowledge representation, and discourse. The late David Marr (1982) argued that vision depends on an unconscious inference from the structure of an image to a mental model that makes explicit the three-dimensional structure of the scene. The inference makes use of a series of mental representations. Pure vision begins with the physical interaction between light focused on the retinae and the visual pigment in retinal cells. It ends with the "two-and-a-half dimensional" sketch, which makes explicit the relative distance and orientation (with respect to the observer) of each visible surface in the scene. In order to move about safely, however, you need to know what things are where in the world. You need a representation of the world that is independent of your point of view. When you walk into a wood and recognize that it contains trees, shrubs, and plants, you can readily navigate your way through it to a particular goal – say, to a distant landmark – even if you have never been in the wood before. You can do so because vision solves three problems: it constructs a mental model that makes explicit the three-dimensional shapes of everything in the scene, it uses these shapes to identify the objects, and it makes explicit their locations in relation to one another.

Marr and his colleagues postulated that the recognition of objects from their shape depends on two steps. First, the visual system represents the shape of an object in terms of its own canonical axes, e.g., a pencil is a long thin cylinder. Second, the system compares this shape with a mental catalogue of the shapes of all known objects. Each entry in the catalogue is itself a model, which decomposes the object into the shapes of its component parts and their interrelations. At the highest level, the gross shape of the object is made explicit, but at lower levels the detailed shapes of its parts are fleshed out. The matching of a percept to a catalogued model is complicated and not well understood. One possibility is that a cue about the shape of an object may trigger access to a model in the catalogue, which is then used to try to match the rest of the percept (cf. Biederman, 1987).

In the late 1970s, cognitive scientists began to talk of general knowledge as represented in mental models, but without any commitment to a particular sort of structure. For example, Hayes (1979) used assertions in the predicate calculus to describe the naïve physics of liquids. His aim was to capture the content of everyday knowledge, and he was not concerned with how inferential processes use this knowledge. Other researchers in artificial intelligence tried to model everyday qualitative reasoning, and de Kleer (1977) distinguished between envisioning a model and running it to simulate behaviour. To envision a model of a device, you have to consider how each component works in isolation, and to combine this knowledge with the structure of the device to infer how it works. Forbus (1985) implemented a program that constructs two-dimensional spatial models in order to draw inferences about the behaviour of bouncing balls. Such models are simpler than the theory of mechanics, and they have an iconic structure that reflects our qualitative experience of the world, although they also contain conventional symbols. They allow the program to determine the relations among objects just as humans can from a diagram (cf. Glasgow, 1993; Kuipers, 1994; Larkin & Simon, 1987).

Psychologists similarly began to study naïve and expert models of various domains, such as mechanics (McCloskey, Caramazza, & Green, 1980), hand-held calculators (Young, 1981), and electrical circuits (Gentner & Gentner, 1983). At the heart of these studies is the idea that people learn how to make mental simulations of phenomena (e.g., Hegarty, 1992; Schwartz & Black, 1996), either in a series of dynamic images in the mind's eye or in more abstract mental models. Researchers studied how children develop mental models (e.g., Halford, 1993; Vosniadou & Brewer, 1992), how models of one domain can serve as analogies for another domain (Holland, Holyoak, Nisbett, & Thagard, 1986), and how to design artifacts and computer systems for which it is easy for users to acquire models (e.g., Ehrlich, 1996; Genter & Stevens, 1983; Moray, 1990, 1999). They studied the role of models in the diagnosis of faults (e.g., Rouse & Hunt, 1984), and algorithms for diagnosis (e.g., Davis & Hamscher, 1988; de Kleer & Williams, 1987). Knowledge indeed appears to be represented in mental models that are as iconic as possible.

Humans construct models of the world, as do other species, but humans also communicate the content of such models. Discourse accordingly enables individuals to experience the world by proxy. The inklings of this idea are in the following passage:

> It is possible that from the meanings of sentences in connected discourse, the listener implicitly sets up a much abbreviated and not especially linguistic model of the narrative . . . Where the model is incomplete, material may even be unwittingly invented to render the memory more meaningful or more plausible – a process that has its parallel in the initial construction of the model. A good writer or raconteur perhaps has the power to initiate a process very similar to the one that occurs when we are actually perceiving (or imagining) events instead of merely reading or hearing about them.
>
> (Johnson-Laird, 1970, p. 269)

Other psychologists had similar intuitions (e.g., Bransford, Barclay, & Franks, 1972), and experiments showed that individuals rapidly forget the surface form of sentences (Sachs, 1967), their underlying syntax (Johnson-Laird & Stevenson, 1970), and even the gist or meaning of individual sentences (Garnham, 1987).

The present author was aware of some of these developments and intrigued by the possibility that reasoning might be based on a representation of the meaning of discourse, and so he tried to integrate comprehension, reasoning, and consciousness in his book on mental models (Johnson-Laird, 1983). He argued that when individuals understand discourse, they can use its meaning to construct a mental model of the situation to which it refers (see Van Dijk & Kintsch, 1983). This representation is remote from the syntactic structure of sentences. It is iconic in the following ways: it contains a token for each referent in the discourse, each token has properties corresponding to the properties of the referent, and the tokens are interrelated according to the relations among the referents. Hence, an indefinite noun phrase that introduces an individual into the discourse, such as: "an ancient monarch", leads to the insertion of a corresponding token into the discourse model, and subsequent references to the same individual, either direct ("the ancient monarch") or indirect ("the old king"), are used to address the same token in order to attach new information to it. Similar ideas were advanced by workers in formal semantics (Kamp, 1981), linguistics (Karttunen, 1976), psycholinguistics (Stenning, 1977), and artificial intelligence (Webber, 1978). That, perhaps, is why the notion of mental models as representations of discourse is uncontroversial. Although many aspects of discourse models remain puzzling, psycholinguists have made progress in discovering how they are constructed as individuals understand discourse (e.g., Garnham, 2001; Garnham & Oakhill, 1996; Glenberg, Meyer, & Lindem, 1987; Stevenson, 1993). The construction of these models depends on the meaning of sentences, on background knowledge, and on knowledge of human communication.

MENTAL MODELS AND REASONING WITH QUANTIFIERS

Piaget and his colleagues were the first modern psychologists to address the question of how people reason. They argued that intellectual development reaches a stage in which, by about the age of 12, children have acquired a set of formal procedures akin to those of a logical calculus (e.g., Inhelder & Piaget, 1958). Subsequent theorists postulated that the mind is equipped with a set of formal rules of inference (Johnson-Laird, 1975; Macnamara, 1986; Osherson, 1974–6), and this view still has its adherents (e.g., Braine & O'Brien, 1998; Rips, 1994). But, if mental models are the end result of vision and the comprehension of discourse, what is more natural than that reasoning should be based on them? This intuition lay behind an attempt to unify discourse and deduction (first mooted in Johnson-Laird, 1975). It adopted the fundamental semantic principle of validity: an inference is valid if its conclusion holds in all the possibilities consistent with the premises. And it aimed to explain reasoning about syllogisms, i.e., those inferences first formulated by Aristotle that are based on two premises that each contain a single quantifier, such as "all", "some", or "none". The following sort of syllogism is child's play:

> Some of the parents are chemists.
> All the chemists are drivers.
> What follows?

Young children can deduce a valid conclusion:

> Some of the parents are drivers.

In contrast, the following sort of syllogism is very difficult:

> None of the readers is a cyclist.
> All the cyclists are women.
> What follows?

Many people draw the invalid conclusion:

> None of the readers is a woman.

Others suppose that there is no valid conclusion. Only the best of adult reasoners draw the valid conclusion:

> Some of the women are not readers.

In the traditional Scholastic account of syllogisms, each premise is in one of four moods:

All X are Y.
Some X are Y.
No X is a Y.
Some X are not Y.

The terms in the premises can have four arrangements (known as "figures"):

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

where B denotes the term common to both premises, e.g., "cyclist" in the example above. Hence, there are 64 possible pairs of premises (4 moods for the first premise, 4 moods for the second premise, and 4 figures). The syllogisms in figure 2 are identical to those in figure 1 apart from the order of the premises, and, allowing for this factor, the premises in figures 3 and 4 each yield only 10 logically distinct syllogisms, i.e., there are 36 logically distinct syllogistic premises. Granted that As, Bs, and Cs exist in the domain of discourse, only 27 out of the 64 forms of premises yield valid conclusions (Johnson-Laird, 1983, pp. 102–103).

Although syllogisms had been investigated for many years, the first study of the inferences that individuals drew from all 64 possible pairs of premises was not carried out until 1971 – a study done in collaboration with Janellen Huttenlocher, but not reported until later (Johnson-Laird & Steedman, 1978). The patterns of performance were robust, and stood in need of explanation. One hypothesis was accordingly that people imagined the possibilities compatible with the premises and drew whatever conclusion, if any, that held in all of them. Johnson-Laird (1975) outlined such an account based on Euler circles. But there was a problem. Euler circles represent each possibility compatible with syllogistic premises. Thus, a premise of the form *All the A are B* has two Euler diagrams: in one the circle representing A is included within the circle representing B corresponding to the proper inclusion of A within B, and in the other the two circles lie one on top of the other to represent that the two sets are co-extensive. There are 16 possible Euler diagrams for the easy inference above but only 6 for the difficult inference. Granted that the number of possibilities ought to correlate with difficulty, either reasoners are not considering all the possibilities compatible with the premises, or their mental models somehow coalesce different possibilities into a single representation. Erickson (1974) accepted the first alternative, suggesting that reasoners never represent more than four Euler diagrams; others chose the second alternative arguing that a single mental model could represent more than one sort of possibility (Johnson-Laird & Bara, 1984; Johnson-Laird & Steedman, 1978). For example, for easy premises of the form:

Some of the A are B.
All the B are C.

reasoners construct a single model in which only one sort of individual must exist:

A [B] C

where the square brackets indicate that the set of Bs has been exhaustively represented in relation to Cs. This model is consistent with the 16 alternative possibilities, i.e., allowing that there may, or may not, be As that are not Bs, for example, and that the only constraint is that Bs must be Cs. The model yields the conclusion:

Some of the A are C.

or its converse, although the figure biases reasoners to the conclusion shown. A difficult syllogism of the form:

None of the A is a B.
All the B are C.

yields an initial model:

[A] ¬ B
[A] ¬ B
[B] C
[B] C
. . .

where "¬" denotes a symbol for negation, and each line denotes a different individual. This model yields the conclusion: *None of A is a C*, or its converse. But these conclusions are refuted by an alternative model created by adding additional tokens of the set that is not exhaustively represented:

[A] ¬ B C
[A] ¬ B
[B] C
[B] C
. . .

The two models together support the conclusion: *Some of the A are not C*, or its converse. The first of these conclusions is refuted by a third model:

[A] ¬ B C
[A] ¬ B C
[B] C
[B] C
. . .

Only the converse conclusion survives unscathed: *Some of the C are not A*.

There are, of course, other possible procedures for reasoning with models. For example, Johnson-Laird and Bara (1984) described a procedure in which one predicate is substituted for another under the control of models. Thus, given a model of the premise, *Some of the A are B*, such as:

A B
A

The model of second premise, *All the B are C*, is used to substitute Cs for Bs in the model:

A C
A

and hence to draw the conclusion: *Some of the A are C*.

Theorists have argued that some individuals use images, whereas other individuals use verbal methods – a view that goes back to Störring (1908). Thus, Ford (1995) vigorously defends the hypothesis that reasoners use either Euler circles or verbal substitutions based on rules of inference. Reasoners could represent syllogisms as Euler circles, especially when they are used with procedures that prevent an explosion in the number of possible diagrams. They then become difficult to distinguish from the mental models above (Stenning, 2002; Stenning & Oberlander, 1995). A study of the external models that reasoners construct yields a more radical possibility: individuals use a variety of strategies (Bucciarelli & Johnson-Laird, 1999). They systematically overlook possible models of individual premises, e.g., they often treat *All X are Y* as referring to two co-extensive sets, which yields an invalid conclusion for the difficult syllogism above. Granted the variety of strategies, which differ from individual to individual and even within individuals, there is a robust generalization: those syllogisms that call for only a single mental model are reliably easier than those that call for more than one model (see also Espino, Santamaria, Meseguer, & Carreiras, 2000). The generalization suggests that the strategies themselves rely on mental models.

One reason for wondering whether naïve individuals spontaneously use Euler circles is that they are likely to have learned about them in school.

Another reason is that the standard system of circles cannot cope with simple inferences based on multiple quantifiers, for example:

Someone has read all these books.
Therefore, all these books have been read by someone.

The converse inference is invalid on the normal interpretation of these assertions (see Wason & Johnson-Laird, 1972). In contrast, mental models can represent these assertions. The first step is to show how they cope with relations. Consider, for example, the following problem (from Byrne & Johnson-Laird, 1989):

The cup is on the right of the spoon.
The plate is on the left of the spoon.
The knife is in front of the cup.
The fork is in front of the plate.
What's the relation between the fork and the knife?

The premises call for the model:

plate	spoon	cup
fork		knife

which represents the entities as though they were arranged symmetrically on top of a table. The model yields the answer to the question:

The fork is on the left of the knife.

If one word in the second premise is changed:

The plate is on the left of the cup.

the premises are consistent with at least two possible layouts:

plate	spoon	cup
fork		knife

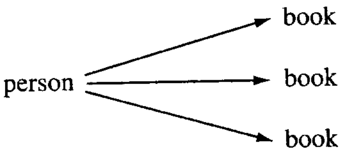
or:

spoon	plate	cup
	fork	knife

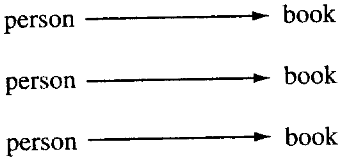
In either case, however, the same conclusion follows as before:

The fork is on the left of the knife.

As the theory predicts, the first problem, which calls for one model, is easier than the second problem, which calls for at least two models. The theory was subsequently extended to temporal and other relations (Carreiras & Santamaria, 1997; Schaeken, Johnson-Laird, & d'Ydewalle, 1996; Vandierendonck & De Vooght, 1996, 1997), and to reasoning based on multiply-quantified relations (Johnson-Laird, Byrne, & Tabossi, 1989). An assertion, such as: *Someone has read all these books*, has the following sort of model:



where the arrows denote the relation of "reading". The assertion: *All these books have been read by someone*, is true in this model, but it also has the following model:



The salient interpretation of the first assertion is false in this model, and so the converse inference is valid.

Reasoners might rely on visual images rather than more abstract models to make relational and quantified inferences. One datum, however, suggests that they use models. Relations that are easy to visualize, but hard to envisage spatially, such as "cleaner than" and "dirtier than", impede reasoning. These relations probably elicit images with vivid details that are irrelevant to reasoning. Hence, they slow the process down in comparison with other sorts of relation, including those that invoke spatial or abstract matters (Knauff & Johnson-Laird, 2002). They are also the only such relations to elicit activity in the areas of the brain that mediate visual associations (Knauff, Fangmeir, Ruff, & Johnson-Laird, 2003).

MENTAL MODELS AND SENTENTIAL REASONING

When the model theory was first formulated, it accounted for relational and quantified reasoning, but not for sentential reasoning, i.e., reasoning that hinges on negation and sentential connectives. A collaboration with Ruth Byrne filled in the lacuna. A disjunctive assertion, such as:

There is a circle or a triangle, or both

calls for models of three possibilities (shown here on separate lines):

o
 Δ
o Δ

whereas the conjunction:

There is a circle and a triangle

calls only for one model (the third of the preceding ones). What was problematic was the representation of conditional assertions, such as:

If there is a circle then there is a triangle.

It was clear that individuals focus on the possibility in which the antecedent is true, and do not think much about other possibilities. We therefore hypothesized that they normally construct two models:

o Δ
. . .

where the second model is a place-holder standing in for the possibilities in which the antecedent of the conditional (there is a circle) is false.

The theory explained the main phenomenon concerning two standard forms of conditional inference. One form is known as *modus ponens*:

If there is a circle then there is a triangle.
There is a circle.
Therefore, there is a triangle.

and the other form is known as *modus tollens*:

If there is a circle then there is a triangle.
There is not a triangle.
Therefore, there is not a circle.

Modus ponens is reliably easier than modus tollens (for a review, see Evans, Newstead, & Byrne, 1993). Yet individuals *can* make a modus tollens inference. So, how is that possible? Plainly, they must be able on occasion to represent explicitly the possibilities in which the antecedent of the conditional is false (see Barrouillet, Grosset, & Lecas, 2000; Barrouillet & Lecas, 1999; Girotto, Mazzocco, & Tasso, 1997; Markovits, 2000). They flesh out their representation into fully explicit models, corresponding either to those of a biconditional (if there is a circle then there is a triangle, and if there isn't a circle then there isn't a triangle):

o Δ
¬o ¬Δ

or to those of a regular conditional (if there is a circle then there is a triangle and if there isn't a circle then there may, or may not, be a triangle):

o Δ
o Δ
¬o ¬Δ

where "¬" denotes a symbol for negation. We therefore distinguished between mental models and fully explicit models. But how do reasoners get from one to the other?

We argued that individuals must make mental footnotes on their models of the conditional (Johnson-Laird & Byrne, 1991), and we adopted a rather cumbersome notation to represent them, akin to the notation for showing that a set has been exhaustively represented (see above). Later, we introduced a more efficient representation in a computer implementation of the theory, which makes footnotes on mental models to indicate what is false in them. For example, the mental models for a conditional of the form:

If A and B then C

are as follows:

A B C
. . .

and the footnote on the implicit model is that it represents the possibilities in which the conjunction, A and B, is false. The models can therefore be fleshed out to represent the seven fully explicit possibilities compatible with this conditional. Naïve reasoners are unable to enumerate these possibilities correctly (Barres & Johnson-Laird, 2002). In general, more models mean more work, and less chance of a correct conclusion (Johnson-Laird & Byrne, 1991; Klauer & Oberauer, 1995). Reasoners tend to focus on as few mental models as possible, and often just on a single model (Richardson & Ormerod, 1997; Sloutsky & Goldvarg, 1999; Sloutsky & Johnson-Laird, 1999). Hence, another possibility is that they use the meaning of the conditional to construct this or that model depending on the circumstances (Johnson-Laird & Byrne, 2002).

The theory of sentential reasoning formulated in Johnson-Laird and Byrne (1991) abided by a principle, whose importance was not realized at first: mental models represent only what is true, not what is false. This principle of truth applies both to premises as a whole and to clauses within them. For premises as a whole, models represent only the possibilities that are true. For each clause in the premises, mental models represent the clause only when it is true in a possibility. For example the exclusive disjunction:

There isn't a circle or else there is a triangle

has the mental models:

¬o
    Δ

The mental models do not represent clauses, whether affirmative or negative, if they are false. Fully explicit models, however, do represent false clauses, using negation where relevant. Hence, the fully explicit models of the preceding disjunction are:

¬o ¬Δ
  o   Δ

Some commentators have argued that the principle of truth is misnamed, because individuals merely represent those propositions that are mentioned in the premises. This view is mistaken. The same propositions can be mentioned in, say, a conjunction and a disjunction, but the mental models of these assertions are very different.

Reasoners focus on what is true and neglect what is false. One consequence is the difficulty of Wason's (1966) selection task. In this task, reasoners have to select evidence relevant to the truth or falsity of an assertion. Given a conditional, such as:

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

they tend to select only an instance of an A, or else instances of an A and of a 2. What they neglect is an instance of a 3. Yet if a 3 occurred in conjunction with an A, the conditional would be false. Any manipulation that helps reasoners to bring to mind the falsifying instance of the conditional improves performance in the selection task (Johnson-Laird, 2001).

Another consequence of the principle of truth was discovered by chance in the output of the program implementing the theory. The neglect of falsity leads to systematic illusions in reasoning. Here is one example (from Goldvarg & Johnson-Laird, 2000):

> Only one of the following premises is true about a particular hand of cards:
> There is a king in the hand or there is an ace, or both.
> There is a queen in the hand or there is an ace, or both.
> There is a jack in the hand or there is a 10, or both.
> Is it possible that there is an ace in the hand?

Intelligent reasoners tend to respond: "Yes". In fact, it is impossible for an ace to be in the hand, because both of the first two premises would then be true, contrary to the rubric that only one premise is true. Most sorts of inference are not illusory, because the neglect of falsity does not affect their validity. But, as the theory predicts, many sorts of illusion do occur in every domain of deduction (e.g., Johnson-Laird & Savary, 1999; Yang & Johnson-Laird, 2000). Experts succumb too. They propose ingenious alternative explanations for their errors. The problems are so complicated, or so artificial, they say, that people are confused. Such explanations, however, overlook that individuals do very well with control problems that are syntactically identical to the illusions.

The illusions corroborate the model theory, but seem wholly inconsistent with other theories of reasoning. Their occurrence is a litmus test for mental models. Certain manipulations alleviate the illusions (e.g., Santamaría & Johnson-Laird, 2000), but the search for a perfect antidote to them has so far been in vain.

Because meaning is central to models, the content of inferences and background knowledge can modulate reasoning (Johnson-Laird & Byrne, 2002). To account for the phenomena, the theory postulates that knowledge takes the form of fully explicit models of those possibilities that are known in detail. The semantic content of a premise is likely to trigger pertinent knowledge of this sort, which is conjoined with the mental models of the assertion, although knowledge normally takes precedence in the case of contradiction. One consequence is the addition of temporal, spatial, and other information to models of assertions. But another important consequence is that knowledge can block the construction of models. The following inference is in the form of a valid modus tollens:

> If Pat is not in Rio then she's in Brazil.
> Pat is not in Brazil.
> Therefore, she is in Rio.

But individuals are reluctant to draw this conclusion. They know that Rio is in Brazil, and so if a person is not in Brazil, then that person cannot be in Rio. Hence, the conditional refers to only two possibilities:

¬ Rio Brazil
Rio Brazil

In contrast, the following inference is easy:

If Pat is in Rio then she is in Brazil.
Pat is not in Brazil.
Therefore, she is not in Rio.

Knowledge readily allows one to draw the conclusion (Johnson-Laird & Byrne, 2002). Analogous phenomena occur as a result of the meaning of clauses, and they extend to disjunctions (Ormerod & Johnson-Laird, 2002).

In logic, connectives such as "if" and "or" have idealized meanings that are truth-functional, that is, the truth or falsity of a sentence they form depends only on the truth or falsity of the clauses they interconnect (Jeffrey, 1981). The preceding examples show that natural language is not truth-functional: the interpretation of a conditional depends on the meanings of its individual clauses and background knowledge. The conditional, *If Pat is not in Rio then she is in Brazil*, rules out the possibility in which both its antecedent and consequent are false, but the conditional, *If Pat is in Rio then she is in Brazil*, does not. Knowledge can also influence the process of reasoning. Reasoners search harder for counterexamples to conclusions that violate their knowledge. This search is compatible with a robust phenomenon: knowledge and beliefs have a bigger effect on invalid inferences than on valid inferences (e.g., Cherubini, Garnham, Oakhill, & Morley, 1998; Evans, 1989; Oakhill, Johnson-Laird, & Garnham, 1989; Santamaría, García-Madruga, & Johnson-Laird, 1998).

STRATEGIES AND COUNTEREXAMPLES

When reasoners make a series of inferences, they develop strategies for coping with them. Different individuals develop different strategies – in sentential reasoning (Byrne & Handley, 1997; Dieussaert, Schaeken, Schroyens, & d'Ydewalle, 2000), in relational reasoning (Roberts, 2000), and in reasoning with quantifiers (Bucciarelli & Johnson-Laird, 1999). Our hypothesis is that individuals develop strategies by trying out different sequences of inferential tactics based on mental models (Van der Henst, Yang, & Johnson-Laird, 2002). Consider, for example, the following problem about marbles in a box:

There is a red marble if and only if there is a green marble.
Either there is a green marble or else there is a blue marble, but not both.
There is a blue marble if and only if there is brown marble.
Does it follow that if there is not a red marble then there is a brown marble?

One strategy is based on following up the consequences of a supposition. Such reasoners say, for instance:

> Assuming there is not a red marble, it follows from the first premise that there is not a green marble. It then follows from the second premise that there is a blue marble. The third premise then implies there is a brown marble. So, yes, the conclusion does follow.

A different strategy is to construct a chain of conditionals leading from one clause in the conditional conclusion to the other. This strategy calls for immediate inferences to convert premises into appropriate conditionals. A "think aloud" protocol contained the following chain, which started invalidly from the consequent of the conditional in the example above:

> If there is a brown marble then there is a blue marble.
> [Immediate inference from premise 3]
> If there is a blue marble then there is not a green marble.
> [Immediate inference from premise 2]
> If there is not a green marble then there is not a red marble.
> [Immediate inference from premise 1]

At this point, the reasoner said that the conclusion followed from the premises.

The strategy that corresponds most directly to the use of mental models is to construct a diagram that integrates all the information from the premises. For example, a participant drew a horizontal line across the middle of the page, and wrote down the two possibilities compatible with the premises in the example above:

Red Green

Blue Brown

Such reasoners work through the premises in whatever order they are stated, taking into account irrelevant premises. When Victoria Bell taught naïve reasoners to use this strategy in a systematic way (in unpublished studies), their reasoning was both faster and more accurate.

Although some strategies are surprising, they can all be based on mental models, and some of them are difficult to explain in any other way. As the model theory predicts, disjunctive premises tend to elicit the incremental diagram strategy, whereas conditional premises tend to elicit the suppositional strategy. Regardless of strategy, however, one-model problems are easier than two-model problems, which in turn are easier than three-model problems. This trend occurs in the correct evaluations of conclusions, and in the validity of conclusions that reasoners draw for themselves (Van der Henst et al., 2002). It supports the hypothesis that all strategies make an underlying use of models.

A rare experimental result contrary to the model theory is Rips's (1994) finding that an inference based on a conjunction was no easier than one based on a disjunction. Rips compared an inference with an initial conjunction, which calls for only one model:

A and B.
If A then C.
If B then C.
Therefore, C.

with an inference with an initial disjunction, which calls for at least two models:

A or B.
If A then C.
If B then C.
Therefore, C.

The participants evaluated the conclusions, and there was no reliable difference between the two sorts of inference. However, in a recent study (García-Madruga, Moreno, Carriedo, Gutiérrez, & Johnson-Laird, 2001) reasoners drew their own conclusions from such premises, and the conjunctive problems were easier than the disjunctive problems. The results also corroborated the model theory when the premises were presented one at a time, and the participants had to evaluate the conclusions. Rips's procedure may therefore have elicited a different strategy from those that the participants developed in the García-Madruga studies.

From its inception, the model theory has postulated that reasoners could reject invalid conclusions on the basis of counterexamples, i.e., models in which the premises are true but the conclusion is false (see also Halpern & Vardi, 1991). But such a model violates the principle of truth, and reasoners do not invariably search for counterexamples (Newstead, Thompson, & Handley, 2002; Polk & Newell, 1995). One way to elicit them is to ask reasoners to evaluate given conclusions that are invalid, for example:

More than half the people in the room speak English.
More than half the people in the room speak Italian.
Does it follow that more than half the people in the room speak both languages?

A typical response is that there could be five people in the room, three speak one language, three speak the other language, but only one person speaks both languages (Neth & Johnson-Laird, 1999).

The use of counterexamples is just one strategy in refuting invalid inferences (Johnson-Laird & Hasson, 2003). With an inference of the following form, for example:

If A then not B.
B or C.
Therefore, A or C.

all the participants in an experiment constructed a counterexample: not-A, B, not-C. But, given an inference of the form:

If A then B.
If B then C.
Therefore, C.

the participants remarked that nothing definite could follow from two conditionals. In other cases, they pointed out the need for a missing premise, or generated a valid conclusion that they contrasted with the given conclusion. As the theory predicts, however, the use of counterexamples is more frequent when an invalid conclusion is consistent with the premises rather than inconsistent with them.

The competence to use counterexamples is contrary to theories of reasoning in which invalidity is established solely by a failure to find a proof (Braine & O'Brien, 1998; Rips, 1994). But it is consistent with the model theory's claim that human rationality rests on the fundamental semantic principle of validity: an inference is valid if its conclusion holds in all the possibilities – the models – consistent with the premises. An application of this principle to invalidity is to construct counterexamples.

CONCLUSIONS

If Craik (1943) is right, then mental models underlie all sorts of thinking from induction to creation. So far, however, the theory has focused on deduction (see the web page maintained by Ruth Byrne and her colleagues at www.tcd.ie/Psychology/Ruth_Byrne/mental_models/). Its three main principles owe something to Craik, something to his precursors, and something to those who inherited his ideas:

(1) Reasoners use the meaning of premises and their knowledge to construct mental models of the possibilities compatible with the premises.
(2) Mental models are iconic as far as possible, but certain components of them are necessarily symbolic.
(3) Mental models represent what is true, but not what is false. Reasoners can – with some difficulty – flesh them out into fully explicit models.

The theory provides a single psychological mechanism for deductions about necessary, probable, and possible conclusions. A conclusion that holds in all possible models of the premises is necessary given the premises. It is not necessary if it has a counterexample, i.e., a model of the premises in which the conclusion is false. A conclusion that holds in most of the models of the premises is probable. Reasoners can estimate the probability of a conclusion based on the proportion of equipossible models in which it holds, or from calculating its probability from models tagged with numerical frequencies or chances of occurrence (Girotto & Gonzalez, 2001; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). A conclusion that holds in at least one model of the premises is possible.

In recent years, the model theory has been extended to many domains:

- Counterfactual reasoning (Byrne & McEleney, 2000; Byrne & Tasso, 1999).
- Reasoning based on suppositions (Byrne & Handley, 1997).
- Modal reasoning about possibilities (Bell & Johnson-Laird, 1998; Evans, Handley, Harper, & Johnson-Laird, 1999).
- Deontic reasoning about obligations (Bucciarelli & Johnson-Laird, 2002; Manktelow & Over, 1995).
- Causal reasoning (Goldvarg & Johnson-Laird, 2001).
- The detection of inconsistencies and their resolution (Girotto, Johnson-Laird, Legrenzi, & Sonino, 2000; Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000).
- Strategic reasoning in games (Steingold & Johnson-Laird, 2002).
- The construction of arguments (Green & McManus, 1995).

Opponents of the theory have often criticized it in helpful ways. Many of its details do need to be clarified. And, despite some steps in the right direction (Bara, Bucciarelli, & Lombardo, 2001), its single biggest weakness is its lack of a comprehensive account of reasoning based on the interplay between quantifiers and connectives.

Yet the theory seems to be on the right lines. Perhaps the best evidence in its favour comes from its unexpected prediction of illusory inferences, which have now been confirmed for many sorts of reasoning. Other theories seem to have no way to account for such systematic errors short of postulating invalid rules of inference – a step with disastrous implications for human rationality. In contrast, failures to construct the correct models are predictable from the principle of truth, but do not impugn rationality based on the semantic principle of validity. The model theory does not imply that reasoners never rely on rules of inference. Intelligent individuals may develop rules spontaneously as a result of experience with many inferences of a similar form. And this step, in principle, can lead to the development of formal logic.

ACKNOWLEDGEMENTS

Preparation of this article was supported by a grant from the National Science Foundation (Grant BCS 0076287) to study strategies in reasoning. The author thanks the community of researchers on deductive reasoning for their help. Their names are in the following section.

REFERENCES

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation, Vol. 8*. London: Academic Press.

Bara, B. G., Bucciarelli, M., & Lombardo, V. (2001). Model theory of deduction: A unified computational approach. *Cognitive Science, 25*, 839–901.

Barres, P., & Johnson-Laird, P. N. (2002). On imagining what is true (and what is false). *Thinking & Reasoning, 9*, 1–42.

Barrouillet, P., & Lecas, J.-F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, 5, 289–302.

Barrouillet, P., Grosset, N., & Lecas, J.-F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition, 75*, 237–266.

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

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

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Binet, A. (1894). *Psychologie des grands calculateurs et joueurs d'echecs*. Paris: Hachette.

Boltzmann, L. (1890/1974). On the significance of physical theories. In B. McGuiness (Ed.), *Ludwig Boltzmann: Theoretical physics and philosophical problems* (pp. 33–36). Boston: Reidel. [Originally published, 1890.]

Boltzmann, L. (1899/1974). On the fundamental principles and equations of mechanics. In B. McGuiness (Ed.), *Ludwig Boltzmann: Theoretical physics and philosophical problems* (pp. 101–128). Boston: Reidel. [Originally published, 1899.]

Bower, G. H. (1970) Analysis of the mnemonic device. *American Scientist, 58*, 496–501.

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

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus an interpretive approach. *Cognitive Psychology, 3*, 193–209.

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

Bucciarelli, M., & Johnson-Laird, P. N. (2002). *Foundations of deontic meaning and reasoning*. Manuscript submitted for publication.

Byrne, R. M. J., & Handley, S. J. (1997). Reasoning strategies for suppositional deductions. *Cognition, 62*, 1–49.

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

Byrne, R. M. J., & McEleney, A. (2000). Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1318–1331.

Byrne, R. M. J., & Tasso, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory & Cognition, 27*, 726–740.

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

Cherubini, P., Garnham, A., Oakhill, J., & Morley, E. (1998). Can any ostrich fly? Some new data on belief bias in syllogistic reasoning. *Cognition, 69*, 179–218.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517.

Craik, K. (1943). *The nature of explanation.* Cambridge: Cambridge University Press.

Davis, R., & Hamscher, W. (1988). Model-based reasoning: Troubleshooting. In H. E. Shrobe & AAAI (Eds.), *Exploring artificial intelligence: Survey talks from the National Conferences on Artificial Intelligence.* San Mateo, CA: Morgan Kaufmann.

de Kleer, J. (1977). Multiple representations of knowledge in a mechanics problem-solver. *International Joint Conference Artificial Intelligence*, 299–304.

de Kleer, J., & Williams, B. C. (1987). Diagnosing multiple faults. *Artificial Intelligence, 32*, 97–130.

Dieussaert, K., Schaeken, W., Schroyens, W., & d'Ydewalle, G. (2000). Strategies during complex conditional inferences. *Thinking & Reasoning, 6*, 125–160.

Ehrlich, K. (1996). Applied mental models in human–computer interaction. In J. Oakhill & A. Garnham (Eds.), *Mental models in cognitive science.* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Erickson, J. R. (1974). A set analysis theory of behaviour in formal syllogistic reasoning tasks. In R. Solso (Ed.), *Loyola Symposium on Cognition, Vol. 2.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Espino, O., Santamaría, C., Meseguer, E., & Carreiras, M. (2000). Eye movements during syllogistic reasoning. In J. A. García-Madruga, N. Carriedo, & M. J. González-Labra (Eds.), *Mental models in reasoning* (pp. 179–188). Madrid: Universidad Nacional de Educación a Distancia.

Evans, J. St. B. T., (1989). *Bias in human reasoning: Causes and consequences.* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

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.* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Feynman, R. P. (1985). *QED: The strange theory of light and matter.* Princeton, NJ: Princeton University Press.

Forbus, K. (1985). Qualitative process theory. In D. G. Bobrow (Ed.), *Qualitative reasoning about physical systems.* Cambridge, MA: MIT Press.

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition, 54*, 1–71.

Frege, G. (1879). *Begriffsschrift, eine der Arithmetischen nachgebildete Formelsprache des reinen Denkens.* Halle: Nebert.

García-Madruga, J. A., Moreno, S., Carriedo, N., Gutiérrez, F., & Johnson-Laird, P. N. (2001). Are conjunctive inferences easier than disjunctive inferences? A comparison of rules and models. *Quarterly Journal of Experimental Psychology, 54A*, 613–632.

Garnham, A. (1987). *Mental models as representations of discourse and text.* Chichester, UK: Ellis Horwood.

Garnham, A. (2001). *Mental models and the interpretation of anaphora.* Hove, UK: Psychology Press.

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: Lawrence Erlbaum Associates Inc.

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of question form and information structure. *Cognition, 78*, 247–276.

Girotto, V., Johnson-Laird, P. N., Legrenzi, P., & Sonino, M. (2000). Reasoning to consistency: How people resolve logical inconsistencies. In J. A. García-Madruga, N. Carriedo & M. J. González-Labra (Eds.), *Mental models in reasoning* (pp. 83–97). Madrid: Universidad Nacional de Educación a Distancia.

Girotto, V., Mazzocco, A., & Tasso, A. (1997). The effect of premise order in conditional reasoning: A test of the mental model theory. *Cognition, 63*, 1–28.

Glasgow, J. I. (1993). Representation of spatial models for geographic information systems. In N. Pissinou (Ed.), *Proceedings of the ACM Workshop on Advances in Geographic Information Systems* (pp. 112–117). Arlington, VA: Association for Computing Machinery.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26*, 69–83.

Goldvarg, Y., & Johnson-Laird, P. N. (2000). Illusions in modal reasoning. *Memory & Cognition, 28*, 282–294.

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

Green, D. W., & McManus, I. C. (1995). Cognitive structural models: Perception of risk and prevention in coronary heart disease. *British Journal of Psychology, 86*, 321–335.

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

Halpern, J. Y., & Vardi, M. Y. (1991). Model checking vs. theorem proving: A manifesto. In J. A. Allen, R. Fikes & E. Sandewall (Eds.), *Principles of knowledge representation and reasoning: Proceedings of the Second International Conference* (pp. 325–334). San Mateo, CA: Kaufmann.

Hayes, P. J. (1979). Naive physics I – ontology for liquids. Mimeo, Centre pour les études Semantiques et Cognitives, Geneva. [Reprinted in Hobbs, J., & Moore, R. (Eds.), *Formal theories of the commonsense world.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.]

Hegarty, M. (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1084–1102.

Holland, J. H. (1998). *Emergence: From chaos to order*. Reading, MA: Perseus Books.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

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

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

Johnson-Laird, P. N. (1970). The perception and memory of sentences. In J. Lyons (Ed.), *New horizons in linguistics* (pp. 261–270). Harmondsworth, UK: Penguin Books.

Johnson-Laird, P. N. (1975). Models of deduction. In R. Falmagne (Ed.), *Reasoning: Representation and process*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Science, 5*, 434–442.

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

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

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

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., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review, 96*, 658–673.

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

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

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

Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition, 71*, 191–229.

Johnson-Laird, P. N., & Steedman, M. J. (1978). The psychology of syllogisms. *Cognitive Psychology, 10*, 64–99.

Johnson-Laird, P. N., & Stevenson, R. (1970). Memory for syntax. *Nature, 227*, 412.

Kamp, H. (1981). A theory of truth and semantic representation. In J. A. G. Groenendijk, T. M. V. Janssen & M. B. J. Stokhof (Eds.), *Formal methods in the study of language* (pp. 277–322). Amsterdam: Mathematical Centre Tracts.

Karttunen, L. (1976). Discourse referents. In J. D. McCawley (Ed.), *Syntax and semantics, Vol. 7: Notes from the linguistic underground*. New York: Academic Press.

Kitchin, R. M. (1994). Cognitive maps: What are they and why study them? *Journal of Environmental Psychology, 14*, 1–19.

Klauer, K. C., & Oberauer, K. (1995). Testing the mental model theory of propositional reasoning. *Quarterly Journal of Experimental Psychology, 48*A, 671–687.

Knauff, M., Fangmeir, 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.

Köhler, W. (1938). *The place of value in a world of facts.* New York: Liveright.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 47–60.

Kuipers, B. (1994). *Qualitative reasoning: Modeling and simulation with incomplete knowledge.* Cambridge, MA: MIT Press.

Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science, 11*, 65–99.

Luria, A. R. (1969). *The mind of a mnemonist.* London: Cape.

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

Maier, N. R. F. (1929). Reasoning in white rats. *Comparative Psychology Monographs, 6*, No. 29.

Manktelow, K. I., & Over. D. E. (1995). Deontic reasoning. In S. E. Newstead & J. St.B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 91–114). Hove, UK: Psychology Press.

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

Markovits, H. (2000). A mental model analysis of young children's conditional reasoning with meaningful premises. *Thinking & Reasoning, 6*, 335–347.

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

Maxwell, J. C. (1911). *Diagram. The Encyclopaedia Britannica, Vol. XVIII.* New York: The Encylopaedia Britannica Company.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motions of objects. *Science, 210*, 1139–1141.

McCulloch, W. S. (1965). *Embodiments of mind.* Cambridge, MA: MIT 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.* Cambridge, MA: MIT Press. [Originally published in Solso, R. L. (Ed.), *Theories in cognitive psychology: The Loyola Symposium.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc, 1974.]

Moray, N. (1990). A lattice theory approach to the structure of mental models. *Philosophical Transactions of the Royal Society of London B, 327*, 577–583.

Moray, N. (1999). Mental models in theory and practice. In D. Gopher & A. Koriat (Eds.), *Attention & performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 223–358). Cambridge, MA: MIT Press.

Neth, H., & Johnson-Laird, P. N. (1999). The search for counterexamples in human reasoning. *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*, 806.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newstead, S. E., Thompson, V. A., & Handley, S. J. (2002). Generating alternatives: A key component in human reasoning? *Memory & Cognition, 30*, 129–137.

Oakhill, J. V., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition, 31*, 117–140.

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

Ormerod, T. C., & Johnson-Laird, P. N. (2002). *How pragmatics modulates the interpretation of sentential connectives*. Manuscript submitted for publication.

Osherson, D. N. (1974–6). *Logical abilities in children, Vols. 1–4*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

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

Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology, 21*, 422–452.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102*, 533–566.

Putnam, H. (1960). Minds and machines. In S. Hook (Ed.), *Dimensions of mind*. New York: New York University Press.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80*, 1–24.

Richardson, J., & Ormerod, T. C. (1997). Rephrasing between disjunctives and conditionals: Mental models and the effects of thematic content. *Quarterly Journal of Experimental Psychology, 50*A, 358–385.

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.

Rouse, W. B., & Hunt, R. M. (1984). Human problem solving in fault diagnosis tasks. In W. B. Rouse (Ed.), *Advances in man–machine systems research*. Greenwich, CN: JAI Press.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*, 437–442.

Santamaría, C., García-Madruga, J. A., & Johnson-Laird, P. N. (1998). Reasoning from double conditionals: The effects of logical structure and believability. *Thinking & Reasoning, 4*, 97–122.

Santamaría, C., & Johnson-Laird, P. N. (2000). An antidote to illusory inferences. *Thinking & Reasoning, 6*, 313–333.

Schaeken, W., 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.

Schwartz, D., & Black, J. B. (1996). Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology, 30*, 154–219.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 71*, 701–703.

Sloutsky, V. M., & Goldvarg, Y. (1999). Effects of externalization on representation of indeterminate problems. *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*, 695–700.

Sloutsky, V. M., & Johnson-Laird. P. N. (1999). Problem representations and illusions in reasoning. *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*, 701–705.

Smith, C., & Wise, M. N. (1989). *Energy and empire: A biographical study of Lord Kelvin*. Cambridge: Cambridge University Press.

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

Steingold, E., & Johnson-Laird, P. N. (2002). Naïve strategic thinking. In W. Gray & C.D. Schunn (Eds), Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society, Fairfax, VA (pp. 845–849). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Stenning, K. (1977). Articles, quantifiers, and their encoding in textual comprehension. In M. Halle, J. Bresnan & G. A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Stenning, K. (2002). *Seeing reason: Image and language in learning to think*. Oxford: Oxford University Press.

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

Stevenson, R. J. (1993). *Language, thought and representation*. New York: Wiley.

Störring, G. (1908). Experimentelle Untersuchungen über einfache Schlussprozesse. *Archiv für die gesamte Psychologie*, *11*, 1–27.

Thorndyke, P. W., & B. Hayes-Roth. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, *14*, 560–589.

Tolman, E. C. (1932). *Purposive behaviour in animals and men*. New York: The Century Co.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208. [Reprinted in Tolman, E. C. (1958). *Behaviour and psychological man: Essays in motivation and learning*. (Originally published as *Collected papers in psychology*, 1951) Berkeley: University of California Press.]

Tolman, E. C. (1959). Principles of purposive behaviour. In S. Koch (Ed.), *psychology: A study of a science, Study 1. Conceptual and systematic, Vol. 2. General systematic formulations, learning, and special processes* (pp. 92–157). New York: McGraw-Hill.

Tolman, E. C., and Honzik, C. H. (1930). "Insight" in rats. *University of California Publications in Psychology*, *4*, 257–275.

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 concepts. *Thinking & Reasoning*, *2*, 249–272.

Vandierendonck, A., & De Vooght, G. (1997). Working memory constraints on linear reasoning with spatial and temporal contents. *Quarterly Journal of Experimental Psychology*, *50*A, 803–820.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

von Frisch, K. (1966). *The dancing bees* (2nd Ed.). London: Methuen.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585.

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology, 21*, 92–107.

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

Wason, P. C., & Johnson-Laird, P. N. (1972). *The psychology of reasoning*. London: Batsford. Cambridge, MA: Harvard University Press.

Webber, B. L. (1978). Description formation and discourse model synthesis. In D. L. Waltz (Ed.), *Theoretical issues in natural language processing, 2*. New York: Association for Computing Machinery.

Wise, M. N. (1979). The mutual embrace of electricity and magnetism. *Science, 203*, 1310–1318.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Routledge & Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

Yang, Y., & Johnson-Laird, P. N. (2000). Illusory inferences in quantified reasoning: How to make the impossible seem possible, and vice versa. *Memory & Cognition, 28*, 452–465.

Young, R. (1981). The machine inside the machine: Users' models of pocket calculators. *International Journal of Man Machine Studies, 15*, 51–85.