CHAPTER 9
Mental Models and Thought

P. N. Johnson-Laird

How do we think? One answer is that we rely on *mental models*. Perception yields models of the world that lie outside us. An understanding of discourse yields models of the world that the speaker describes to us. Thinking, which enables us to anticipate the world and to choose a course of action, relies on internal manipulations of these mental models. This chapter is about this theory, which it refers to as the *model* theory, and its experimental corroborations. The theory aims to explain all sorts of thinking about propositions, that is, thoughts capable of being true or false. There are other sorts of thinking – the thinking, for instance, of a musician who is improvising. In daily life, unlike the psychological laboratory, no clear demarcation exists between one sort of thinking and another. Here is a protocol of a typical sequence of everyday thoughts:

> *I had the book in the hotel's restaurant, and now I've lost it. So, either I left it in the restaurant, or it fell out of my pocket on the way back to my room, or it's somewhere here in my room. It couldn't have fallen from my pocket – my pockets are deep and I walked slowly back to my room – and so it's here or in the restaurant.*

Embedded in this sequence is a logical deduction of the form:

> A or B or C.
> Not B.
> Therefore, A or C.

The conclusion is *valid:* It must be true given that the premises are true. However, other sorts of thinking occur in the protocol (e.g., the inference that the book could not have fallen out of the protagonist's pocket).

A simple way to categorize thinking about propositions is in terms of its effects on semantic information (Johnson-Laird, 1993). The more possibilities an assertion rules out, the greater the amount of semantic information it conveys (Bar-Hillel & Carnap, 1964). Any step in thought from current premises to a new conclusion therefore falls into one of the following categories:

- The premises and the conclusion eliminate the same possibilities.

- The premises eliminate at least one more possibility over those the conclusion eliminates.
- The conclusion eliminates at least one more possibility over those the premises eliminate.
- The premises and conclusion eliminate disjoint possibilities.
- The premises and conclusion eliminate overlapping possibilities.

The first two categories are deductions (see Evans, Chapter 11). The third category includes all the traditional cases of induction, which in general is definable as any thought yielding such an increase in semantic information (see Sloman & Lagnado, Chap. 3). The fourth category occurs only when the conclusion is inconsistent with the premises. The fifth case occurs when the conclusion is consistent with the premises but refutes at least one premise and adds at least one new proposition. Such thinking goes beyond induction. It is associative or creative (see Sternberg, Chap. 13).

The model theory aims to explain all propositional thinking, and this chapter illustrates its application to the five preceding categories. The chapter begins with the history of the model theory. It then outlines the current theory and its account of deduction. It reviews some of the evidence for this account. It shows how the theory extends to probabilistic reasoning. It then turns to induction, and it describes the unconscious inferences that occur in understanding discourse. It shows how models underlie causal relations and the creation of explanations. Finally, it assesses the future of the model theory.

The History of Mental Models

In the seminal fifth chapter of his book, *The Nature of Explanation*, Kenneth Craik (1943) 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.*

This same process of internal imitation of the external world, Craik wrote, is carried out by mechanical devices such as Kelvin's tidal predictor. Craik died in 1945, before he could develop his ideas. Several earlier thinkers had, in fact, anticipated him (see Johnson-Laird, 2003). Nineteenth-century physicists, including Kelvin, Boltzmann, and Maxwell, stressed the role of models in thinking. In the twentieth century, physicists downplayed these ideas with the advent of quantum theory (but cf. Deutsch, 1997).

One principle of the modern theory is that the parts of a mental model and their structural relations correspond to those which they represent. This idea has many antecedents. It occurs in Maxwell's (1911) views on diagrams, in Wittgenstein's (1922) "picture" theory of meaning, and in Köhler's (1938) hypothesis of an isomorphism between brain fields and the world. However, the nineteenth-century grandfather of the model theory is Charles Sanders Peirce.

Peirce coinvented the main system of logic known as *predicate* calculus, which governs sentences in a formal language containing idealized versions of negation, sentential connectives such as "and" and "or," and quantifiers such as "all" and "some." Peirce devised two diagrammatic systems of reasoning, not to improve reasoning, but to display its underlying mental steps (see Johnson-Laird, 2002). He wrote:

> *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 (Peirce, 1.66; this standard notation refers to paragraph 66 of Volume 1 of Peirce, 1931–1958).

Diagrams can be *iconic*, in other words, have the same structure as what they represent (Peirce, 4.447). It is the inspection of an iconic diagram that reveals truths other than those of the premises (2.279, 4.530). Hence, Peirce anticipates Maxwell, Wittgenstein, Köhler, and the model theory. Mental models are as iconic as possible (Johnson-Laird, 1983, pp. 125, 136).

A resurgence of mental models in cognitive science began in the 1970s. Theorists proposed that knowledge was represented in mental models, but they were not wed to any particular structure for models. Hayes (1979) used the predicate calculus to describe the naive physics of liquids. Other theorists in artificial intelligence proposed accounts of how to envision models and use them to simulate behavior (de Kleer, 1977). Psychologists similarly examined naive and expert models of various domains, such as mechanics (McCloskey, Caramazza, & Green, 1980) and electricity (Gentner & Gentner, 1983). They argued that vision yields a mental model of the three-dimensional structure of the world (Marr, 1982). They proposed that individuals use these models to simulate behavior (e.g., Hegarty, 1992; Schwartz & Black, 1996). They also studied how models develop (e.g., Vosniadou & Brewer, 1992; Halford, 1993), how they serve as analogies (e.g., Holland, Holyoak, Nisbett, & Thagard, 1986; see Holyoak, Chap. 6), and how they help in the diagnosis of faults (e.g., Rouse & Hunt, 1984). Artifacts, they argued, should be designed so users easily acquire models of them (e.g., Ehrlich, 1996; Moray, 1990, 1999).

Discourse enables humans to experience the world by proxy, and so another early hypothesis was that comprehension yields models of the world (Johnson-Laird, 1970). The models are iconic in these ways: They contain a token for each referent in the discourse, properties corresponding to the properties of the referents, and relations corresponding to the relations among the referents. Similar ideas occurred in psycholinguistics (e.g., Bransford, Barclay, & Franks, 1972), linguistics (Karttunen, 1976), artificial intelligence (Webber, 1978), and formal semantics (Kamp, 1981). Experimental evidence corroborated the hypothesis, showing that individuals rapidly forget surface and underlying syntax (Johnson-Laird & Stevenson, 1970), and even the meaning of individual sentences (Garnham, 1987). They retain only models of who did what to whom. Psycholinguists discovered that models are constructed from the meanings of sentences, general knowledge, and knowledge of human communication (e.g., Garnham, 2001; Garnham & Oakhill, 1996; Gernsbacher, 1990; Glenberg, Meyer, & Lindem, 1987).

Another early discovery was that content affects deductive reasoning (Wason & Johnson-Laird, 1972; see Evans, Chap. 8), which was hard to reconcile with the then dominant view that reasoners depend on *formal* rules of inference (Braine, 1978; Johnson-Laird, 1975; Osherson, 1974–1976). Granted that models come from perception and discourse, they could be used to reason (Johnson-Laird, 1975): An inference is *valid* if its conclusion holds in all the models of the premises because its conclusion must be true granted that its premises are true. The next section spells out this account.

Models and Deduction

Mental models represent entities and persons, events and processes, and the operations of complex systems. However, what is a mental model? The current theory is based on principles that distinguish models from linguistic structures, semantic networks, and other proposed mental representations (Johnson-Laird & Byrne, 1991). The first principle is

> *The principle of* iconicity: *A mental model has a structure that corresponds to the known structure of what it represents.*

Visual images are iconic, but mental models underlie images. Even the rotation of

mental images implies that individuals rotate three-dimensional models (Metzler & Shepard, 1982), and irrelevant images impair reasoning (Knauff, Fangmeir, Ruff, & Johnson-Laird, 2003; Knauff & Johnson-Laird, 2002). Moreover, many components of models cannot be visualized.

One advantage of iconicity, as Peirce noted, is that models built from premises can yield new relations. For example, Schaeken, Johnson-Laird, and d'Ydewalle (1996) investigated problems of temporal reasoning concerning such premises as

John eats his breakfast before he listens to the radio.

Given a problem based on several premises with the form:

A before B.
B before C.
D while A.
E while C.

reasoners can build a mental model with the structure:

A	B	C
D		E

where the left-to-right axis is time, and the vertical axis allows different events to be contemporaneous. Granted that each event takes roughly the same amount of time, reasoners can infer a new relation:

D before E.

Formal logic less readily yields the conclusion. One difficulty is that an infinite number of conclusions follow validly from any set of premises, and logic does not tell you *which* conclusions are useful. From the previous premises, for instance, this otiose conclusion follows:

A before B, *and* B before C.

Possibilities are crucial, and the second principle of the theory assigns them a central role:

The principle of possibilities: *Each mental model represents a possibility.*

Table 9.1. The Truth Table for Exclusive Disjunction

A	*B*	*A or else B, but not both*
True	True	False
True	False	True
False	True	True
False	False	False

This principle is illustrated in *sentential* reasoning, which hinges on negation and such sentential connectives as "if" and "or." In logic, these connectives have idealized meanings: They are *truth-functional* in that the truth-values of sentences formed with them depend solely on the truth-values of the clauses that they connect. For example, a disjunction of the form: *A or else B but not both* is true if *A* is true and *B* is false, and if *A* is false and *B* is true, but false in any other case. Logicians capture these conditions in a truth table, as shown in Table 9.1. Each row in the table represents a different possibility (e.g., the first row represents the possibility in which both *A* and *B* are true), and so here the disjunction is false.

Naive reasoners do not use truth tables (Osherson, 1974–1976). *Fully explicit* models of possibilities, however, are a step toward psychological plausibility. The fully explicit models of the exclusive disjunction, *A or else B but not both*, are shown here on separate lines:

A	¬B
¬A	B

where "¬" denotes negation. Table 9.2 presents the fully explicit models for the main sentential connectives. Fully explicit models correspond exactly to the true rows in the truth table for each connective. As the table shows, the conditional *If A then B* is treated in logic as though it can be paraphrased as *If A then B, and if not-A then B or not-B*. The paraphrase does not do justice to the varied meanings of everyday conditionals (Johnson-Laird & Byrne, 2002). In fact, no connectives in natural language are truth

Table 9.2. Fully Explicit Models and Mental Models of Possibilities Compatible with Sentences Containing the Principal Sentential Connectives

Sentences	*Fully Explicit Models*		*Mental Models*	
A and B:	A	B	A	B
Neither A nor B:	¬A	¬B	¬A	¬B
A or else B but not both:	A	¬B	A	
	¬A	B		B
A or B or both:	A	¬B	A	
	¬A	B		B
	A	B	A	B
If A then B:	A	B	A	B
	¬A	B	...	
	¬A	¬B		
If, and only if A, then B:	A	B	A	B
	¬A	¬B	...	

functional (see the section on implicit induction and the modulation of models).

Fully explicit models yield a more efficient reasoning procedure than truth tables. Each premise has a set of fully explicit models, for example, the premises:

1. A or else B but not both.
2. Not-A.

have the models:

(Premise 1)		(Premise 2)
A	¬B	¬A
¬A	B	

Their conjunction depends on combining each model in one set with each model in the other set according to two main rules:

- A contradiction between a pair of models yields the null model (akin to the empty set).
- Any other conjunction yields a model of each proposition in the two models.

The result is:

Input from (1)		Input from (2)	Output	
A	¬B	¬A	null model	
¬A	B	¬A	¬A	B

or in brief:

¬A B

Because an inference is valid if its conclusion holds in all the models of the premises, it follows that: *B*. The same rules are used recursively to construct the models of compound premises containing multiple connectives.

Because infinitely many conclusions follow from any premises, computer programs for proving validity generally evaluate conclusions given to them by the user. Human reasoners, however, can draw conclusions for themselves. They normally abide by two constraints (Johnson-Laird & Byrne, 1991). First, they do not throw semantic information away by adding disjunctive alternatives. For instance, given a single premise, *A*, they never spontaneously conclude, *A or B or both*. Second, they draw novel conclusions that are parsimonious. For instance, they never draw a conclusion that merely conjoins the premises, even though such a deduction is valid. Of course, human performance rapidly degrades with complex problems, but the goal of parsimony suggests that intelligent programs should draw conclusions that succinctly express all the information in the premises. The model theory yields an algorithm that draws

such conclusions (Johnson-Laird & Byrne, 1991, Chap. 9).

Fully explicit models are simpler than truth tables but place a heavy load on working memory. *Mental* models are still simpler because they are limited by the third principle of the theory:

> *The principle of* truth: *A mental model represents a true possibility, and it represents a clause in the premises only when the clause is true in the possibility.*

The simplest illustration of the principle is to ask naive individuals to list what is possible for a variety of assertions (Barrouillet & Lecas, 1999; Johnson-Laird & Savary, 1996). Given an exclusive disjunction, *not-A or else B*, they list two possibilities corresponding to the mental models:

¬A
B

The first mental model does not represent *B*, which is false in this possibility; and the second mental model does not represent *not-A*, which is false in this possibility, in other words, *A* is true. Hence, people tend to neglect these cases. Readers might assume that the principle of truth is equivalent to the representation of the propositions *mentioned* in the premises. However, this assumption yields the same models of *A* and *B* regardless of the connective relating them. The right way to conceive the principle is that it yields pared-down versions of fully explicit models, which in turn map into truth tables. As we will see, the principle of truth predicts a striking effect on reasoning.

Individuals can make a mental footnote about what is false in a possibility, and these footnotes can be used to flesh out mental models into fully explicit models. However, footnotes tend to be ephemeral. The most recent computer program implementing the model theory operates at two levels of expertise. At its lowest level, it makes no use of footnotes. Its representation of the main sentential connectives is summarized in Table 9.2. The mental models of a conditional, *if A then B*, are

A B
. . .

The ellipsis denotes an *implicit* model of the possibilities in which the antecedent of the conditional is false. In other words, there are alternatives to the possibility in which *A* and *B* are true, but individuals tend not to think explicitly about what holds in these possibilities. If they retain the footnote about what is false, then they can flesh out these mental models into fully explicit models. The mental models of the biconditional, *If, and only if, A then B*, as Table 9.2 shows, are identical to those for the conditional. What differs is that the footnote now conveys that both *A* and *B* are false in the implicit model. The program at its higher level uses fully explicit models and so makes no errors in reasoning.

Inferences can be made with mental models using a procedure that builds a set of models for a premise and then updates them according to the other premises. From the premises,

A or else B but not both.
Not-A.

the disjunction yields the mental models

A
B

The categorical premise eliminates the first model, but it is compatible with the second model, yielding the valid conclusion, *B*. The rules for updating mental models are summarized in Table 9.3.

The model theory of deduction began with an account of reasoning with quantifiers as in *syllogisms* such as:

Some actuaries are businessmen.
All businessmen are conformists.
Therefore, some actuaries are conformists.

A plausible hypothesis is that people construct models of the possibilities compatible with the premises and draw whatever conclusion, if any, holds in all of them. Johnson-Laird (1975) illustrated such an account with Euler circles. A premise of the form, *Some A are B*, however, is compatible with four distinct possibilities, and the previous premises are compatible with 16 distinct possibilities. Because the inference is easy, reasoners may fail to consider

Table 9.3. The procedures for forming a conjunction of a pair of models. Each procedure is presented with an accompanying example. Only mental models may be implicit and therefore call for the first two procedures

1: The conjunction of a pair of implicit models yields the implicit model:

... and ... yield ...

2: The conjunction of an implicit model with a model representing propositions yields the null model (akin to the empty set) by default, for example,

... and B C yield nil.

But, if none of the atomic propositions (B C) is represented in the set of models containing the implicit model, then the conjunction yields the model of the propositions, for example,

... and B C yield B C.

3: The conjunction of a pair of models representing respectively a proposition and its negation yield the null model, for example,

A ¬B and ¬A yield nil.

4: The conjunction of a pair of models in which a proposition, B, in one model is not represented in the other model depends on the set of models of which this other model is a member. If B occurs in at least one of these models, then its absence in the current model is treated as negation, for example,

A B and A yields nil.

However, if B does not occur in one of these models (e.g., only its negation occurs in them), then its absence is treated as equivalent to its affirmation, and the conjunction (following the next procedure) is

A B and A yields A B.

5: The conjunction of a pair of fully explicit models free from contradiction update the second model with all the new propositions from the first model, for example,

¬A B and ¬A C yield ¬A B C.

all the possibilities (Erickson, 1974), or they may construct models that capture more than one possibility (Johnson-Laird & Bara, 1984). The program implementing the model theory accordingly constructs just one model for the previous premises:

actuary	[businessman]	conformist
actuary		
	[businessman]	conformist
	. . .	

where each row represents a different sort of individual, the ellipsis represents the possibility of other sorts of individual, and the square brackets represent that the set of businessmen has been represented exhaustively – in other words, no more tokens representing businessmen can be added to the model. This model yields the conclusion that *Some actuaries are conformists*. There are many ways in which reasoners might use such models, and Johnson-Laird and Bara (1984) described two alternative strategies. Years of tinkering with the models for syllogisms suggest that reasoning does not rely on a single deterministic procedure. The following principle applies to thinking in general but can be illustrated for reasoning:

> *The principle of strategic variation: Given a class of problems, reasoners develop a variety of strategies from exploring manipulations of models (Bucciarelli & Johnson-Laird, 1999).*

Stenning and his colleagues anticipated this principle in an alternative theory of syllogistic reasoning (e.g., Stenning & Yule, 1997). They proposed that reasoners focus on individuals who necessarily exist given the premises (e.g., given the premise *Some A are B*, there must be an *A* who is *B*). They implemented this idea in three different algorithms that all yield the same inferences. One algorithm is based on Euler circles supplemented with a notation for

necessary individuals, one is based on tokens of individuals in line with the model theory, and one is based on verbal rules, such as

> *If there are two existential premises, that is, that contain "some", then respond that there is no valid conclusion.*

Stenning and Yule concluded from the equivalence of the outputs from these algorithms that a need exists for data beyond merely the conclusions that reasoners draw, and they suggested that reasoners may develop different representational systems, depending on the task. Indeed, from Störring (1908) to Stenning (2002), psychologists have argued that some reasoners may use Euler circles and others may use verbal procedures.

The *external* models that reasoners constructed with cut-out shapes corroborated the principle of strategic variation: Individuals develop various strategies (Bucciarelli & Johnson-Laird, 1999). They also overlook possible models of premises. Their search may be organized toward finding necessary individuals, as Stenning and Yule showed, but the typical representations of premises included individuals who were not necessary; for example, the typical representation of *Some A are B* was

A B
A B
A

A focus on necessary individuals is a particular strategy. Other strategies may call for the representation of other sorts of individuals, especially if the task changes – a view consistent with Stenning and Yule's theory. For example, individuals readily make the following sort of inference (Evans, Handley, Harper, & Johnson-Laird, 1999):

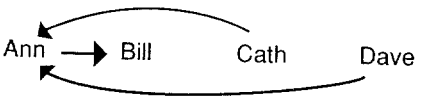
> Some A are B.
>
> Some B are C.
>
> Therefore, it is possible that Some A are C.

Such inferences depend on the representation of possible individuals.

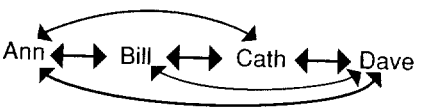
The model theory has been extended to some sorts of inference based on premises containing more than one quantifier (Johnson-Laird, Byrne, & Tabossi, 1989). Many such inferences are beyond the scope of Euler circles, although the general principles of the model theory still apply to them. Consider, for example, the inference (Cherubini & Johnson-Laird, 2004):

> There are four persons: Ann, Bill, Cath, and Dave.
>
> Everybody loves anyone who loves someone.
>
> Ann loves Bill.
>
> What follows?

Most people can envisage this model in which arrows denote the relation of *loving*:



Hence, they infer that everyone loves Ann. However, if you ask them whether it follows that Cath loves Dave, they tend to respond "no." They are mistaken, but the inference calls for using the quantified premise again. The result is this model (strictly speaking, all four persons love themselves, too):



It follows that Cath loves Dave, and people grasp its validity if it is demonstrated with diagrams. No complete model theory exists for inferences based on quantifiers and connectives (cf. Bara, Bucciarelli, & Lombardo, 2001). However, the main principles of the theory should apply: iconicity, possibilities, truth, and strategic variation.

Experimental Studies of Deductive Reasoning

Many experiments have corroborated the model theory (for a bibliography, see the Web page created by Ruth Byrne: www.tcd.ie/Psychology/People/Ruth_Byrnelmental_

models/). This section outlines the corroborations of five predictions.

Prediction 1: The fewer the models needed for an inference, and the simpler they are, the less time the inference should take and the less prone it should be to error. Fewer entities do improve inferences (e.g., Birney & Halford, 2002). Likewise, fewer models improve spatial and temporal reasoning (Byrne & Johnson-Laird, 1989; Carreiras & Santamaría, 1997; Schaeken, Johnson-Laird, & d'Ydewalle, 1996; Vandierendonck & De Vooght, 1997). Premises yielding one model take less time to read than corresponding premises yielding multiple models; however, the difference between two and three models is often so small that it is unlikely that reasoners construct all three models (Vandierendonck, De Vooght, Desimpelaere, & Dierckx, 2000). They may build a single model with one element represented as having two or more possible locations.

Effects of number of models have been observed in comparing one sort of sentential connective with another and in examining batteries of such inferences (see Johnson-Laird & Byrne, 1991). To illustrate these effects, consider the "double disjunction" (Bauer & Johnson-Laird, 1993):

Ann is in Alaska or else Beth is in Barbados, but not both.
Beth is in Barbados or else Cath is in Canada, but not both.
What follows?

Reasoners readily envisage the two possibilities compatible with the first premise, but it is harder to update them with those from the second premise. The solution is

Ann in Alaska Cath in Canada
Beth in Barbados

People represent the spatial relations: Models are not made of words. The two models yield the conclusion: *Either Ann is in Alaska and Cath is in Canada or else Beth is in Barbados.* An increase in complexity soon overloads working memory. This problem defeats most people:

Ann is in Alaska or Beth is in Barbados, or both.
Beth is in Barbados or Cath is in Canada, or both.
What follows?

The premises yield five models, from which it follows: *Ann is in Alaska and Cath is in Canada, or Beth is in Barbados, or all three.* When the order of the premises reduces the number of models to be held in mind, reasoning improves (García-Madruga, Moreno, Carriedo, Gutiérrez, & Johnson-Laird, 2001; Girotto, Mazzocco, & Tasso, 1997; Mackiewicz & Johnson-Laird, 2003).

Because one model is easier than many, an interaction occurs in *modal* reasoning. It is easier to infer that a situation is possible (one model of the premises suffices as an example) than that it is *not* possible (all the models of the premises must be checked for a counterexample to the conclusion). In contrast, it is easier to infer that a situation is *not* necessary (one counterexample suffices) than that it is necessary (all the models of the premises must be checked as examples). The interaction occurs in both accuracy and speed (Bell & Johnson-Laird, 1998; see also Evans et al., 1999).

Prediction 2: Reasoners should err as a result of overlooking models of the premises. Given a double disjunction (such as the previous one), the most frequent errors were conclusions consistent with just a single model of the premises (Bauer & Johnson-Laird, 1993). Likewise, given a syllogism of the form,

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

reasoners infer: *None of the A is a* C (Newstead & Griggs, 1999). They overlook the possibility in which Cs that are not *B*s are *A*s, and so the valid conclusion is

Some of the C *are not A.*

They may have misinterpreted the second premise, taking it also to mean that *all*

the C are B (Newstead & Griggs, 1999), but many errors with syllogisms appear to arise because individuals consider only a single model (Bucciarelli & Johnson-Laird, 1999; Espino, Santamaría, & García-Madruga, 2000). Ormerod proposed a "minimal completion" hypothesis according to which reasoners construct only the minimally necessary models (see Ormerod, Manktelow, & Jones, 1993; Richardson & Ormerod, 1997). Likewise, Sloutsky postulated a process of "minimalization" in which reasoners tend to construct only single models for all connectives, thereby reducing them to conjunctions (Morris & Sloutsky, 2002; Sloutsky & Goldvarg, 1999). Certain assertions, however, do tend to elicit more than one model. As Byrne and her colleagues showed (e.g., Byrne, 2002; Byrne & McEleney, 2000; Byrne & Tasso, 1999), counterfactual conditionals such as

If the cable hadn't been faulty then the printer wouldn't have broken

tend to elicit models of both what is factually the case, that is,

cable faulty printer broken

and what holds in a counterfactual possibility

¬ cable faulty ¬ printer broken

Prediction 3: Reasoners should be able to refute invalid inferences by envisaging counterexamples (i.e., models of the premises that refute the putative conclusion). There is no guarantee that reasoners will find a counterexample, but, where they do succeed, they *know* that an inference is invalid (Barwise, 1993). The availability of a counterexample can suppress fallacious inferences from a conditional premise (Byrne, Espino, & Santamaría, 1999; Markovits, 1984; Vadeboncoeur & Markovits, 1999). Nevertheless, an alternative theory based on mental models has downplayed the role of counterexamples (Polk & Newell, 1995), and reasoners' diagrams have sometimes failed to show their use (e.g., Newstead, Handley, & Buck, 1999). However, when reasoners had to construct external models (Bucciarelli & Johnson-Laird, 1999), they used counterexamples (see also Neth & Johnson-Laird, 1999; Roberts, in press).

There are two sorts of invalid conclusions. One sort is invalid because the conclusion is disjoint with the premises; for example,

A or B or both.
B or else C but not both.
Therefore, not-A and C.

The premises have three fully explicit models:

A	¬ B	C
¬ A	B	¬ C
A	B	¬ C

The conclusion is inconsistent with the premises because it conflicts with each of their models. But, another sort of invalid conclusion is consistent with the premises but does not follow from them such as the conclusion *A and not*-C from the previous premises. It is consistent with the premises because it corresponds to their third model, but it does not follow from them because the other two models are counterexamples. Reasoners usually establish the invalidity of the first sort of conclusion by detecting its inconsistency with the premises, but they refute the second sort of conclusion with a counterexample (Johnson-Laird & Hasson, 2003). An experiment using functional magnetic resonance imaging showed that reasoning based on numeric quantifiers, such as *at least five* – as opposed to arithmetical calculation based on the same premises – depended on the right frontal hemisphere. A search for counterexamples appeared to activate the right frontal pole (Kroger, Cohen, & Johnson-Laird, 2003).

Prediction 4: Reasoners should succumb to *illusory* inferences, which are compelling but invalid. They arise from the principle of

truth and its corollary that reasoners neglect what is false. Consider the problem:

Only one of the following assertions 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 ten, or both.

Is it possible that there is an ace in the hand?

Nearly everyone responds, "yes" (Goldvarg & Johnson-Laird, 2000). They grasp that the first assertion allows two possibilities in which an ace occurs, so they infer that an ace is possible. However, it is impossible for an ace to be in the hand because both of the first two assertions would then be true, contrary to the rubric that only one of them is true. The inference is an illusion of possibility: Reasoners infer wrongly that a card is possible. A similar problem to which reasoners tend to respond "no" and thereby commit an illusion of impossibility is created by replacing the two occurrences of "there is an ace" in the problem with, "there is not an ace." When the previous premises were stated with the question

Is it possible that there is a jack?

the participants nearly all responded "yes," again. They considered the third assertion, and its mental models showed that there could be a jack. However, this time they were correct: The inference is valid. Hence, the focus on truth does not always lead to error, and experiments have accordingly compared illusions with matching control problems for which the neglect of falsity should not affect accuracy.

The computer program implementing the theory shows that illusory inferences should be sparse in the set of all possible inferences. However, experiments have corroborated their occurrence in reasoning about possibilities, probabilities, and causal and deontic relations. Table 9.4 illustrates some different illusions. Studies have used remedial procedures to reduce the illusions (e.g., Santamaría & Johnson-Laird, 2000). Yang taught participants to think explicitly about what is true and what is false. The difference between illusions and control problems vanished, but performance on the control problems fell from almost 100% correct to around 75% correct (Yang & Johnson-Laird, 2000). The principle of truth limits understanding, but it does so without participants realizing it. They were highly confident in their responses, no less so when they succumbed to an illusion than when they responded correctly to a control problem.

The rubric, "one of these assertions is true and one of them is false," is equivalent to an exclusive disjunction between two assertions: *A or else B, but not both.* This usage leads to compelling illusions that seduce novices and experts alike, for example,

If there is a king then there is an ace, or else if there isn't a king then there is an ace.

There is a king.

What follows?

More than 2000 individuals have tackled this problem (see Johnson-Laird & Savary, 1999), and nearly everyone responded, "there is an ace." The prediction of an illusion depends not on logic but on how other participants interpreted the relevant connectives in simple assertions. The preceding illusion occurs with the rubric: *One of these assertions is true and one of them is false* applying to the conditionals. That the conclusion is illusory rests on the following assumption, corroborated experimentally: If a conditional is false, then one possibility is that its antecedent is true and its consequent is false. If skeptics think that the illusory responses are correct, then how do they explain the effects of a remedial procedure? They should then say that the remedy produced illusions. Readers may suspect that the illusions arise from the artificiality of the problems, which

Table 9.4. Some illusory inferences in abbreviated form, with percentages of illusory responses. Each study examined other sorts of illusions and matched control problems

Premises	*Illusory responses*	*Percentages of illusory responses*
1. If A then B or else B. A.	B.	100
2. Either A and B, or else C and D. A.	B.	87
3. If A then B or else if C then B. A and B.	Possibly both are true.	98
4. A or else not both B and C. A and not B.	Possibly both are true.	91
5. One true and one false: not-A or not-B, or neither. Not-C and not-B.	Possibly not-C and not-B.	85
6. Only one is true: At least some A are not B. No A are B.	Possibly No B are A.	95
7. If one is true so is the other: A or else not B. A.	A is more likely than B.	95
8. If one is true so is the other: A if and only if B. A.	A is equally likely as B.	90

Note: 1 is from Johnson-Laird and Savary (1999), 2 is from Walsh and Johnson-Laird (2003), 3 is from Johnson-Laird, Legrenzi, Girotto, and Legrenzi (2000), 4 is from Legrenzi, Girotto, and Johnson-Laird (2003), 5 is from Goldvarg and Johnson-Laird (2000), 6 is from Experiment 2, Yang and Johnson-Laird (2000), and 7 and 8 are from Johnson-Laird and Savary (1996).

never occur in real life and therefore confuse the participants. The problems may be artificial, although analogs do occur in real life (see Johnson-Laird & Savary, 1999), and artificiality fails to explain the correct responses to the controls or the high ratings of confidence in both illusory and control conclusions.

Prediction 5: Naive individuals should develop different reasoning strategies based on models. When they are tested in the laboratory, they start with only rough ideas of how to proceed. They can reason, but not efficiently. With experience but no feedback about accuracy, they spontaneously develop various strategies (Schaeken, De Vooght, Vandierendonck, & d'Ydewalle, 1999). Deduction itself may be a strategy (Evans, 2000), and people may resort to it more in Western cultures than in East Asian cultures (Peng & Nisbett, 1999). However, deduction itself leads to different strategies (Van der Henst, Yang, & Johnson-Laird, 2002). Consider a problem in which each premise is *compound*, that is, contains a connective:

A if and only if B.
Either B or else C, but not both.
C if and only if D.
Does it follow that if not A then D?

where A, B, ... refer to different colored marbles in a box. Some individuals develop a strategy based on suppositions. They say, for example,

> Suppose *not A. It follows from the first premise that not B. It follows from the second premise that C. The third premise then implies D. So, yes, the conclusion follows.*

Some individuals construct a chain of conditionals leading from one clause in the conclusion to the other – for example: *If D then C, If C then not B, If not B then not A*. Others develop a strategy in which they enumerate the different possibilities compatible with the premises. For example, they draw a horizontal line across the page and write down the possibilities for the premises:

A	B
	C
	D

When individuals are taught to use this strategy, as Victoria Bell showed in unpublished studies, their reasoning is faster and more accurate. The nature of the premises and the conclusion can bias reasoners to adopt a predictable strategy (e.g., conditional premises encourage the use of suppositions, whereas disjunctive premises

encourage the enumeration of possibilities) (Van der Henst et al., 2002).

Reasoners develop diverse strategies for relational reasoning (e.g., Goodwin & Johnson-Laird, in press; Roberts, 2000), suppositional reasoning (e.g., Byrne & Handley, 1997), and reasoning with quantifiers (e.g., Bucciarelli & Johnson-Laird, 1999). Granted the variety of strategies, there remains a robust effect: Inferences from one mental model are easier than those from more than one model (see also Espino, Santamaría, Meseguer, & Carreiras, 2000). Different strategies could reflect different mental representations (Stenning & Yule, 1997), but those so far discovered are all compatible with models. Individuals who have mastered logic could make a strategic use of formal rules. Given sufficient experience with a class of problems, individuals begin to notice some formal patterns.

Probabilistic Reasoning

Reasoning about probabilities is of two sorts. In *intensional* reasoning, individuals use heuristics to infer the probability of an event from some sort of index, such as the availability of information. In *extensional* reasoning, they infer the probability of an event from a knowledge of the different ways in which it might occur. This distinction is due to Nobel laureate Daniel Kahneman and the late Amos Tversky, who together pioneered the investigation of heuristics (Kahneman, Slovic, & Tversky, 1982; see Kahneman & Frederick, Chap. 12). Studies of extensional reasoning focused at first on "Bayesian" reasoning in which participants try to infer a conditional probability from the premises. These studies offered no account of the foundations of extensional reasoning. The model theory filled the gap (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999), and the present section outlines its account.

Mental models represent the extensions of assertions (i.e., the possibilities to which they refer). The theory postulates

> *The principle of* equiprobability: *Each mental model is assumed to be equiprobable, unless there are reasons to the contrary.*

The probability of an event accordingly depends on the proportion of models in which it occurs. The theory also allows that models can be tagged with numerals denoting probabilities or frequencies of occurrence, and that simple arithmetical operations can be carried out on them. Shimojo and Ichikawa (1989) and Falk (1992) proposed similar principles for Bayesian reasoning. The present account differs from theirs in that it assigns equiprobability, not to actual events, but to mental models. And equiprobability applies only by default. An analogous principle of "indifference" occurred in classical probability theory, but it is problematic because it applies to events (Hacking, 1975).

Consider a simple problem such as

> In the box, there is a green ball or a blue ball or both.
>
> What is the probability that both the green and the blue ball are there?

The premise elicits the mental models:

green	
	blue
green	blue

Naive reasoners follow the equiprobability principle, and infer the answer, "1/3." An experiment corroborated this and other predictions based on the mental models for the connectives in Table 9.2 (Johnson-Laird et al., 1999).

Conditional probabilities are on the borderline of naive competence. They are difficult because individuals need to consider several fully explicit models. Here is a typical Bayesian problem:

> *The patient's PSA score is high. If he doesn't have prostate cancer, the chances of such a value is 1 in 1000. Is he likely to have prostate cancer?*

Many people respond, "yes." However, they are wrong. The model theory predicts the error: Individuals represent the conditional

probability in the problem as one explicit model and one implicit model tagged with their chances:

¬ prostate cancer	high PSA	1
. . .		999

The converse conditional probability has the same mental models, and so people assume that if the patient has a high PSA the chances are only 1 in 1000 that he does not have prostate cancer. Because the patient has a high PSA, then he is highly likely to have prostate cancer (999/1000). To reason correctly, individuals must envisage the complete partition of possibilities and chances. However, the problem fails to provide enough information. It yields only:

¬ prostate cancer	high PSA	1
¬ prostate cancer	¬ high PSA	999
prostate cancer	high PSA	?
prostate cancer	¬ high PSA	?

There are various ways to provide the missing information. One way is to give the base rate of prostate cancer, which can be used with Bayes's theorem from the probability calculus to infer the answer. However, the theorem and its computations are beyond naive individuals (Kahneman & Tversky, 1973; Phillips & Edwards, 1966). The model theory postulates an alternative:

> *The* subset *principle: Given a complete partition, individuals infer the conditional probability, P(A | B), by examining the subset of B that is A and computing its proportion (Johnson-Laird* et al., *1999).*

If models are tagged with their absolute frequencies or chances, then the conditional probability equals their value for the model of *A and B* divided by their sum for all the models containing *B*. A complete partition for the patient problem might be

¬ prostate cancer	high PSA	1
¬ prostate cancer	¬ high PSA	999
prostate cancer	high PSA	2
prostate cancer	¬ high PSA	0

The subset of chances of prostate cancer within the two possibilities of a high PSA (rows 1 and 3) yields the conditional probability: P(prostate cancer | high PSA) = 2/3. It is high, but far from 999/1000.

Evolutionary psychologists postulate that natural selection led to an innate "module" in the mind that makes Bayesian inferences from naturally occurring frequencies. It follows that naive reasoners should fail the patient problem because it is about a unique event (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). In contrast, as the model theory predicts, individuals cope with problems about unique or repeated events provided they can use the subset principle and the arithmetic is easy (Girotto & Gonzalez, 2001).

The model theory dispels some common misconceptions about probabilistic reasoning. It is *not* always inductive. Extensional reasoning can be deductively valid, and it need not depend on a tacit knowledge of the probability calculus. It is not always correct because it can yield illusions (Table 9.4).

Induction and Models

Induction is part of everyday thinking (see Sloman & Lagnado, Chap. 5). Popper (1972) argued, however, that it is not part of scientific thinking. He claimed that science is based on explanatory conjectures, which observations serve only to falsify. Some scientists agree (e.g., Deutsch, 1997, p. 159). However, many astronomical, meteorological, and medical observations are not tests of hypotheses. Everyone makes inductions in daily life. For instance, when the starter will not turn over the engine, your immediate thought is that the battery is dead. You are likely to be right, but there is no guarantee. Likewise, when the car ferry, *Herald of Free Enterprise*, sailed from Zeebrugge on March 6, 1987, its master made the plausible induction that the bow doors had been closed. They had always been closed in the past, and there was no evidence to the contrary. However, they had not been closed,

the vessel capsized and sank, and many people drowned. Induction is a common but risky business.

The textbook definition of induction – alas, all too common – is that it leads from the particular to the general. Such arguments are indeed inductions, but many inductions such as the preceding examples are inferences from the particular to the particular. That is why the "Introduction" offered a more comprehensive definition: Induction is a process that increases semantic information. As an example, consider again the inference:

The starter won't turn.

Therefore, the battery is dead.

Like all inductions, it depends on knowledge and, in particular, on the true conditional:

If the battery is dead, then the starter won't turn.

It is consistent with the possibilities:

battery dead	¬ starter turn
¬ battery dead	¬ starter turn
¬ battery dead	starter turn

The premise of the induction eliminates the third possibility, but the conclusion goes beyond the information given because it eliminates the second of them. The availability of the first model yields an intensional inference of a high probability, but its conclusion rejects a real possibility. Hence, it may be false. Inductions are vulnerable because they increase semantic information.

Inductions depend on knowledge. As Kahneman and Tversky (1982) showed, various heuristics constrain the use of knowledge in inductions. The *availability* heuristic, illustrated in the previous example, relies on whatever relevant knowledge is available (e.g., Tversky & Kahneman, 1973). The *representativeness* heuristic yields inferences dependent on the representative nature of the evidence (e.g., Kahneman & Frederick, 2002; also see Kahneman & Frederick, Chap. 12). The present account presupposes these heuristics but examines the role of models in induction. Some inductions are *implicit:* They are rapid, involuntary, and unconscious (see Litman & Reber, Chap. 18). Other inductions are *explicit:* They are slow, voluntary, and conscious. This distinction is familiar (e.g., Evans & Over, 1996; Johnson-Laird & Wason, 1977, p. 341; Sloman, 1996; Stanovich, 1999). The next part considers implicit inductions, and the part thereafter considers explicit inductions and the resolution of inconsistencies.

Implicit Induction and the Modulation of Models

Semantics is central to models, and the content of assertions and general knowledge can modulate models. Psychologists have proposed many theories about the mental representation of knowledge, but knowledge is about what is possible, and so the model theory postulates that it is represented in fully explicit models (Johnson-Laird & Byrne, 2002). These models, in turn, modulate the mental models of assertions according to

> *The principle of* modulation: *The meanings of clauses, coreferential links between them, general knowledge, and knowledge of context, can modulate the models of an assertion. In the case of inconsistency, meaning and knowledge normally take precedence over the models of assertions.*

Modulation can add information to mental models, prevent their construction, and flesh them out into fully explicit models. As an illustration of semantic modulation, consider the following conditional:

If it's a game, then it's not soccer.

Its fully explicit models (Table 9.2), if they were unconstrained by coreference and semantics, would be

game	¬ soccer
¬ game	¬ soccer
¬ game	soccer

The meaning of the noun *soccer* entails that it is a game, and so an attempt to construct

the third model fails because it would yield an inconsistency. The conditional has only the first two models.

The pragmatic effects of knowledge have been modeled in a computer program, which can be illustrated using the example

If the match is struck properly, then it lights.

The match is soaking wet and it is struck properly.

What happens?

In logic, it follows that the match lights, but neither people nor the program draws this conclusion. Knowledge that wet matches do not light overrides the model of the premises. The program constructs the mental model of the premises:

match wet	match struck	match lights [the model of the premises]

If a match is soaking wet, it does not light, and the program has a knowledge base containing this information in fully explicit models:

match wet	¬ match lights
¬ match wet	¬ match lights
¬ match wet	match lights

The second premise states that the match is wet, which triggers the matching possibility in the preceding models:

match wet	¬ match lights

The conjunction of this model with the model of the premises would yield a contradiction, but the program follows the principle of modulation and gives precedence to knowledge yielding the following model:

match wet	match struck	¬ match lights

and so the match does not light. The model of the premises also triggers another possibility from the knowledge base:

¬ match wet	match lights

This possibility and the model of the premises are used to construct a counterfactual conditional:

If it had not been the case that match wet *and given* match struck, *then it might have been the case that* match lights.

Modulation is rapid and automatic, and it affects comprehension and reasoning (Johnson-Laird & Byrne, 2002; Newstead, Ellis, Evans, & Dennis, 1997; Ormerod & Johnson-Laird, in press). In logic, connectives such as conditionals and disjunctions are *truth functional*, and so the truth value of a sentence in which they occur can be determined solely from a knowledge of the truth values of the clauses they interconnect. However, in natural language, connectives are not truth functional: It is always necessary to check whether their content and context modulate their interpretation.

Explicit Induction, Abduction, and the Creation of Explanations

Induction is the use of knowledge to increase semantic information: Possibilities are eliminated either by adding elements to a mental model or by eliminating a mental model altogether. After you have stood in line to no avail at a bar in Italy, you are likely to make an explicit induction:

In Italian bars with cashiers, you pay the cashier first and then take your receipt to the bar to make your order.

This induction is a general description. You may also formulate an explanation:

The barmen are too busy to make change, and so it is more efficient for customers to pay a cashier.

Scientific laws are general descriptions of phenomena (e.g., Kepler's third law describes the elliptical orbits of the planets). Scientific theories explain these regularities in terms of more fundamental considerations (e.g., the general theory of relativity explains planetary orbits as the result of the sun's mass curving space-time). Peirce (1903)

called thinking that leads to explanations *abduction*. In terms of the five categories of the "Introduction," abduction is creative when it leads to the revision of beliefs.

Consider the following problem:

If a pilot falls from a plane without a parachute, the pilot dies. This pilot did not die, however. Why not?

Most people respond, for example, that

The plane was on the ground.
The pilot fell into a deep snow drift.

Only a minority draws the logically valid conclusion:

The pilot did not fall from the plane without a parachute.

Hence, people prefer a causal explanation repudiating the first premise to a valid deduction, albeit they may presuppose that the antecedent of the conditional is true. Granted that knowledge usually takes precedence over contradictory assertions, the explanatory mechanism should dominate the ability to make deductions.

In daily life, the propensity to explain is extraordinary, as Tony Anderson and this author discovered when they asked participants to explain the inexplicable. The participants received pairs of sentences selected *at random* from separate stories:

John made his way to a shop that sold TV sets.
Celia had recently had her ears pierced.

In another condition, the sentences were modified to make them coreferential:

Celia made her way to a shop that sold TV sets.
She had recently had her ears pierced.

The participants' task was to explain what was going on. They readily went beyond the given information to account for what was happening. They proposed, for example, that Celia was getting reception in her earrings and wanted the TV shop to investigate, that she wanted to see some new earrings on closed circuit TV, that she had won a bet by having her ears pierced and was spending the money on a TV set, and so on. Only rarely were the participants stumped for an explanation. They were almost as equally ingenious with the sentences that were not coreferential.

Abduction depends on knowledge, especially of causal relations, which according to the model theory refer to temporally ordered sets of possibilities (Goldvarg & Johnson-Laird, 2001; see Cheng & Buehner, Chapter 5.). An assertion of the form C *causes E* is compatible with three fully explicit possibilities:

C	E
¬ C	E
¬ C	¬ E

with the temporal constraint that *E* cannot precede C. An "enabling" assertion of the form C *allows E* is compatible with the three possibilities:

C	E
C	¬ E
¬ C	¬ E

This account, unlike others, accordingly distinguishes between the meaning and logical consequences of causes and enabling conditions (pace, e.g., Einhorn & Hogarth, 1978; Hart & Honoré, 1985; Mill, 1874). It also treats causal relations as determinate rather than probabilistic (pace, e.g., Cheng, 1997; Suppes, 1970). Experiments support both these claims: Participants listed the previous possibilities, and they rejected other cases as impossible, contrary to probabilistic accounts (Goldvarg & Johnson-Laird, 2001). Of course, when individuals induce a causal relation from a series of observations, they are influenced by relative frequencies. However, on the present account, the meaning of any causal relation that they induce is deterministic.

Given the cause from a causal relation, there is only one possible effect, as the previous models show; however, given the effect, there is more than one possible cause. Exceptions do occur (Cummins, Lubart, Alksnis, & Rist, 1991; Markovits, 1984),

but the principle holds in general. It may explain why inferences from causes to effects are more plausible than inferences from effects to causes. As Tversky and Kahneman (1982) showed, conditionals in which the antecedent is a cause such as

A girl has blue eyes if her mother has blue eyes.

are judged as more probable than conditionals in which the antecedent is an effect:

The mother has blue eyes if her daughter has blue eyes.

According to the model theory, when individuals discover inconsistencies, they try to construct a model of a cause and effect that resolves the inconsistency. It makes possible the facts of the matter, and the belief that the causal assertion repudiates is taken to be a counterfactual possibility (in a comparable way to the modulation of models by knowledge). Consider, for example, the scenario:

If the trigger is pulled then the pistol will fire. The trigger is pulled, but the pistol does not fire. Why not?

Given 20 different scenarios of this form (in an unpublished study carried out by Girotto, Legrenzi, & Johnson-Laird), most explanations were causal claims that repudiated the conditional. In two further experiments with the scenarios, the participants rated the statements of a cause and its effect as the most probable explanations; for example,

A prudent person had unloaded the pistol and there were no bullets in the chamber.

The cause alone was rated as less probable, but as more probable than the effect alone, which in turn was rated as more probable than an explanation that repudiated the categorical premise; for example,

The trigger wasn't really pulled.

The greater probability assigned to the conjunction of the cause and effect than to either of its clauses is an instance of the "conjunction" fallacy in which a conjunction is in error judged to be more probable than its constituents (Tversky & Kahneman, 1983).

Abductions that resolve inconsistencies have been implemented in a computer program that uses a knowledge base to create causal explanations. Given the preceding example, the program constructs the mental models of the conditional:

trigger pulled	pistol fires
. . .	

The conjunction of the categorical assertion yields

trigger pulled	pistol fires	[the model of the premises]

That the pistol did not fire is inconsistent with this model. The theory predicts that individuals should tend to abandon their belief in the conditional premise because its one explicit mental model conflicts with the fact that the pistol did not fire (see Girotto, Johnson-Laird, Legrenzi, & Sonino, 2000, for corroborating evidence). Nevertheless, the conditional expresses a useful idealization, and so the program treats it as the basis for a counterfactual set of possibilities:

trigger pulled	¬pistol fires	[the model of the facts]
trigger pulled	pistol fires	[the models of counterfactual possibilities]
. . .		

People know that a pistol without bullets does not fire, and so the program has in its knowledge base the models:

¬ bullets in pistol	¬ pistol fires
bullets in pistol	¬ pistol fires
bullets in pistol	pistol fires

The model of the facts triggers the first possibility in this set, which modulates the model of the facts to create a possibility:

¬ bullets in pistol	trigger pulled	¬ pistol fires

The new proposition in this model triggers a causal antecedent from another set of models in the knowledge base, which explains the inconsistency: A person emptied the pistol and so it had no bullets. The counterfactual possibilities yield the claim: If the person had not emptied the pistol, then it would have had bullets, and . . . it would have fired. The fact that the pistol did not fire has been used to reject the conditional premise, and available knowledge has been used to create an explanation and to modulate the conditional premise into a counterfactual. There are, of course, other possible explanations.

In sum, reasoners can resolve inconsistencies between incontrovertible evidence and the consequences of their beliefs. They use their available knowledge – in the form of explicit models – to try to create a causal scenario that makes sense of the facts. Their reasoning may resolve the inconsistency, create an erroneous account, or fail to yield any explanation whatsoever.

Conclusions and Further Directions

Mental models have a past in the nineteenth century. The present theory was developed in the twentieth century. In its application to deduction, as Peirce anticipated, if a conclusion holds in all the models of the premises, it is necessary given the premises. If it holds in a proportion of the models, then, granted that they are equiprobable, its probability is equal to that proportion. If it holds in at least one model, then it is possible. The theory also applies to inductive reasoning – both the rapid implicit inferences that underlie comprehension and the deliberate inferences yielding generalizations. It offers an account of the creation of causal explanations. However, if Craik was right, mental models underlie all thinking with a propositional content, and so the present theory is radically incomplete.

What of the future of mental models? The theory is under intensive development and intensive scrutiny. It has been corroborated in many experiments, and it is empirically distinguishable from other theories. Indeed, there are distinguishable variants of the theory itself (see, e.g., Evans, 1993; Ormerod, Manktelow, & Jones, 1993; Polk & Newell, 1995). The most urgent demands for the twenty-first century are the extension of the theory to problem solving, decision making, and strategic thinking when individuals compete or cooperate.

Acknowledgments

This chapter was made possible by a grant from the National Science Foundation (Grant BCS 0076287) to study strategies in reasoning. The author is grateful to the editor, the community of reasoning researchers, and his colleagues, collaborators, and students – many of their names are found in the "References" section.

References

Bar-Hillel, Y., & Carnap, R. (1964). An outline of a theory of semantic information. In Y. Bar-Hillel (Ed.), *Language and information processing*. Reading, MA: Addison-Wesley.

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

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

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.

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

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

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. (in press). Naïve deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology*.

Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences*, 6, 426–431.

Byrne, R. M. J., Espino, O., & Santamaría, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, 40, 347–373.

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 and Cognition*, 27, 726–740.

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

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367–405.

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

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1–73.

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

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory and Cognition*, 19, 274–282.

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

Deutsch, D. (1997). *The fabric of reality: The science of parallel universes – and its implications*. New York: Penguin Books.

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

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, 85, 395–416.

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: Erlbaum.

Espino, O., Santamaría, C., & García-Madruga, J. A. (2000). Activation of end terms in syllogistic reasoning. *Thinking and Reasoning*, 6, 67–89.

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. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition*, 48, 1–20.

Evans, J. St. B. T. (2000). What could and could not be a strategy in reasoning. In W. S. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies*. (pp. 1–22) Mahwah, NJ: Erlbaum.

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., & Over, D. E. (1996). *Rationality and reasoning*. Hove, East Sussex: Psychology Press.

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition*, 43, 197–223.

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

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: Erlbaum.

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: Erlbaum.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency format. *Psychological Review, 102*, 684–704.

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. González-Labra (Eds.), *Mental models in reasoning* (pp. 83–97). Madrid: Universidad Nacióńal de Educacion a Distanzia.

Girotto, V., Legrenzi, P., & Johnson-Laird, P. N. (Unpublished studies).

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 and 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.

Goodwin, G., & Johnson-Laird, P. N. (in press). Reasoning about the relations between relations.

Hacking, I. (1975). *The emergence of probability*. Cambridge, UK: Cambridge University Press.

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

Hart, H. L. A., & Honoré, A. M. (1985). *Causation in the law* (2nd ed.). Oxford, UK: Clarendon Press. (First edition published in 1959.)

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.). (1985). *Formal theories of the commonsense world*. Hillsdale, NJ: Erlbaum.)

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.

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

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

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

University Press; Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (1993). *Human and machine thinking*. Hillsdale, NJ: Erlbaum.

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

Johnson-Laird, P. N. (in press). The history of mental models. In K. Manktelow (Ed.), *Psychology of reasoning: Theoretical and historical perspectives*. London: Psychology Press.

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: 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., 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 and Cognition*, 31, 1105–1113.

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., & 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. (1996). Illusory inferences about probabilities. *Acta Psychologica*, 93, 69–90.

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

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

Johnson-Laird, P. N., & Wason, P. C. (Eds.). (1977). *Thinking*. Cambridge, UK: Cambridge University Press.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics of intuitive judgment: Extensions and applications*. New York: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.

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.

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 and Cognition*, 30, 363–371.

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

Kroger, J. K., Cohen, J. D., & Johnson-Laird, P. N. (2003). A double dissociation between logic and mathematics. Unpublished MS.

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

Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (2003). Models of consistency. *Psychological Science*, 14, 131–137.

Mackiewicz, R., & Johnson-Laird, P. N. (in press). Deduction, models, and order of premises.

Markovits, H. (1984). Awareness of the "possible" as a mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology*, 75, 367–376.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

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

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.

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. (Originally published in Solso, R. L. (Ed.). (1974). *Theories in cognitive psychology: The Loyola Symposium*. Hillsdale, NJ: Erlbaum).

Mill, J. S. (1874). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific evidence* (8th ed.). New York: Harper. (First edition published 1843.)

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. 223–258). Cambridge, MA: MIT Press.

Morris, B. J., & Sloutsky, V. (2002). Children's solutions of logical versus empirical problems: What's missing and what develops? *Cognitive Development*, 16, 907–928.

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.

Newstead, S. E., Ellis, M. C., Evans, J. St. B. T., & Dennis, I. (1997). Conditional reasoning with realistic material. *Thinking and Reasoning*, 3, 49–76.

Newstead, S. E., & Griggs, R. A. (1999). Premise misinterpretation and syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, 52*A*, 1057–1075.

Newstead, S. E., Handley, S. J., & Buck, E. (1999). Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory and Cognition*, 27, 344–354.

Ormerod, T. C., & Johnson-Laird, P. N. (in press). How pragmatics modulates the meaning of sentential connectives.

Ormerod, T. C., Manktelow, K. I., & Jones, G. V. (1993). Reasoning with three types of conditional: Biases and mental models. *Quarterly Journal of Experimental Psychology*, 46*A*, 653–678.

Osherson, D. N. (1974–1976) *Logical abilities in children, vols. 1–4*. Hillsdale, NJ: Erlbaum.

Peirce, C. S. (1903). Abduction and induction. In J. Buchler (Ed.), *Philosophical writings of Peirce*. New York: Dover, 1955.

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

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, 54, 741–754.

Phillips, L., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72, 346–354.

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

Popper, K. R. (1972). *Objective knowledge*. Oxford, UK: Clarendon.

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.

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

Roberts, M. J. (in press). Falsification and mental models: It depends on the task. In W. Schaeken, A. Vandierendonck, W. Schroyens, & G. d'Ydewalle (Eds.), *The mental models theory of reasoning: Refinement and extensions*. Mahwah, NJ: Erlbaum.

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, CT: JAI Press.

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

Schaeken, W. S., De Vooght, G., Vandierendonck, A., & d'Ydewalle, G. (Eds.). (1999). *Deductive reasoning and strategies*. Mahwah, NJ: Erlbaum.

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

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

Shimojo, S., & Ichikawa, S. (1989). Intuitive reasoning about probability: Theoretical and experimental analyses of the 'problem of three prisoners'. *Cognition*, 32, 1–24.

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.), *Proceedings of the 21 st annual conference of the Cognitive Science Society* (pp. 695–700). Mahwah, NJ: Erlbaum.

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

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

Stenning, K., & Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology*, 34, 109–159.

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.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North-Holland.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.

Tversky, A., & Kahneman, D. (1982). Causal schemas in judgements under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds), *Judgement under uncertainty: Heuristics and biases* (pp. 117–128). Cambridge, Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 292–315.

Vadeboncoeur, I., & Markovits, H. (1999). The effect of instructions and information retrieval on accepting the premises in a conditional reasoning task. *Thinking and Reasoning*, 5, 97–113.

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

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.

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

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

Wason, P. C., & Johnson-Laird, P. N. (1972). *The psychology of reasoning*. 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* (vol. 2). New York: Association for Computing Machinery.

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

Yang, Y., & Johnson-Laird, P. N. (2000). How to eliminate illusions in quantified reasoning. *Memory and Cognition*, 28, 1050–1059.