Psychological Review
2002, Vol. 109, No. 4, 646–678

0033-295X/02/$5.00 DOI: 10.1037//0033-295X.109.4.646

Conditionals: A Theory of Meaning, Pragmatics, and Inference

P. N. Johnson-Laird
Princeton University

Ruth M. J. Byrne
University of Dublin

The authors outline a theory of conditionals of the form *If A then C* and *If A then possibly C.* The 2 sorts of conditional have separate core meanings that refer to sets of possibilities. Knowledge, pragmatics, and semantics can modulate these meanings. Modulation can add information about temporal and other relations between antecedent and consequent. It can also prevent the construction of possibilities to yield 10 distinct sets of possibilities to which conditionals can refer. The mental representation of a conditional normally makes explicit only the possibilities in which its antecedent is true, yielding other possibilities implicitly. Reasoners tend to focus on the explicit possibilities. The theory predicts the major phenomena of understanding and reasoning with conditionals.

You reason about conditional relations because much of your knowledge is conditional. If you get caught speeding, then you pay a fine. If you have an operation, then you need time to recuperate. If you have money in the bank, then you can cash a check. Conditional reasoning is a central part of thinking, yet people do not always reason correctly. The lawyer Jan Schlictmann in a celebrated trial (see Harr, 1995, pp. 361–362) elicited the following information from an expert witness about the source of a chemical pollutant trichloroethylene (TCE):

> If the TCE in the wells had been drawn from out of the river, then there'd be TCE in the riverbed.
> But there isn't any TCE in the riverbed.

Schlictmann then argued that these premises were consistent with the proposition that no contamination came from the river. What he overlooked is that the conclusion is not merely consistent with the premises but follows necessarily from them. Psychologists disagree about the cause of such oversights and about conditional reasoning in general (e.g., Braine & O'Brien, 1998; Cheng & Holyoak, 1985; Johnson-Laird, 1986; Rips, 1983). Philosophers, logicians, and linguists also disagree about the meaning of conditionals (Adams, 1975; Lewis, 1973; Stalnaker, 1968). No consensus exists about *if.*

Our concern is everyday conditionals, including all sentences of the form *If A then C* or *C if A,* where *A* and *C* are declarative clauses. Naive individuals can grasp the meaning of such assertions, and they can use it to reason. The term *naive* refers here merely to people who have not studied logic in any depth. In the 1970s, psychologists assumed that such individuals reason using formal rules of inference like those of a logical calculus. The challenge was to pin down the particular rules (e.g., Braine, 1978; Johnson-Laird, 1975; Osherson, 1974–1976). Such theorizing neglected a discovery made by Wason (1966). Intelligent adults in his *selection* task regularly committed a logical error. He laid out four cards in front of them:

A B 2 3

They knew that each card had a letter on one side and a number on the other side. He showed them a conditional: If a card has the letter *A* on one side then it has the number *2* on the other side. Their task was to select those cards that had to be turned over to discover whether the conditional was true or false about the four cards. Most people selected the *A* and *2* cards, or the *A* card alone. They failed to select the *3* card. However, if the *3* has an *A* on its other side, the conditional is false. Indeed, nearly everyone judges it to be false in this case. Individuals also generate this case when they are asked to make a conditional false (Oaksford & Stenning, 1992), and they judge that the probability of this case is zero given

P. N. Johnson-Laird, Department of Psychology, Princeton University; Ruth M. J. Byrne, Department of Psychology, University of Dublin, Dublin, Ireland.

This research was made possible in part by a grant to P. N. Johnson-Laird from the National Science Foundation (Grant BCS 0076287) to study strategies in reasoning. We thank Jon Baron and Ray Nickerson for their helpful criticisms of an earlier version of this article. We thank Yingrui Yang for providing technical help and advice, Patricia Barres for carrying out Experiment 2, Rachel McCloy for carrying out a study of modus tollens, Dan Zook for writing a preliminary version of the program for pragmatic modulation, and David Over for engaging in many discussions of philosophical theories of conditionals. We are also grateful to many people: Peter Wason, now in retirement, for pioneering the psychological study of conditionals; Jonathan Evans and David Over for a critique of our account, which they presented at the Workshop on Mental Models in Brussels, March 2001, and also for many discussions of conditionals; Luca Bonatti, the late Martin Braine, Keith Holyoak, Howard Margolis, David O'Brien, and Lance Rips for their vigorous criticisms of the model theory; Monica Bucciarelli, Orlando Espino, Vittorio Girotto, Eugenia Goldvarg, Paolo Legrenzi, Maria Sonino Legrenzi, Juan García Madruga, Henry Markovits, Mary Newsome, Cristina Quelhas, Sergio Moreno Rios, Carlos Santamaría, Fabien Savary, Walter Schaeken, Walter Schroyens, Alessandra Tasso, and Valerie Thompson for their collaboration in studies of conditional reasoning. We are also grateful to the following individuals for many useful ideas: Bruno Bara, Victoria Bell, David Green, Uri Hasson, Mark Keane, Markus Knauff, Jim Kroger, Robert Mackiewicz, Hans-Georg Neth, Tom Ormerod, Rosemary Stevenson, Dan Sperber, Vladimir Sloutsky, and Jean-Baptiste van der Henst.

Correspondence concerning this article should be addressed to P. N. Johnson-Laird, Department of Psychology, Princeton University, Green Hall, Princeton, New Jersey 08544. E-mail: phil@princeton.edu

the truth of the conditional (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). When Wason and Shapiro (1971) changed the assertion to a sensible everyday conditional, many people made the correct selection. Theories based on formal rules have difficulty in explaining both the error with neutral conditionals and its correction with sensible content.

Our theory is not formal. It assumes instead that reasoners use the meaning of premises, and general knowledge, to imagine the possibilities under consideration—that is, to construct mental models of them (see Johnson-Laird & Byrne, 1991; Polk & Newell, 1995). Mental models can be constructed from perception, imagination, or the comprehension of discourse. They underlie visual images, but they can also be abstract, representing situations that cannot be visualized. Each mental model represents a possibility. It is akin to a diagram in that its structure is analogous to the structure of the situation that it represents, unlike, say, the structure of logical forms used in formal rule theories.

Some sentential connectives such as *and* and *or* have logical meanings that are easy to define. For example, consider the following:

> Either the printer is broken or else the cable is disconnected, but not both.

The truth of an exclusive disjunction such as this one is a function of the truth of its constituent *atomic* propositions, that is, those propositions that contain neither *not* nor connectives (the printer is broken, the cable is disconnected). The exclusive disjunction is true in two cases: first, where it is true that the printer is broken but false that the cable is disconnected, and, second, where it is false that the printer is broken but true that the cable is disconnected. Logicians lay out these conditions in a *truth table,* as shown in Table 1, in which we abbreviate the atomic propositions. Each row in the truth table represents a different possibility. Thus, the first row represents the possibility in which both the printer is broken and the cable is disconnected; in this possibility, the exclusive disjunction is false. The connective *and* can be defined in an analogous way. A conjunction of the form *A and C* is true in just one row of its truth table, namely, where *A* is true and *C* is true, and it is false in every other case. Those sentential connectives that can be defined in a truth table are known as *truth functional*—that is, their meanings are functions that take the truth values of propositions as inputs and deliver a truth value as an output. The logic of these truth-functional meanings is captured in the *sentential* calculus (see, e.g., Jeffrey, 1981).

Psychological theories based on formal rules (hereinafter *rule* theories) and the theory based on mental models (hereinafter the *model* theory) run in parallel for many aspects of reasoning. Insofar as both sorts of theory reflect logical principles, however, they have potential limitations (cf. Sloman, 1996). And at the heart of these limitations is the vexed question of conditionals. Are they truth functional or not? The answer is not clear. Some theorists have argued that conditionals have the same meaning as the truth-functional connective of material implication (e.g., Grice, 1975; Wilson, 1975), which is shown in Table 2. Other theorists have rejected this analysis (e.g., Gazdar, 1979; O'Brien, 1999). The profundity of the puzzle correlates with the length of its literature. One aim of the present article is to resolve this puzzle.

Table 1
A Truth Table for the Exclusive Disjunction "Either the Printer Is Broken or Else the Cable is Disconnected"

Printer is broken	Cable is disconnected	Either printer is broken or else cable is disconnected
True	True	False
True	False	True
False	True	True
False	False	False

Simple inferences from conditionals are of four main sorts:

1. *Modus ponens:*

If there is a circle then there is a triangle.	If A then C.
There is a circle.	A.
∴ There is a triangle.	∴ C.

2. *Modus tollens:*

If there is a circle then there is a triangle.	If A then C.
There is not a triangle.	Not C.
∴ There is not a circle.	∴ Not A.

3. *Denial of the antecedent:*

If there is a circle then there is a triangle.	If A then C.
There is not a circle.	Not A.
∴ There is not a triangle.	∴ Not C.

4. *Affirmation of the consequent:*

If there is a circle then there is a triangle.	If A then C.
There is a triangle.	C.
∴ There is a circle.	∴ A.

Modus ponens and modus tollens are valid for both conditionals and biconditionals (If and only if there is a circle then there is a triangle), whereas the other two inferences are valid only for biconditionals. There are more experimental studies of these and other inferences with *if* than of any other deductive term (for reviews, see Baron, 1994; J. Evans, Newstead, & Byrne, 1993; Garnham & Oakhill, 1994). Likewise, there are probably more semantic analyses of *if* than of any other single word in English (e.g., Harper, Stalnaker, & Pearce, 1981; Jackson, 1987; Lewis, 1973; Traugott, ter Meulen, Reilly, & Ferguson, 1986). Some authors have despaired of formulating the conditions in which conditionals are true and have argued instead that they have merely conditions in which it is justifiable to assert them (see, e.g., Adams, 1970, 1975; cf. Lewis, 1976). Such skepticism is in striking contrast to the ease with which individuals understand conditionals in daily life and judge them to be true or false.

Our goal is to resolve the problems of conditionals. We therefore formulate a theory of the meaning of conditionals, of how this meaning is modulated by semantics and pragmatics, and of its use in reasoning. Braine and O'Brien (1991) offered such an account based on a rule theory. The present article offers an account based on the model theory. Of course, the model theory has had something to say about conditionals in the past (e.g., Johnson-Laird & Byrne, 1991), but its account was incomplete. Critics have identified two principal lacunae (see, e.g., Fillenbaum, 1993; Holyoak

Table 2
A Truth Table for a Material Implication, "The Printer Works Implies That the Cable Is Connected"

Printer works	Cable is connected	The printer works implies that the cable is connected
True	True	True
True	False	False
False	True	True
False	False	True

& Cheng, 1995). The theory said nothing about the semantic mechanism that allows the meanings of a conditional's constituent clauses to influence its interpretation. Likewise, the theory said nothing about the pragmatic mechanism that allows context and knowledge to influence the interpretation of a conditional. Rule theories do not account for such effects, but as Bonatti (1994a,b) has pointed out, the model theory is in the same difficulty. The new theory fills the lacunae. It makes sense of the everyday comprehension of conditionals and of the phenomena that occur when they are used to make inferences. It also makes some unexpected predictions, which have led to the discovery of new phenomena.

The article begins with a theory of the core meanings of conditionals, both indicative and subjunctive. It then outlines the theory of mental models and extends this theory to the mental representation of basic conditionals. It next formulates a principle of semantic modulation and a principle of pragmatic modulation. These mechanisms modulate the core meanings of conditionals. The article then describes how the new theory accounts for reasoning with conditionals, for the selection task, and for so-called *illusory* inferences based on conditionals. Finally, it draws some general conclusions about conditionals and conditional reasoning.

The Core Meanings of Basic Conditionals

Philosophers and logicians distinguish between the meaning of an expression and its reference. The classic illustration contrasts *the Morning star* and *the Evening star*, which differ in meaning but have the same reference, namely, the planet Venus. We assume that the meaning of a sentence when it is used in a particular context functions to refer to a situation or to a set of situations. Most sentences, however, can be used to express many different meanings, depending on their context. The sentence "We are here," for instance, expresses different meanings depending on who asserts it and when and where they assert it. It is laborious to keep writing, "the meaning of a sentence when it is used in a particular context," and so unless the distinction matters we use *assertion* to refer to sentences and to the meanings that they express.

We develop a model theory of conditionals below, but in our account of their meaning, we refer to *possibilities*, not to *models*. Our aim is to make the theory of meaning independent of the theory of comprehension, which relies on models as mental representations. To frame the theory of meaning, we need to consider the ontology of the situations that occur in everyday life. A major dichotomy is between *factual* claims about situations and *modal* claims about them. A conditional such as "If Pat goes to the party then Viv will go too" makes a factual claim. In the case that Pat goes and Viv does not, the conditional is false. Modal claims are of various sorts, but we focus on the most familiar, which are epistemic claims about what is possible and deontic claims about what is permissible. A conditional such as "If Pat goes to the party then it is permissible for Viv to go" makes a deontic claim. In the case that Pat goes, and Viv does not, the conditional is not necessarily false. It is false only if Viv has no such permission conditional on the truth of the antecedent. There are other sorts of modal claim. For example, the assertion "If you're going to Venice then you ought to see the Doge's palace" expresses a recommendation. Likewise, the assertion "If he is wearing a red tie, then he *must* be wearing a tie" makes a logical claim. To keep matters simple, however, we focus on factual and deontic conditionals.

The ontology of daily life distinguishes between factual possibilities, such as (at the time of writing) "Al Gore is elected President of the United States in 2004," and counterfactual possibilities, such as "Al Gore was elected President of the United States in 2000." A counterfactual possibility refers to a situation that once was a factual possibility but that did not occur. A factual possibility that did occur thereby becomes a fact, that is, an actual state of affairs. The ability to envisage counterfactual possibilities is an important part of how one evaluates what actually happens (Byrne, 1997; Byrne & McEleney, 2000; Byrne, Segura, Culhane, Tasso, & Berrocal, 2000; Kahneman & Miller, 1986). The distinction between factual and counterfactual possibilities is reflected in the English language, and in the construction of conditionals, though their syntax is complex (see Dudman, 1988). Consider the outcome of an election that has been decided but that is not known to the speaker. Reference to a factual possibility calls for the present tense of the modal auxiliary *have:* "If Gore has won the election." However, where the outcome is known, reference to a counterfactual possibility calls for the past tense of the modal auxiliary *have:* "If Gore had won the election."

Assertions can describe a fact, a factual possibility, a counterfactual possibility, and even an impossibility. These four categories also occur in fictional discourse. Thus, given the situation at the end of the first act of Hamlet, there is a factual possibility that Hamlet will kill the king. At the end of the second act, there is a counterfactual possibility that Hamlet had killed the king while the latter was at prayer. Systems of modal logic do not make these distinctions and are accordingly difficult to apply to the analysis of natural language (Karttunen, 1972). However, granted the existence of the four sorts of situation, we can formulate a semantic theory of conditionals.

The Core Semantics

We begin our analysis with *basic* conditionals, which are those with a neutral content that is as independent as possible from context and background knowledge, and which have an antecedent and consequent that are semantically independent apart from their occurrence in the same conditional. We consider two sorts of basic conditional, those of the form *If A then C* and those of the form *If A then possibly C*, and we argue that each of them has a core meaning.

The antecedent *A* of a basic conditional is a description—almost always a partial description—of a possibility. The consequent *C* can serve any illocutionary role and is interpreted as though it were

an isolated main clause in a context that satisfies the antecedent. It follows that the antecedent of a conditional must be a declarative clause in order to describe a situation. In contrast, the consequent of a conditional can have any illocutionary force. It can make an assertion, ask a question, give a command, or make a request. Thus, all of the following are acceptable conditionals:

If Vivien leaves, then you will leave.
If Vivien leaves, then are you allowed to leave?
If Vivien leaves, then please leave.

However, conditionals with antecedents that ask questions, make requests, or perform any other illocution apart from assertion are ungrammatical, for example, "If *are* you allowed to leave, then Vivien will leave." One minor exception occurs: An antecedent can form a *wh* question provided that it echoes the form of a relevant precursor, for example:

If Vivien arrived then Evelyn left.
Evelyn left if *who* arrived?
A basic conditional, *If A then . . .* , asserts that its antecedent is a possibility:

Factual possibilities:	a
	¬ a

a denotes the possibility satisfying the antecedent, *A;* ¬ denotes negation; and so ¬ *a* denotes a possibility that satisfies the negation of the antecedent, *not A.* Here and throughout the article, we use lowercase letters to represent possibilities corresponding to atomic propositions. A conditional can be qualified to assert that the antecedent is definitely the case: "If there is a circle, which there *is,* then . . ." The antecedent and its qualification refer to the following where ○ denotes the occurrence of a circle:

Fact:	○
Counterfactual possibility:	¬ ○

A conditional can also be qualified to rule out the possibility of the antecedent, but in this case the antecedent must be couched in the subjunctive in order to convey that it is a counterfactual possibility: "If there had been a circle, which there wasn't, then . . ." The antecedent and its qualification now refer to the following:

Counterfactual possibility:	○
Fact:	¬ ○

In sum, the antecedent of a conditional establishes two possibilities, either two factual possibilities or a fact and a counterfactual possibility.

The antecedent refers to a possibility, and the consequent is interpreted in that context. The principle is corroborated by the existence of conditionals with no antecedents. They occur when the situation makes a possibility so obvious that it is unnecessary to describe it. For example, a spouse observing some incipient misbehavior can assert, "I'll divorce you," where the force of the assertion is "If you do that, I'll divorce you." Conditionals with no consequents also occur when the context makes the outcome so obvious that it is unnecessary to describe it—for example, "If you do that again . . ." Hereinafter, we focus on complete conditionals in which both antecedent and consequent make assertions.

The antecedent of a basic conditional describes a possibility almost completely, that is, it describes all that one needs to know in order to interpret the conditional. Thus, a typical basic conditional is "If the weather is fine, then the sun shines." It asserts that the sun is bound to shine in fine weather. The consequent may refer explicitly to a necessity: "If the weather is fine, then the sun must shine." The set of possibilities is exactly the same as the preceding factual assertion. Unlike modal logic, the force of *must,* if anything, is weaker than a factual claim. It signals an inference or expectation rather than a matter of fact. In contrast, the consequent can assert merely a possibility given the antecedent: "If the weather is fine then the sun may shine." Hence, there are two sorts of basic conditional. In the case of *If A then C,* the consequent has to occur given the antecedent:

Possibilities:	a	c

In the case of *If A then possibly C,* the consequent may occur given the antecedent:

Possibilities:	a	c
	a	¬ c

What about the possibilities in which the antecedent of a conditional is not satisfied? Our account does not constrain them. Hence, each of the two sorts of basic conditional has its own core meaning. The core meaning of *If A then C* refers to the following:

Factual possibilities:	a	c
	¬ a	c
	¬ a	¬ c

We call this interpretation the *conditional* meaning. Its three possibilities are described completely by the following conjunction of conditionals: *If A then C, and if not A then either C or not C.* Indeed, the conditional meaning applied to this conjunction of conditionals also yields the preceding set of possibilities. Necessity and sufficiency are nothing more than these possibilities in which *A* is sufficient (but not necessary) for *C* and *C* is necessary (but not sufficient) for *A.* A basic deontic conditional of the form *If A then C is obligatory* has an analogous interpretation:

Factual possibilities:	a	c	:Deontic possibilities
	¬ a	c	
	¬ a	¬ c	

Here, the antecedent refers to factual possibilities and the consequent refers to deontic possibilities, that is, what is permissible in each possibility.

The core meaning of *If A then possibly C* refers to the following:

Factual possibilities:	a	c
	a	¬ c
	¬ a	c
	¬ a	¬ c

We call this interpretation the *tautological* meaning, because it allows any possibility, and so the conditional cannot be false. A deontic conditional has the same interpretation except that the consequent refers to deontic possibilities. The conditional can be paraphrased as *If A then possibly C, and if not A then possibly C.*

As shown below, the pragmatics of the situation often rules out one of the possibilities, particularly the possibility in which *C* occurs in the absence of *A*.

We integrate these assumptions in the first principle of the five that will make up our theory of conditionals:

> *Principle 1.* The principle of *core meanings:* The antecedent of a basic conditional describes a possibility, at least in part, and the consequent can occur in this possibility. Each of the two sorts of basic conditional accordingly has a core meaning referring to a set of factual or deontic possibilities. The core meaning of *If A then C* is the *conditional* interpretation, which refers to the possibilities:
>
> a c
> ¬ a c
> ¬ a ¬ c
>
> The core meaning of *If A then possibly C* is the tautological interpretation, which refers to the possibilities:
>
> a c
> a ¬ c
> ¬ a c
> ¬ a ¬ c

The conjunction of two basic conditionals *If A then C, and if not A then not C* can be interpreted as a conjunction of two *conditional* meanings, which yield the following:

Factual possibilities: a c
¬ a ¬ c

In other words, *A* is both necessary and sufficient for *C*. The following conjunction also yields the same set: *If A then C, and if C then A.* The set corresponds to the *biconditional* interpretation expressed by *If, and only if, A then C.*

Alternative Theories

One corroboration of the present semantics comes from a comparison with an alternative theory (formulated by Quine, 1952; defended by Wason & Johnson-Laird, 1972, p. 90; and maintained by K. Holyoak, personal communication, 1999, and O'Brien, 1999). According to this alternative, conditionals have a "defective" truth table with no truth value when their antecedents are false, that is, they are partial truth functions with no result in this case. The account is plausible at first sight but founders on the case of biconditionals. As we saw earlier, the biconditional *If, and only if, A then C* is synonymous with the conjunction *If A then C, and if not A then not C.* With a defective truth table, the first of these conditionals has no truth value when *A* is false, and the second of them has no truth value when *A* is true. However, the biconditional in daily life is true when both *A* and *C* are true and when neither is true and is false in any other case. This complete truth table for the biconditional cannot be equivalent to a conjunction in which there is always one conjunct lacking a truth value. The biconditional can also be paraphrased as *If A then C, and if C then A.* Suppose neither *A* nor *C* is true. The biconditional is true, but neither of the two conditionals in this conjunction has a truth value. The idea that a conjunction of two assertions having no truth value should somehow yield a true assertion is a recipe for nonsense. The principle of core meanings is therefore corroborated at the expense of defective truth tables. Yet, as is shown below, the present theory captures the intuition underlying the appeal of defective truth tables.

The reader is invited to evaluate an inference based on the following premise:

> In this hand of cards, there is an ace or there is a king, or both.

Given the truth of this premise, does it follow validly that

> If there isn't an ace in this hand of cards, then there is a king?

The reader should judge whether there could be any circumstances in which the premise is true but the conclusion false. Everyone to whom we have given this inference informally has accepted its validity (see also Ormerod, Manktelow, & Jones, 1993; Richardson & Ormerod, 1997). The premise is consistent with the following:

Factual possibilities: ace ¬ king
¬ ace king
ace king

Hence, it follows that if there isn't an ace, there is a king. The validity of the inference demonstrates the existence of the *conditional* interpretation of the "if . . . then" premise.

Another alternative account of conditionals concerns the conditions in which speakers are justified in asserting them. A philosophical tradition deriving from Adams (1975) postulates that the degree to which a conditional *If A then C,* is "assertible" equals the conditional probability of *C* given *A,* that is, $p(C|A)$. A related view (Stalnaker, 1970; Stevenson & Over, 1995, pp. 617–618) is that the probability of a conditional, $p(\textit{If A then C})$, is close to the conditional probability $p(C|A)$. The seductive nature of the claim depends in part on the syntax of English. Certain sentential modifiers are taken to apply only to main clauses but not to subordinate clauses such as the antecedents of conditionals. As a corollary, consider the following:

> What's the probability that if Paolo has the king then he also has the ace?

It is easy to interpret this question as meaning

> If Paolo has the king then what's the probability that he also has the ace?

Such an interpretation, of course, is a direct way of asking for the conditional probability. Investigators must take extra pains to ask for the absolute probability of a conditional instead of this conditional probability, or else participants are in danger of confusing the two. Moreover, naive individuals are likely to base their answers not on the actual possibilities but on mental models of them (see the section The Interpretation of Basic Conditionals). In an unpublished study, Girotto & Johnson-Laird (2002) examined, for example, the following problem:

> There are three cards face down on a table: 3, 6 and 8. Paolo takes one card at random, and then he takes another at

random. Maria says: "If Paolo has the 8 then he has also the 3." What is the probability that Maria's assertion is true?

There are three possibilities for Paolo's hand, and only one of them violates Maria's assertion granted a conditional interpretation, and so the correct answer is 2/3, but if individuals rely on mental models, then they will think only of the case in which the antecedent is true as satisfying the conditional and, accordingly, infer that the probability is 1/3. In contrast, the following question asks for the conditional probability:

Paolo shows his card: It is the 8. What is the probability that he has the 3?

The correct answer is 1/2. The results showed that naive individuals do not tend to give the same estimates for the two probabilities. The majority give the correct answer for the conditional probability, but the modal response for the probability of the conditional was the one based on mental models, and a handful of participants even responded 2/3.

Yet another alternative to our theory is that conditionals do not merely refer to sets of possibilities but rather assert that a relation holds between the antecedent and consequent (e.g., Barwise, 1986). One putative relation is that the antecedent is the cause of the consequent (but cf. Goldvarg & Johnson-Laird, 2001). Another is that the antecedent logically implies the consequent, perhaps taking into account other premises (e.g., Braine & O'Brien, 1991; D. Sperber, personal communication, September 1994). Again, we corroborate our account by showing that any such relation is not part of the meaning of a basic conditional.

Several sorts of counterexample exist to the claim that the consequents of conditionals can be inferred from their antecedents. One sort of counterexample occurs with the *relevance* interpretation of conditionals (see the Ten Sets of Possibilities for Conditionals section). It does not yield the modus ponens inference. For example, the following inference is bizarre:

If you're interested in *Vertigo,* it is on TV tonight.
You're interested in *Vertigo.*
Therefore, *Vertigo* is on TV tonight.

And modus tollens is still odder:

If you're interested in *Vertigo,* it is on TV tonight.
In fact, it is not on TV tonight.
Therefore, you're not interested in *Vertigo.*

Another sort of counterexample is a conditional expressing a deontic relation, such as "If a person is drinking beer, then the person is 19 years of age." Given that, say, Fred is drinking beer, it does not follow that Fred is 19 years of age but only that he ought to be 19 years of age. Yet another sort of counterexample is a conditional with an interrogative consequent, for example, "If the president phones, will you talk to him?" Questions are not inferable from premises.

We do not deny that many conditionals are interpreted as conveying a relation between their antecedents and consequents. However, the core meaning alone does not signify any such relation. If it did, then to deny the relation while asserting the conditional would be to contradict oneself. Yet, the next example is not a contradiction:

If there was a circle on the board, then there was a triangle on the board, though there was no relation, connection, or constraint, between the two—they merely happened to co-occur.

In summary, the basic conditionals *If A then C* and *If A then possibly C* refer to sets of possibilities, and each has its own core meaning: the *conditional* interpretation in which *A* is *sufficient* for *C* and *C* is *necessary* for *A* and the *tautological* interpretation compatible with any possibility. Other interpretations of conditionals do occur, but as we show later, they result from the modulating effects of semantics and pragmatics.

The "Paradoxes" of Implication

Does the principle of core meanings yield a truth-functional account of the meaning of conditionals? On the one hand, the possibilities in the conditional interpretation correspond to the truth-functional connective of material implication. The truth-functional analysis, however, is one that many logicians and philosophers reject (e.g., Stalnaker, 1975). It leads to seeming paradoxes, it cannot be applied to counterfactual conditionals, and it yields interpretations that depart from theorists' intuitions. On the other hand, our theory is based on possibilities, not truth values. In this section, we resolve the apparent paradoxes. In subsequent sections, we show that conditionals are not truth functional.

The so-called *paradoxes* of implication arise because on an analysis of conditionals as material implications, the mere falsity of the antecedent, or the mere truth of the consequent, suffices to establish the truth of the conditional as a whole (see Table 2). Consider the conditional "If there is a circle, then there is a triangle." When it is false that there is a circle, the conditional is true. Likewise, when it is true that there is a triangle, the conditional is true. However, the skeptics say that the following inferences, which are valid on this account, are dubious:

There isn't a circle.
Therefore, if there is a circle then there is a triangle.

And:

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

The paradoxes are readily resolved. The preceding inferences are valid, but their oddity has nothing to do with conditionals. Analogous arguments that naive individuals resist can be derived from disjunctions. For example, the inclusive disjunction "There is not a circle or there is a triangle" is true provided that at least one of its two disjuncts is true; otherwise, it is false. The following two arguments parallel the paradoxes of implication, and they are also rejected by naive reasoners:

There is not a circle.
Therefore, there is not a circle or there *is* a triangle.

And:

> There is a triangle.
> Therefore, there is *not* a circle or there is a triangle.

The causes of the paradoxes are twofold. First, they throw semantic information away. Their premises contain more semantic information, that is, rule out more possibilities, than do their conclusions. Naive reasoners do not spontaneously draw conclusions that throw information away by adding disjunctive alternatives (Johnson-Laird & Byrne, 1991). Second, the judgment of the truth or falsity of assertions containing connectives, such as conditionals and disjunctions, is a meta-ability. That is, it calls for a grasp of the metalinguistic predicates *true* and *false*, which refer to relations between assertions and the world (see, e.g., Jeffrey, 1981). In contrast, a task that taps directly into the interpretation of assertions is to judge what is possible. For instance, suppose that the following conditional holds: "If there is a circle, then there is a triangle." Is it possible that there is triangle? And is it possible that there isn't a circle? Naive individuals answer both questions in the affirmative, and they do not find the task difficult (Barrouillet & Lecas, 1999; Johnson-Laird & Savary, 1996).

Subjunctive and Counterfactual Conditionals

When modal auxiliaries, such as *have* and *will*, are used in the past tense, they can express the subjunctive mood in English. Thus, the following conditional is subjunctive: "If there had been a circle, then there would have been a triangle." One interpretation of this conditional is counterfactual. That is, there was neither a circle nor a triangle, but had there been a circle there would have been a triangle. Such assertions are known as *counterfactual* conditionals. Some counterfactuals are true, for example, "If you hadn't been born, then you wouldn't be alive now." It is false that you weren't born, and it is false that you are not alive now, yet the conditional is true. And some counterfactuals are false, for example, "If you weren't alive now, then you wouldn't have been born." It is false that you aren't alive now, and it is false that you weren't born, but the conditional is false. You could have died yesterday, and yet you would have been born. Hence, among counterfactuals with false antecedents and false consequents, there are some that are true and some that are false. Counterfactual conditionals therefore cannot be truth functional (Quine, 1952).

Philosophers have struggled to frame the truth conditions of counterfactual conditionals. One stratagem owes much to Ramsey's (1929/1990) idea of using a thought experiment to evaluate conditionals. Ramsey wrote as follows in a footnote:

> If two people are arguing "If p will q?" and both are in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q; so that in a sense "If p, q" and "If p, not q" are contradictories. We can say they are fixing their degrees of belief in q given p. If p turns out to be false, these degrees of belief are rendered *void.* (p. 155)

Stalnaker (1968) extended this idea to deal with counterfactuals: You add the antecedent to your set of current beliefs, adjust these beliefs where necessary for consistency, and check whether the consequent of the conditional holds. If it does, the conditional is true; if it does not, the conditional is false. This account has been applied to the circumstances in which conditionals are assertible (e.g., Adams, 1975). It has also been applied within the framework of a "possible worlds" semantics for modal logic (see, e.g., Kripke, 1963). A possible world is, in effect, a possible state of affairs, though one that is completely specified, and the actual world is treated as a member of the set of possible worlds. On Stalnaker's (1968) analysis, the truth of a conditional in a world, w1, depends on another world, w2, that is accessible to w1 and that is the most similar possible world to w1 except that the antecedent of the conditional holds in it. If the consequent is true in w2, then the conditional as a whole is true; otherwise, it is false. It is not obvious that there is always a single world most similar to the actual world except that a counterfactual antecedent is true. If there is always a unique world of this sort, then one of the following pair of assertions is true and one of them is false:

> If Verdi and Bizet had been compatriots, then they would both have been Italian.
> If Verdi and Bizet had been compatriots, then they would both have been French.

Yet, it is impossible to decide which assertion is the true one.

In the light of such examples, Lewis (1973) argued that a conditional is true just in case there is a world in which the antecedent and consequent are true that is closer to the actual world than any world in which the antecedent is true and the consequent false. Thus, consider the truth value in our world of the conditional: "If Quayle had been the Republicans' 1992 Presidential candidate, then Clinton would have lost." It is necessary to find a world that is as alike to the actual world as possible except that Quayle was the Republicans' Presidential candidate. There will be other differences in consequence of this change. If this world supports the truth of the consequent of the conditional, then the counterfactual is true; otherwise, it is false.

These analyses are sophisticated but problematical. One difficulty is illustrated by our example. One could assume that Quayle, as we know him, was nominated as a result of some extraordinary events at the Republican convention. He would have lost the election, and so the conditional is false. Alternatively, one could assume that Quayle underwent an extraordinary mental transformation as a prerequisite for the Republicans to nominate him. He turned out, say, to be wise, though dyslexic. In this case, it is conceivable that he would have beaten Clinton. There is no objective way to determine which of these two worlds is more similar to the actual world (Kratzer, 1989). One cannot observe counterfactual states, and so the truth or falsity of a counterfactual conditional about a contingent matter may never be ascertained. Perhaps that is why counterfactual speculations so intrigue historians, novelists, and sports fans.

Our concern is the meaning of conditionals, not their truth or falsity. The principle of core meanings postulates that indicative conditionals refer to sets of possibilities, and subjunctive conditionals call for only a slight addition to the principle:

> *Principle 2.* The principle of *subjunctive meanings.* A subjunctive conditional refers to the same set of possibilities as the corresponding indicative conditional, but the set consists either of factual possibilities or of a fact in which the ante-

cedent and consequent did not occur and counterfactual possibilities in which they did occur.

To illustrate the principle, consider the subjunctive conditional "If there had been a circle, then there would have been a triangle." The principle implies that this conditional can refer to a set of past factual possibilities, such as the following:

Factual possibilities: ○ △
¬ ○ △
¬ ○ ¬ △

Such conditionals are typically used when a speaker does not know what happened in some past situation but asserts a conditional relation, for example, "If there had been any deserters at Waterloo, then they would have been shot." The subjunctive principle also allows for a counterfactual interpretation:

Fact: ¬ ○ ¬ △
Counterfactual possibilities: ○ △
¬ ○ △

Such conditionals are typically used when a speaker knows what happened but asserts a conditional about an alternative to reality, for example, "If there had been any deserters at Waterloo, not that there were, then they would have been shot." There are subjunctive conditionals corresponding to the tautological interpretation, for example, "If there had been a circle then there might have been a triangle." This conditional refers to

Factual possibilities: ○ △
○ ¬ △
¬ ○ △
¬ ○ ¬ △

It can also refer to a set in which the fourth possibility is a fact (or perhaps the third) and the rest are counterfactual possibilities.

In this section, we have argued that basic indicative conditionals have core meanings that refer to sets of possibilities and that basic subjunctive conditionals also have such core meanings. A natural question to raise is how such meanings are mentally represented. Before we can answer it, however, we need to outline a general theory of comprehension and representation.

The Theory of Mental Models

The theory of mental models was originally postulated in order to explain the comprehension of discourse and elementary deductive reasoning. In this section, we sketch the theory and its computer implementation (see Johnson-Laird & Savary, 1999). By definition, a mental model of an assertion represents a possibility given the truth of the assertion. Hence, a set of mental models represents a set of possibilities. Each model corresponds to a true row in a truth table, though, as shown below, a mental model may not contain all the information in the corresponding row in the truth table. There may also be true rows in the truth table that are not represented explicitly in any mental models. The fundamental representational principle of the theory is as follows:

> The principle of *truth:* Each mental model of a set of assertions represents a possibility given the truth of the assertions, and each mental model represents a clause in these assertions only when it is true in that possibility.

The principle implies that mental models represent only what is true and not what is false. Moreover, each mental model represents a clause in the premises only when the clause is true within the possibility that the model represents. This point can best be explained by way of an example. There are three mental models of an inclusive disjunction, such as "The battery is dead or the circuit is *not* connected, or both." These models are as follows:

Factual possibilities: dead
¬ connected
dead ¬ connected

Each line denotes a separate model, *dead* denotes a model of the clause *the battery is dead* and ¬ *connected* denotes a model of the negative clause *the circuit is not connected.* The first model accordingly represents the possibility in which it is true that the battery is dead, but the model does not make explicit that it is false that the circuit is not connected in this possibility, that is, the circuit is connected. Similarly, the second model represents the possibility in which it is true that the circuit is not connected, but the model does not make explicit that it is false that the battery is dead in this possibility. As the example illustrates, mental models represent negations when they are true, but they do not represent false clauses, whether they are affirmative or negative. Reasoners make mental "footnotes" to capture the information about what is false. The computer program implementing the theory almost literally uses footnotes. To represent the first of the models above, it uses the notation ((dead)(t14)), where t14 is an automatically generated variable with a value of (connected). What happens to footnotes, as we show below, depends on the level of expertise at which the program is operating.

Each entry in a truth table represents the truth or falsity of an assertion given a particular possibility. In contrast, each mental model in a set represents a possibility. A corollary is that possibilities are psychologically basic, not truth values. Discourse about the truth or falsity of propositions is at a higher level than mere descriptions of possibilities. Logicians say that reference to truth and falsity occurs not in the language under analysis—the so-called *object* language—but in a *metalanguage,* which is a language for talking about the object language. Johnson-Laird (1990) argued that children first learn to use language to refer to possibilities. Later, they develop a metalinguistic ability to use the predicates *true* and *false.* Adults are likewise often puzzled by the truth or falsity of assertions containing connectives. For example, if they are told that the assertion "Gawain is a knight, and Lancelot is a knave" is false, then they sometimes assume that it is false that Gawain is a knight and false that Lancelot is a knave (Byrne & Handley, 1992). They overlook that a conjunction can be false even though one of its conjuncts is true. In contrast, they realize that a conjunction is compatible with a single possibility—the one in which both conjuncts hold. Natural language is its own metalanguage. One consequence is that conditionals can make assertions about their own truth or falsity, for example, "If this conditional is true, then its antecedent is true but its consequent is false." Similarly, assertions can make explicit reference to what is possi-

ble or impossible, for example, "If it is impossible that Mallory climbed Everest, then it is impossible that Irvine climbed it."

According to the principle of truth, mental models represent true possibilities. The principle does not imply, however, that individuals never represent false cases. They can use their mental footnotes about what is false to construct *fully explicit* models of what is true. Hence, the inclusive disjunction above has the following fully explicit models:

Factual possibilities:	dead	connected
	¬ dead	¬ connected
	dead	¬ connected

These models can in turn be used to infer what is false, namely, the complement of the set:

Factual impossibility:	¬ dead	connected

The task is difficult because people tend to forget mental footnotes and because it is hard to construct the complement of a set of models (see Barres & Johnson-Laird, 1997).

How are inferences made with mental models? The next example illustrates a simple method. Consider these premises:

> The battery is dead or the circuit is not connected, or both.
> In fact, the battery is not dead.

The disjunction has the mental models presented at the start of this section, and the categorical premise eliminates the first and third of them. The remaining model yields the conclusion "The circuit is not connected." This conclusion is valid, that is, it is necessarily true given the truth of the premises, because it holds in all the models—in this case, the single model—consistent with the premises.

The model theory provides a unified account of inference. If a conclusion holds in all the models of the premises, it is necessary given the premises, that is, it is deductively valid. If it holds in at least one model of the premises, then it is possible (see Bell & Johnson-Laird, 1998). And the probability of a conclusion depends on the proportion of equiprobable models in which it holds or on the sum of the probabilities of the models in which it holds (Johnson-Laird, Legrenzi, et al., 1999). Experimental evidence has corroborated the model theory (e.g., Johnson-Laird & Byrne, 1991). Inferences based on one model are easier than inferences based on multiple models. Reasoners tend to neglect models, and so their systematic errors correspond to a proper subset of the models, typically just a single model. Perhaps the strongest evidence, however, is the phenomenon of *illusory inferences*, which we describe later in the article. They are systematic fallacies that cannot be explained by theories based solely on valid rules of inference.

The computer program implementing the model theory operates at different levels of expertise (Johnson-Laird & Savary, 1999). At its most rudimentary level, Level 1, it forgets mental footnotes. Consider an inclusive disjunction of the form

> A and B, or C and D.

This inclusive disjunction yields the following models:

```
a  b
      c  d
```

Here, as usual, lowercase letters denote models of the corresponding uppercase propositions. At this level, the program does not distinguish between exclusive and inclusive interpretations of a disjunction, which, as the example shows, it represents in two models. It is able, however, to make the correct responses to the 61 so-called *direct* reasoning problems of Braine, Reiser, and Rumain (1984), and the number of models it constructs predicts their difficulty just as well as Braine's theory even though, unlike that theory, it does not depend on parameters estimated from the data (Johnson-Laird, Byrne, & Schaeken, 1992).

At Level 2, the program takes mental footnotes into account in combining models. It distinguishes between inclusive and exclusive disjunctions, and its models for the assertion above are

```
a  b
      c  d
a  b  c  d
```

In this case, there is a footnote that C or D, or both, are false in the first model, and a footnote that A or B, or both, are false in the second model. The contents of footnotes do not emerge into the explicit content of any resulting models.

At Level 3, footnotes not only control a conjunction of sets of models, but now their content emerges into the resulting models. The models for the assertion above are accordingly still more complex:

```
   a     b   ¬ c     d
   a     b     c   ¬ d
   a     b   ¬ c   ¬ d
¬  a     b     c     d
   a   ¬ b     c     d
¬  a   ¬ b     c     d
   a     b     c     d
```

Finally, at Level 4, the program goes beyond human performance. It always constructs fully explicit models, and so it always makes correct deductions. We now turn to the extension of the theory to conditionals.

The Interpretation of Basic Conditionals

The theory postulates the following:

> *Principle 3.* The principle of *implicit models:* Basic conditionals have mental models representing the possibilities in which their antecedents are satisfied, but only implicit mental models for the possibilities in which their antecedents are not satisfied. A mental footnote on the implicit model can be used to make fully explicit models (at Levels 2 and above), but individuals are liable to forget the footnote (Level 1) and even to forget the implicit model itself for complex compound assertions.

According to this principle, the mental models for the conditional interpretation of *If A then C* are as follows:

Factual possibilities: a c
. . .

where the ellipsis denotes the implicit model, which has no explicit content, and which distinguishes a conditional from a conjunction, *A and C.* Similarly, the mental models for the tautological interpretation of *If A then possibly C* are as follows:

Factual possibilities: a c
a ¬ c
. . .

The principle of implicit models also applies to subjunctive conditionals. When a subjunctive conditional, such as "If there had been a circle then there would have been a triangle" refers to unknown past possibilities, it has the following mental models:

Factual possibilities: ○ △
. . .

However, the counterfactual interpretation has two explicit models, one of the fact and one of the counterfactual possibility:

Fact: ¬ ○ ¬ △
Counterfactual possibilities: ○ △
. . .

The presence of two explicit mental models, as we show below, has consequences for conditional inferences.

The models for specific conditional assertions, such as "If Pat has malaria then she has a fever," differ from those for universal conditional assertions, such as "If a patient has malaria then she has a fever" (see Langford, 1992). The specific assertion has the following mental models:

Factual possibilities: Pat malaria fever
. . .

The footnote on the implicit model denotes that it represents the possibilities in which it is false that Pat has malaria. Hence, the conditional's fully explicit models are as follows:

Factual possibilities: Pat malaria fever
¬ malaria fever
¬ malaria ¬ fever

Here we adopt the notational convention that a token only at the start of the first line, such as *Pat* in this case, is an entity for which each model specifies an alternative set of possible properties. Hence, the set of models represents Pat's possible properties, and they allow that Pat could have a fever without malaria. The universal assertion has the following logical form: "For any x, if patient(x) and has(x malaria) then has(x fever)." Variables do not occur in models (Johnson-Laird & Byrne, 1991, chapter 9). The set of patients with malaria is accordingly represented using a small but arbitrary number of tokens:

Factual possibilities: patient malaria fever
patient malaria fever
. . .

Each row in this diagram, unlike previous examples, represents a different individual, and the diagram as a whole represents the set of patients with malaria. A mental footnote indicates that the ellipsis represents entities other than patients with malaria. If the cardinality of the set matters, then models can be tagged with numerals, just as they can be tagged to represent numerical probabilities (see Johnson-Laird, Legrenzi, et al., 1999).

Models of general assertions can be unified with models of particular facts. Consider, for example, the following premises:

If a patient has malaria then she has a fever.
Pat is a patient with malaria.

Given these premises, it follows that Pat has a fever. This inference can be made by unifying the preceding models of Pat's properties and of the set of patients. The result is the following model

Fact: Pat patient malaria fever

From this model, it follows that Pat has a fever.

The Corroboration of the Principle of Implicit Models

A developmental trend supports the model theory. Young children interpret basic conditionals as conjunctions, slightly older children interpret them as biconditionals, and adolescents and adults are able to make the conditional interpretation (see, e.g., Taplin, Staudenmayer, & Taddonio, 1974; cf. Markovits, 1993; Russell, 1987). This result is a nice confirmation of the model theory: Conjunctions have one fully explicit model, which corresponds to the single explicit mental model of the analogous conditional; biconditionals have two fully explicit models; and conditionals have three fully explicit models. Sloutsky and Morris (1999) also observed that children tend to ignore the second clause of a compound premise, so that the premise calls for exactly one model. They are most likely to ignore the second clause of a tautology or contradiction, less likely to ignore it in a disjunction, and least likely to ignore it in a conjunction. Barrouillet (1997) and his colleagues have recently demonstrated the developmental trend in several studies (see Barrouillet & Lecas, 1998; Lecas & Barrouillet, 1999). For instance, Barrouillet and Lecas (1999) carried out a study of 90 children (30 each at the mean ages of about 9, 11½, and nearly 15 years). The children's task was to list the possibilities consistent with basic conditionals. The participants also carried out a task that measured the capacity of their working memories. They counted out loud the number of red dots on sequences of cards, and then at the end of the sequence they had to recall in correct order the number of dots on each card. The results confirmed the predictions. The children showed the developmental sequence of conjunction, biconditional, and conditional. The measure of working memory capacity correlated strongly and significantly with these interpretations ($r = .78$), and it did so even when school grade was partialed out (cf. Simon, 1982).

Other evidence shows that young children interpret conditionals as akin to conjunctions (see, e.g., Delval & Riviere, 1975; Kuhn, 1977; Paris, 1973; Politzer, 1986). O'Brien and his colleagues, however, have argued that *if* is not understood as *and* (O'Brien, 1999, p. 399; O'Brien, Dias, & Roazzi, 1998). Yet, Schroyens, Schaeken, and d'Ydewalle (2000) carried out a meta-analysis of developmental studies, and its results corroborated the model

theory, not the rule theories. Likewise, O'Brien himself proposed the conjunctive interpretation in the past. He wrote "Braine, O'Brien, and Connell found that young children persist in making conjunction-like responses even when provided with evidence to the contrary" (O'Brien, 1987, pp. 74–75). The principle of implicit models implies that individuals should tend to treat basic conditionals as conjunctions, because they discount implicit possibilities in which the antecedent of an indicative conditional is false. The principle has some unexpected consequences that emerged from the computer program implementing the model theory. These predictions led to the discovery of a phenomenon.

Experiments 1 and 2

The computer program implementing the model theory predicts that the mental models of conjunctions, disjunctions, and basic conditionals should interact in a surprising way. Basic conditionals, as readers will recall from the section The Core Meanings of Basic Conditionals, are those that have a neutral content independent of context and background knowledge. The conjunction of two basic conditionals of the form *If A then B, and if C then D* yields the following mental models (at Level 2):

Factual possibilities:	a	b		
			c	d
	a	b	c	d
		...		

The ellipsis denotes an implicit model with a footnote indicating that the antecedents of the two conditionals are both false. However, an inclusive disjunction of the two conditionals yields exactly the same set of mental models. If the first conditional is true, it has the following mental models:

Factual possibilities:	a	b	
	...		

These already occur in the models of the conjunction of the two conditionals. If the second conditional is true, it has the following mental models:

Factual possibilities:	c	d	
	...		

These already occur in the models of the conjunction. And if both conditionals are true, they have the models for the conjunction above. In short, the disjunction requires no additional mental models over and above those elicited by the conjunction (at Levels 1 and 2). However, the fully explicit models of the conditionals, which take into account the falsity of clauses, yield 9 explicit models for the conjunction and 15 explicit models for the disjunction. A disjunction of conjunctions, such as *A and B, or C and D,* has the following mental models:

Factual possibilities:	a	b		
			c	d
	a	b	c	d

The only difference between this set and the set for the two preceding assertions is the lack of an implicit model (denoted by the ellipsis). However, as the developmental trend shows, individuals are likely to forget the implicit model. Indeed, if young children treat conditionals as though they were conjunctions, then complex conditionals should cause adults to regress to conjunctions, too (cf. Case, 1985). They should forget the implicit model of basic conditionals and, in consequence, treat them as akin to conjunctions. The model theory, accordingly, predicts that the three following assertions should tend to be interpreted in the same way even though their real meanings are distinct:

1. If A then B, and if C then D.
2. If A then B, or if C then D.
3. A and B, or C and D.

The theory also predicts that it should be difficult to envisage the possibilities in which these assertions are false: The participants have to infer them from the true possibilities.

There was no evidence in the literature, and so we carried out two experiments to examine these predictions. The first experiment was carried out as a class exercise in an undergraduate course in which each experimenter was a student and tested a separate naive individual. None of the acting experimenters knew the predictions of the model theory. The second experiment was carried out in a conventional way with just one experimenter, and it did not include the disjunction of conjunctions. Otherwise, the two experiments were similar, and so we report them together.

Method. The participants acted as their own controls and listed the sets of possibilities for four assertions. In Experiment 1, the assertions were as follows:

1. If there is a *A* on the board then there is a *2,* and if there is a *C* on the board then there is a *3.*
2. If there is a *D* on the board then there is a *5,* or if there is an *E* on the board then there is a *6,* and both conditionals may be true.
3. There is a *J* on the board and there is a *9,* or there is a *L* on the board and there is a *7,* and both conjunctions may be true.

There was also a control assertion, which was a conjunction of conjunctions. Experiment 2 used Assertions 1 and 2, and two control assertions, a conjunction of disjunctions and a disjunction of disjunctions. The assertions were presented in different random orders to the participants, who wrote down a list of the possibilities given the truth of each assertion, that is, pairings of letters and numbers, such as *A2.* When they had completed this task, they went through the assertions again and wrote down the cases in which they would be false. In Experiment 1, there were 25 participants from a class on cognitive science at Princeton University. In Experiment 2, there were 21 participants, who were students at Princeton University, and who received either a course credit or a monetary reward.

Results and discussion. The results corroborated the predictions. Table 3 presents the percentages of responses to the three experimental assertions, and no other responses occurred reliably. The numbers of participants in Experiment 1 who listed the responses in the order predicted by the computer model were as follows: 9 for Assertion 1, 17 for Assertion 2, and 16 for Assertion 3. There are 16 possible selections, and so the chance probability of making the three selections in their predicted order is very small, that is, the chance probability for one such selection is $1/16 \times 1/15 \times 1/14 < .0003$, and so the binomial probability for, say, 9 out of 25 participants' making this selection is minuscule. The results for Experiment 2 were almost as robust. In both experiments, the control sentences were accurately interpreted, but the task of generating false possibilities was very hard, yielding many varied selections. It seemed that the participants—with considerable difficulty—inferred the false possibilities from the true

Table 3
Percentages of Responses to the Experimental Sentences in Experiments 1 and 2

Possibilities	Assertion 1: If A then B, and if C then D	Assertion 2: If A then B, or if C then D	Assertion 3: A and B, or C and D
A B	84 (62)	92 (90)	92
C D	80 (52)	96 (95)	88
A B C D	80 (76)	96 (52)	76

Note. Values shown are the percentages of participants who listed the predicted possibilities given the truth of the assertions in Experiment 1 ($n = 25$) and Experiment 2 ($n = 21$). The percentages for Experiment 2 are shown in parentheses. The assertions and possibilities are stated as though they had the same lexical materials.

ones. Few participants accounted for all 16 contingencies in their combined lists of true and false possibilities.

The results confirmed the model theory of the interpretation of basic conditionals. Naive individuals take a conjunction of conditionals to have the same possibilities as a disjunction of conditionals and as a disjunction of conjunctions. Hence, their interpretation of basic conditionals regresses to a childlike conjunction. Alternative psychological theories of conditionals, including rule theories, either do not deal with the meaning of connectives or else treat their meanings as captured in the rules of inference that govern them (e.g., Braine & O'Brien, 1991). Because conditionals, disjunctions, and conjunctions are governed by different formal rules of inference, these theories fail to explain why the three sorts of assertion in our experiments are interpreted in the same way.

Another line of evidence corroborates the regression to conjunction-like interpretations of conditionals. Johnson-Laird, Legrenzi et al. (1999) carried out a study in which adults were asked to judge the probability of one assertion given the truth of another initial assertion. The evidence showed that the participants based their estimates on the equiprobability of mental models of the assertions, for example, the participants gave estimates of 33% for one of the three possibilities compatible with an inclusive disjunction. On some trials, the initial assertion was a basic conditional such as "There is a box in which if there is a yellow card then there is a brown card." The conditional has two mental models, one explicit and the other implicit:

yellow brown
. . .

The participants had to infer the probability of the conjunction "In the box there is a yellow card and a brown card," which has just a single mental model corresponding to the explicit model of the conditional. Most participants inferred a probability of 50%, which is predicted from the two models of the conditional. However, over one third of the participants gave an estimate of 100%. This judgment and others in the task show that people often forget the implicit model and, as a result, treat the conditional as though it were a conjunction, which has only a single mental model.

Judgments of Truth and Falsity

Judgments of the truth or falsity of basic conditionals also corroborate the principle of implicit models. Such judgments call for a higher order analysis than merely generating possibilities compatible with a conditional (see the section The Theory of Mental Models). Given a conditional, such as "If there is a circle then there is a triangle," participants usually judge that the conditional is neither true nor false when its antecedent is false. Instead, they deem it irrelevant (J. Evans, 1972; Johnson-Laird & Tagart, 1969). As we have seen, theorists have defended a defective truth table in which a conditional has no truth value when its antecedent is false. The model theory, however, accounts for the result without having the undesirable consequences of the defective truth table (see the section The Theory of Mental Models). There is no explicit mental model representing the possibilities in which the antecedent is false, and so naive individuals and theorists deem the conditional neither true nor false but irrelevant.

Another consequence of conditionals' having only one explicit mental model is that it is easy to confuse one basic conditional, such as "If there is circle then there is a triangle," with its converse, "If there is a triangle then there is a circle." One way for people to grasp the difference between the two conditional interpretations is to generate an additional explicit model of the first conditional,

Factual possibility: ¬ ○ △

They can then note that this model is impossible according to the second conditional. Otherwise, they will tend to confuse the two conditionals. The same confusion can occur in daily life. Thus, for example, it is easy to confuse "If the DNA matches the sample, then the suspect is guilty" with "If the suspect is guilty, then the DNA matches the sample." This confusion may in part be responsible for analogous confusions between the conditional probability of a DNA match given guilt and the conditional probability of guilt given a DNA match (see Johnson-Laird, Legrenzi, et al., 1999).

Unlike indicative conditionals, counterfactual conditionals have two explicit mental models. For example, the counterfactual "If there had been a circle, then there would have been a triangle" has the following mental models:

Fact: ¬ ○ ¬ △
Counterfactual possibilities: ○ △
. . .

It follows that if individuals are asked to list the possibility that best fits a basic conditional, they should tend to list the possibility corresponding to the explicit model of an indicative conditional (the circle and triangle). And they should tend to list the factual case (neither a circle nor a triangle) more often for counterfactual conditionals than for indicative conditionals. Byrne and Tasso (1999) corroborated both these predictions.

The Principle of Semantic Modulation

The model theory is committed to the *compositionality* of the process of interpreting sentences, that is, the meanings of words and phrases are combined according to their syntactic relations to yield the meaning of a sentence (see Johnson-Laird & Byrne, 1991, chapter 9). The computer program implementing the theory

is compositional in exactly this sense. However, the meanings of words are often interrelated, and in this section we describe how the content of the antecedent and consequent can modulate a conditional's core meaning. In the next section, we describe how a conditional's context can affect its interpretation, and then we outline the different possible interpretations of conditionals that these effects of semantic and pragmatic modulation can yield. In both cases, by definition, we are no longer dealing with basic conditionals.

The model theory makes the following assumption:

> *Principle 4.* The principle of *semantic modulation:* The meanings of the antecedent and consequent, and coreferential links between these two clauses, can add information to models, prevent the construction of otherwise feasible models of the core meaning, and aid the process of constructing fully explicit models.

Here is an example of how this principle works. Consider the following conditional:

> If Vivien entered the elevator, then Evelyn left it one floor up.

The coreferential pronoun *it* establishes a spatial and a temporal relation between the two events: Evelyn left the elevator one floor above and after the floor on which Vivien entered it. The conditional is accordingly not truth functional. Given that its antecedent is true, its truth depends on the consequent event occurring in the appropriate spatial and temporal relation to the antecedent event. Another conditional illustrates how semantics can prevent the construction of models:

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

The fully explicit models of the conditional interpretation if they were unconstrained by semantics would be as follows:

Factual possibilities:	It	game	¬ soccer
		¬ game	¬ soccer
		¬ game	soccer

These three models represent the different possibilities for the referent of *it*, assuming that its two occurrences in the conditional are coreferential. However, the meaning of the noun *soccer* entails that it is a game (for different implementations of this assumption, see, e.g., Fodor, Garrett, Walker, & Parkes, 1980; Miller & Johnson-Laird, 1976; Quillian, 1968). Hence, an attempt to construct the third model would yield an inconsistency, because *it* would refer to something that is both a game and not a game. The assertion therefore has just these fully explicit models:

Factual possibilities:	It	game	¬ soccer
		¬ game	¬ soccer

Various referential relations can hold between the antecedent and consequent of a conditional. For example, a pronoun in the consequent can refer back anaphorically to a reference established in the antecedent: "If Fido is tame, then he is obedient." Alternatively, a pronoun in the antecedent, which is a subordinate clause, can refer forward cataphorically to a reference established in the consequent: "If he is tame, then Fido is obedient." Similarly, universal claims, such as "If a dog is tame, then it is obedient," contain a pronoun that functions as a bound variable in the predicate calculus: For any x, if x is a dog and x is tame, then x is obedient. We showed earlier how models can represent such assertions (see the section The Interpretation of Basic Conditionals). Another use of pronouns is merely to stand in place of an earlier *expression* in a sentence, for example, "If he paid tax to the British Inland Revenue, then now he pays it to the American Internal Revenue Service." In this case, *it* is not coreferential with the tax referred to in the antecedent, and the usage is a case of what Geach (1962) called "a pronoun of laziness" (pp. 124–125). The interpretation of the assertion accordingly calls for substituting *tax* for *it* in the consequent clause.

A use of pronouns that is problematic for rule theories is illustrated in the following conditional: "If there is a car in the garage, then Mr. Toad will drive it." The pronoun is not a simple case of coreference, because the antecedent does not establish a definite referent to which the pronoun can refer. The force of the pronoun is to refer to *the car in the garage if there is one.* However, the natural interpretation of this assertion cannot be captured in the first-order predicate calculus (see, e.g., G. Evans, 1980). That is, the following interpretation is not viable: If there exists an x such that x is a car and x is in the garage, then Mr. Toad drives x. The scope of the quantifier, *there exists an x,* is restricted to the antecedent of the conditional, and so the x in the consequent is unbound. One solution is to represent the assertion with a universal quantifier (see, e.g., Reinhart, 1986): For any x, if x is a car and x is in the garage, then Mr. Toad drives x. According to the model theory, however, if there is a car in the garage, the pronoun refers to it, and if there is not a car in the garage, the pronoun has no referent. The assertion accordingly refers to the factual possibilities:

A car in garage	Toad drives it
¬ A car in garage	

Readers may wonder about the different interpretations of conditionals that can arise from the semantic modulation of core meanings. We return to this question after we examine pragmatic modulation.

The Principle of Pragmatic Modulation

Pragmatics concerns the effects on interpretation of the linguistic context of an utterance, its social and physical situation, background knowledge, and the conventions of discourse (Levinson, 1983). We refer to these factors collectively as the *context* of an utterance, and they include the participants' task in psychological experiments (Thompson, 2000). Context can play a part in determining the particular proposition that a sentence expresses, and this proposition in turn determines what inferences reasoners can make from the sentence. Conditionals are notoriously influenced by their context, and so our goals are to describe the underlying mechanism that allows contextual knowledge to modulate sets of models and to outline a computer program that implements the mechanism. Our assumption is that the mechanism is likely to intercede in the case of implicatures based on the conventions governing discourse (see, e.g., Fillenbaum, 1977; Geis & Zwicky, 1971; Grice, 1975; Sperber & Wilson, 1995).

For an illustration of the role of context, consider the following well-known example, which has perplexed many theorists (e.g., Lewis, 1973; Lycan, 1991; Over & Evans, 1997; Stalnaker, 1968):

> If you strike a match properly, then it lights.
> Therefore, if a match is soaking wet and you strike it properly, then it lights.

This inference is valid given the conditional interpretation of the two assertions. Yet, obviously, it is unacceptable in everyday life. There are equivalent inferential problems with disjunctive assertions, for example,

> You put sugar on your porridge, or it doesn't taste sweet.
> Therefore, you put sugar or diesel oil on your porridge, or it doesn't taste sweet.

One reaction to such examples is that their premises are not strictly true. They should be reformulated to state all the relevant information: "If a match has not been soaked in water or treated in any other way that affects its potential for lighting and it is struck properly, then it lights." However, there can never be any guarantee that all the relevant conditions have been captured in the antecedent. The solution accordingly degenerates into a vacuous claim: "If a match is in such a state that it lights when it is struck properly and it is struck properly, then it lights." Moreover, inferences in daily life are made in the absence of complete information. One jumps to a conclusion that reality may overturn. Beliefs such as that properly struck matches light and that sugar makes porridge taste sweet are useful idealizations.

A standard way to treat such idealizations in artificial intelligence is as default assumptions (Minsky, 1975). If a match is struck properly and there is no information to the contrary, then, by default, it lights. Researchers have devised formal systems that incorporate rules or axioms to accommodate reasoning from defaults (see Brewka, Dix, & Konolige, 1997). However, a more psychological approach may emerge from the way that the model theory deals with the indeterminacy of discourse. A model can be built on the basis of default assumptions, and revised, if necessary, by abandoning the default values in the light of subsequent information. Indeed, the theory of mental models relies on this procedure (Johnson-Laird & Byrne, 1991, chapter 9). An analogous idea underlies our approach to the pragmatics of conditionals.

According to the principle of core meanings, the antecedent of a conditional may provide only a partial description of the possibility in which to evaluate the consequent. Richardson and Ormerod (1997) showed the effects of familiarity and causality on paraphrasing conditionals as disjunctions and vice versa. They suggested that familiarity can only aid the process of turning existing partial models into fully explicit models, whereas causality can lead to the fleshing out of implicit possibilities into fully explicit models. Our proposal does not draw this distinction, which may be restricted to the paraphrase task, but makes a more general claim about the effects of knowledge:

> *Principle 5.* The principle of *pragmatic modulation:* The context of a conditional depends on general knowledge in long-term memory and knowledge of the specific circumstances of its utterance. This context is normally represented in explicit models. These models can modulate the core interpretation of a conditional, taking precedence over contradictory models. They can add information to models, prevent the construction of otherwise feasible models, and aid the process of constructing fully explicit models.

No one knows how the mind represents knowledge. It could take the form of assertions in a mental language or a semantic network (e.g., Fodor et al., 1980); content-specific rules, procedures, or productions (e.g., Newell, 1990); or distributed representations (e.g., Rumelhart & McClelland, 1986). However, people often have knowledge about situations, that is, they know what the different possibilities are. The model theory accordingly assumes that such knowledge is represented in explicit models, which modulate the mental models of a core interpretation. In the case of an inconsistency, the resulting unification gives precedence to an explicit model from general knowledge.

We have developed a computer program that implements the essential principle of pragmatic modulation. It operates at three levels of expertise (see the section The Theory of Mental Models), and it uses a knowledge base of explicit models of possibilities to modulate the interpretation of assertions. As an illustration, consider how the program works with the conditional "If a match is struck properly, then it lights," which is represented as

> If match-struck, then match-lights.

And suppose that a match is soaked in water and then struck:

> Match-soaked and match-struck.

What happens? The program begins by building the mental models of the conditional, which at Level 1 are:

> Match-struck Match-lights
> . . .

It adds the information in the second premise to yield the following:

> Match-soaked Match-struck Match-lights

The program's knowledge base includes the fact that if a match is soaking wet, it will not light, which is represented in fully explicit models:

> Match-soaked ¬ Match-lights
> ¬ Match-soaked ¬ Match-lights
> ¬ Match-soaked Match-lights

According to the premises, the match was soaked, and this fact triggers the matching possibility in the knowledge base:

> Match-soaked ¬ Match-lights

The conjunction of this model with the model of the premises would yield a contradiction (and the null model), but the program follows the principle of pragmatic modulation and gives precedence to general knowledge. Hence, the possibility from general knowledge is combined with the model of the premises to yield the following reinterpretation:

Match-soaked, and match-struck, and so it is not the case that match-lights.

The model of the premises also triggers another possibility from the knowledge base:

¬ Match-soaked Match-lights

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

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

As a further illustration, consider the following example suggested by a reviewer:

If today is Monday, then today is Tuesday.
Today is Monday.

The program's knowledge base represents the fact that the days of the week are mutually exclusive:

Monday ¬ Tuesday
¬ Monday Tuesday
¬ Monday ¬ Tuesday

The first of these models takes precedence over the model of the premises:

Monday Tuesday

The program uses the model in the knowledge base to draw the conclusion:

Monday, and so it is not the case that Tuesday.

It uses the model of the premise to trigger the second model in the knowledge base, which it uses to frame the counterfactual conditional:

If it had not been the case that Monday, then it might have been the case that Tuesday.

The models in the knowledge base are compatible with the correct conditional claim:

If today is Monday, then today is *not* Tuesday.

Indeed, because naive individuals judge a conditional as true just in case its antecedent and consequent are both true (see the *Judgments of Truth and Falsity* section), they judge the original conditional as false because it conflicts with this case.

Logically speaking, a set of assertions of the following form is inconsistent:

If A, then C.
If B, then not C.
A and B.

If neither conditional is definitive, then indeed people will treat them as inconsistent, for example,

If he is a friend of Pat, then he is honest.
If he is a friend of Viv, then he is not honest.
He is a friend of Pat and Viv.

People try to reason their way to consistency; that is, if they detect the inconsistency, they try to determine which premise to abandon and to create a diagnosis of the situation (for model accounts of these processes, see Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000; M. S. Legrenzi, Girotto, Legrenzi, & Johnson-Laird, 2000). If one conditional is stated to have a more probable consequent than the other, then reasoners are likely to favor it (George, 1999). If only one conditional rests on definite knowledge, then it takes precedence over the other. And if both conditionals are based on general knowledge, then reasoners may also know which of them takes precedence, for example,

If the boxer punches his opponent below the belt, then he will be disqualified.
If the boxer punches his opponent above the belt, then he will not be disqualified.
The boxer punches his opponent above the belt and below the belt.
Therefore, he will be disqualified.

People know that an illegal action is not offset by a legal action, and so the first conditional takes precedence over the second. One might argue that the second conditional is not strictly true and attempt to amend it. We spelled out the difficulties for this approach in our earlier discussion of the case of the soaked match. It leads ultimately to the need for a vacuous reformulation of the conditional: "If the boxer punches his opponent above the belt and does nothing that results in his disqualification, then he will not be disqualified."

Ten Sets of Possibilities for Conditionals

The principle of core meanings implies that any conditional referring to two or more possibilities must include one that satisfies the antecedent and consequent and that any conditional referring to only one possibility cannot refer to the possibility that falsifies (or violates) the conditional. Semantics and pragmatics, however, allow any other modulation of the two core interpretations. A conditional of the form *If A then possibly C* may refer to all four possibilities of the tautological interpretation, and a conditional of the form *If A then C* may refer to the three possibilities of the conditional interpretation. Modulation can yield either of the other two sets of three possibilities containing the possibility of *A* and *C*, any of the three sets of two possibilities containing the possibility of *A* and *C*, and any of the three single possibilities excluding *A* and *not C*. The theory accordingly predicts that conditionals can refer to 10 distinct sets of possibilities out of the 16 a priori sets for binary connectives. In this section, we illustrate the 10 sets of possibilities, and we also corroborate the occurrence of this sort of modulation. In addition, however, modulation can establish an indefinite number of different temporal, spatial, and coreferential relations between the antecedent and consequent of a conditional.

1. The *tautological* interpretation. This is a core interpretation of the basic conditional *If A then possibly C*. When the antecedent is satisfied, the consequent is possible, and when the antecedent is

not satisfied, the consequent is also possible. For example, the assertion "If there are lights over there, then there may be a road" can refer to the following:

Factual possibilities:	lights	road
	lights	¬ road
	¬ lights	road
	¬ lights	¬ road

On this interpretation, the conditional is a tautology compatible with any state of affairs. The possibility satisfying the antecedent can be factual or counterfactual. A special case of a tautology arises with a conditional such as "If it rains, then it rains," which allows for two possibilities:

Factual possibilities:	rain
	¬ rain

Such a conditional can be used to express events beyond the speaker's control.

2. The *conditional* interpretation. This is a core interpretation of the basic conditional *If A then C.* The antecedent is sufficient for the consequent; the consequent is necessary for the antecedent. For example, the assertion "If the patient has malaria, then she has a fever" can refer to the following:

Factual possibilities:	patient	malaria	fever
		¬ malaria	fever
		¬ malaria	¬ fever

Goldvarg and Johnson-Laird (2001) argued that the meaning of causation is equivalent to the conditional interpretation combined with the temporal constraint that the antecedent possibility does not follow the consequent possibility. The possibilities satisfying the antecedent can be factual, as above, or counterfactual.

3. The *enabling* interpretation. With the tautological interpretation for *If A then possibly C,* the two possibilities in which the antecedent holds are conveyed by the modal qualification in the consequent. However, a common interpretation is that *C* does not occur in the absence of *A.* Thus, the antecedent is necessary for the consequent: *A* is the only enabling condition for *C.* For example, the assertion "If you log on to the computer, then you may be able to receive e-mail" implies that it is impossible to receive e-mail unless you are logged on:

Factual possibilities:	you	log on	receive
		log on	¬ receive
		¬ log on	¬ receive

Staudenmayer (1975) dubbed this interpretation the *reverse conditional.* In an appropriate context, it can be elicited without a modal auxiliary in the consequent (see, e.g., Cosmides, 1989). The possibilities satisfying the antecedent can be factual or counterfactual, for example, "If you had logged on to the computer, then you might have been able to receive e-mail."

4. The *disabling* interpretation. Content or context prevents the construction in the tautological meaning of the possibility in which neither the antecedent nor the consequent occurs. Hence, when the antecedent is satisfied, the negation of the consequent is possible, and when the antecedent is not satisfied, the negation of the consequent is impossible. A "recipe" for generating this interpretation is to fill in the sentence frame *Even if A then C may still occur,* for example, "Even if the workers settle for lower wages then the company may still go bankrupt." The implicature (in the sense of Grice, 1975) is that the firm is bound to go bankrupt if the workers do not settle for lower wages, but settling may disable this outcome:

Factual possibilities:	settle	bankrupt
	settle	¬ bankrupt
	¬ settle	bankrupt

The subjunctive mood allows a counterfactual interpretation.

5. The *biconditional* interpretation. The antecedent is both necessary and sufficient for the consequent. We saw earlier that this interpretation occurs for explicit biconditionals, but many conditionals in everyday life elicit a biconditional interpretation as a result of their content or context. Consider the following assertion: "If he drives the car, then he will crash it," where one knows that he cannot crash the car unless he drives it. Context, however, can exert more subtle effects. P. Legrenzi (1970), in a classic study, showed that people tend to interpret a conditional as a biconditional in a "binary" universe. He used the conditional "If the ball rolls to the left, then the red light comes on" in a situation in which the ball could roll either to the left or right and the light was either red or green. In this case, reasoners' knowledge is as follows:

Factual possibilities:	rolls left	red light
	rolls right	green light

However, even when content and context are neutral, conditionals are often interpreted as biconditionals (see J. Evans, 1982; Staudenmayer, 1975; Staudenmayer & Bourne, 1978; Wason & Johnson-Laird, 1972). Such a biconditional can be counterfactual, for example, "If the ball had rolled to the left, then the red light would have come on."

6. The *strengthened antecedent* interpretation. The theory predicts that a feasible interpretation of a conditional, *If A, then C,* should be one that has just the two models:

Factual possibilities:	a	c
	a	¬ c

Some conditionals with a negative antecedent have this interpretation. Their effect is to imply that at the very least the antecedent is true and that an even stronger claim made in the consequent may be true, for example, "If it doesn't rain, then it'll pour." The verb *to pour* (with rain) means that it will rain heavily, and so the force of the utterance is that it will rain and may rain heavily. The conditional accordingly has the following two models:

Factual possibilities:	rain	pour
	rain	¬ pour

The interpretation can occur with affirmative conditionals when the truth of the antecedent is known to speaker and hearer. For example, the assertion "If there is gravity (which there is), then your apples may fall tomorrow" has the following models:

Factual possibilities:	gravity	apples fall
	gravity	¬ apples fall

7. The *relevance* interpretation. Content or context precludes the possibility in the conditional interpretation in which the consequent does not occur. Typical examples occur when the antecedent asserts merely a condition to which the consequent may be relevant, for example, "If you are interested in *Vertigo,* then it is on TV tonight." Individuals know that the antecedent possibility has no bearing on the occurrence of the consequent, and so the consequent holds in any case:

Factual possibilities:	interested	on TV
	¬ interested	on TV

The conditional can be paraphrased as "If you are interested or not, *Vertigo* is on TV tonight." Other conditionals call for the same set of possibilities on the basis of their content, such as "If you are in contact with the infectious disease, then you're immune":

Factual possibilities:	contact	immune
	¬ contact	immune

Once again, the possibilities can be either factual or counterfactual, for example, "Even if you had been in contact with the infectious disease, you are still immune." This example is a case of a *semifactual* conditional. It has a counterfactual possibility satisfying the antecedent, but the consequent is a fact.

8. The *tollens* interpretation. Conditionals consistent with only a single possibility usually depend on common knowledge between the speaker and hearer. These facts of the matter can be expressed either literally or ironically. Thus, when a consequent is obviously false, it conveys that the antecedent is false too (by analogy with a modus tollens inference). Individuals can use their knowledge of the consequent's falsity to prevent the construction of all models apart from one in which neither the antecedent nor consequent is satisfied, for example, "If that experiment works, then I'll eat my hat." The conditional has only a single model:

Fact: ¬ works ¬ eat hat

However, a counterfactual claim of the sort "If that experiment had worked, then I would've eaten my hat" reintroduces two possibilities:

Fact:	¬ worked	¬ eaten hat
Counterfactual possibility:	worked	eaten hat

What is unusual is that the counterfactual possibility is doubly false, once because of the counterfactual interpretation and once because of irony.

9. The *ponens* interpretation. When an antecedent is obviously true, it conveys that the consequent is true (by analogy with modus ponens). A typical example is "If my name's Alex, then Viv is engaged" when the speaker's name is known to be Alex. The same information can be asserted explicitly: "If my name's Alex, which it is, then Viv is engaged." The conditional has just a single model in which both the antecedent and consequent are satisfied:

Fact: Alex Viv engaged

This sort of conditional cannot be sensibly asserted as a counterfactual. For instance, "If my name had been Alex, then Viv would have been engaged" is bizarre given a speaker whose name is known to be Alex.

10. *Denial of the antecedent and affirmation of the consequent.* An ironic assertion of the following sort is an instance of this interpretation: "If Bill Gates needs money, then I'll be happy to lend it to him." The notion that the world's richest man needs money is plainly false, but the speaker is nevertheless happy to lend him money. The antecedent is false and the consequent is true:

Fact: ¬ need happy to lend

The corresponding counterfactual, "If Bill Gates had needed some money, then I'd have been happy to lend him some," reintroduces an extra possibility:

Fact:	¬ need	happy to lend
Counterfactual possibility:	need	happy to lend

What the theory rules out is the occurrence of affirmative conditionals *If A then C* that have any of the following interpretations (where *nil* signifies the null model, which arises from a contradiction):

	1.								
Factual possibilities:	a	¬ c							
	¬ a	c							
	¬ a	¬ c							
	2.		3.			4.			
Factual possibilities:	a	¬ c	a	¬ c	¬ a	c			
	¬ a	c	¬ a	¬ c	¬ a	¬ c			
	5.		6.						
Factual possibility:	a	¬ c	nil						

We used irony to illustrate some of the predicted interpretations. This maneuver is dangerous, because it could yield interpretations contrary to the theory. Is it possible, for example, to make an ironic interpretation of a conditional that yields an interpretation in which the antecedent is satisfied but the consequent is not? The following sort of assertion ought to yield this interpretation: "If my name is Alex, then Attila the Hun was kindhearted," where it is known that the speaker's name is Alex and that Attila was not kindhearted. Likewise, an assertion of the sort "If you're interested, then Attila the Hun was kindhearted" ought to yield a variant on the relevance interpretation in which the consequent is false. In fact, neither of these interpretations appears to be viable. Likewise, no interpretation of a conditional is a self-contradiction. A conditional of the form *If not A then A* has the following model:

Fact: a

A conditional of the form *If both A and not A then both C and not C* has an interpretation in which both the antecedent and consequent are false. No conditional expresses a self-contradiction, but the negation of a conditional expressing a tautology is self-contradictory, for example, "It is not the case that if it rains, then it rains."

Deontic Conditionals

The antecedents of conditionals refer to possibilities, but the consequents can refer to deontic possibilities. A *deontic possibility* is a situation that is permissible, and a *deontic necessity* is a

situation that is obligatory. Thus, the conditional interpretation of the assertion "If there is a circle, then a triangle is obligatory" refers to the following:

Factual possibilities: ○ △ :Deontic possibilities
¬ ○ △
¬ ○ ¬ △

A circle could occur without a triangle, but it would violate this deontic rule. The deontic conditional "If there is a circle, then a triangle is permissible" has the tautological interpretation. However, if the antecedent is the only possibility in which the consequent is permissible, then the conditional refers to the following:

Factual possibilities: ○ △ :Deontic possibilities
○ ¬ △
¬ ○ ¬ △

If a triangle occurs without a circle, then it violates what is permissible according to this rule.

Do deontic conditionals yield all 10 sets of possibilities open to factual conditionals? The answer is, indeed, that they do, though some of the interpretations are rare. Table 4 includes examples in each category, and we comment only on two that require contextual cues to their meaning. The interpretation that strengthens the antecedent can occur when the truth of the antecedent is common knowledge. Thus, when the speaker and hearer both know that the hearer is allowed to drink alcohol, the deontic assertion "If you're allowed to drink, then you can have a beer" has the following models:

Factual possibilities: drink beer :Deontic possibilities
drink ¬ beer

The interpretation that denies the antecedent and affirms the consequent calls for irony. If it is mutual knowledge that Viv has been cruel toward Pat but that Pat is a devout Christian, a speaker can assert "If Viv has been so *kind* to Pat, then Pat must turn the other cheek and forgive Viv." The antecedent is ironic, whereas the consequent is not. The force of the utterance is, accordingly, as follows:

Factual
possibility: ¬ Viv kind Pat forgive :Deontic possibility

Previous accounts have postulated some of the 10 sets of possibilities (e.g., Veltman, 1986), but to our knowledge no other theory postulates all of them. Table 4 summarizes the 10 sets, giving examples of both factual and deontic conditionals.

Table 4
The Interpretations of Conditionals of the Form If A Then C

Number of possibilities	The ten interpretations		
Four	Tautology a c a ¬ c ¬ a c ¬ a ¬ c If there are lights over there then there may be a road. If she owns the house then she may look out of the window.		
Three	Conditional a c ¬ a c ¬ a ¬ c If the patient has malaria then she has a fever. If he promised then he must take the kids to the zoo.	Enabling a c a ¬ c ¬ a ¬ c If oxygen is present then there may be a fire. If it's her book then she is allowed to give it away.	Disabling a c a ¬ c ¬ a c If the workers settle for lower wages then the company may still go bankrupt. If you're married then you have the right to remain silent.
Two	Biconditional a c ¬ a ¬ c If he drives the car then he will crash it. If she owes money then she must repay it.	Strengthen antecedent a c a ¬ c If there is gravity (which there is) then your apples may fall. If you're allowed to drink then you can have a beer.	Relevance a c ¬ a c If you're interested in seeing Vertigo then it is on TV tonight. If you're interested then he must pay the fine.
One	Tollens ¬ a ¬ c If it works then I'll eat my hat. If it works then I'll be obligated to jump in the lake.	Ponens a c If my name is Alex then Viv is engaged. If I'm a soldier then I must fight.	Deny antecedent and affirm consequent ¬ a c If Bill Gates needs money then I'll lend it to him. If Viv has been so kind to Pat then Pat as a devout person must forgive Viv.

Note. The table presents the set of possibilities referred to by each sort of conditional and everyday examples of factual and deontic conditionals.

Experiment 3: The Corroboration of Modulation

The model theory implies that individuals normally rely on the mental models of conditionals. They can construct fully explicit models, but modulation should yield different interpretations, that is, different fully explicit models depending on semantics and pragmatics. We have carried out several studies in which we manipulated content in order to modulate the interpretation of conditionals. Their results corroborated the theory. These experiments showed that reasoners had no difficulty in listing the appropriate possibilities for conditionals, that modulation affected their performance, but that it was not always easy to elicit certain interpretations. We report here an experiment in which the participants listed what was possible and impossible for each of the seven sorts of conditional that have two or three fully explicit models (see Table 4). We avoided conditionals that have only one model, because their interpretation often depends on irony. Four of the conditionals had consequents containing the modal auxiliary *may:*

Tautology: If there are lights over there then there may be a road.
Enabling: If you log on to the computer then you may be able to receive e-mail.
Disabling: If it is sunny then it may also be cloudy.
Strengthened antecedent: If there is gravity (which there is) then your apples may fall tomorrow.

Three of the conditionals had consequents without the modal auxiliary:

Conditional: If the patient has malaria, then she has a fever.
Biconditional: If it's heated, then this butter will melt.
Relevance: If you're interested, then Letterman is on TV tonight.

Method. The participants acted as their own controls and listed what was possible and impossible (in any order) for each of the seven conditionals, which were presented in a different random order to each of the participants. Three of the conditionals were of the form *If A then C,* and four of them were of the form *C if A;* and the participants were assigned alternately to one of two counterbalanced assignments of these forms to the sentences. We tested 22 Princeton University undergraduates who were fulfilling a course requirement. One participant failed to carry out the task properly, stating possibilities that referred only to one of the clauses in the conditional, and so we replaced him with another participant.

Results and discussion. Table 5 presents the most frequent interpretations for the seven conditionals. Overall, the results corroborated the model theory's predictions. As the table shows, there was a reliable effect of modulation, and the participants tended to make an interpretation corresponding to its predicted effects. The mean number of different interpretations that the participants made was 5.05, and the mean number of interpretations that fit the predicted modulation was 3.7. Because there are at least seven possible interpretations of a conditional, any participant who made at least two of the predicted interpretations showed a bias toward the predictions. In fact, all 22 participants showed such a bias (binomial $p = .5^{22}$); likewise, 20 of the participants were more likely to make a predicted interpretation than not, 1 participant went against the trend, and there was one tie (binomial $p < .000001$). All the participants tended to begin with the possibility corresponding to an explicit mental model of the conditional (binomial $p = .5^{22}$).

In general, all the predicted interpretations did occur (see Table 5). For five of the seven conditionals they occurred more often than any other interpretation, and for six of the seven conditionals they occurred more often than expected by chance (binomial probabilities ranged from $p < .03$ to $p < 1$ in 10 million, assuming that the chance probability of the predicted interpretation is 1/7). The disabling conditional elicited its predicted interpretation only four times. Although it was the only conditional to yield this interpretation, it elicited more often the tautological interpretation and the enabling interpretation. With hindsight, our choice of the sentence "If it is sunny, then it may also be cloudy" was injudicious. The participants considered that it was possible to have no clouds even when it was not sunny, as at nighttime. The biconditional sentence elicited a conditional interpretation more often than its predicted interpretation. The participants considered it possible for butter to melt when it was not heated. Another of our studies, however, readily elicited the biconditional interpretation using the following assertion: "If he drives the car, then he will crash it."

Table 5
The Number of Participants Making the Most Frequent Interpretations in Experiment 3 (N = 22)

	The true possibilities in the participants' interpretations							
	a c	a c	a c	a c	a c	a c	a c	
	a ¬ c	a ¬ c	a ¬ c	a ¬ c	¬ a c	¬ a ¬ c	¬ a c	
	¬ a c	¬ a ¬ c	¬ a c		¬ a ¬ c			
The seven sorts of sentence	¬ a ¬ c							Total
Tautology	11	6			2			19
Enabling	4	17						21
Disabling	8	6	4					18
Strengthening antecedent		2		11				13
Conditional					17	2		19
Biconditional					11	7		18
Relevance						4	14	18

Note. The table shows only those interpretations made by more than 1 participant. The predicted interpretations are given along the diagonal.

Modulation is a robust phenomenon, though its effects are sometimes unpredictable. The occurrence of the conditional interpretation is another piece of evidence against the hypothesis that conditional assertions have a defective truth table. The conditional interpretation is equivalent to material implication, though we elicited it in judgments of possibility and impossibility rather than of truth and falsity.

Conditional Reasoning

Many studies have been made of the four main conditional inferences: modus ponens, modus tollens, affirmation of the consequent, and denial of the antecedent (see the Introduction). Only modus ponens and modus tollens are valid for the conditional interpretation, whereas all four inferences are valid for the biconditional interpretation. Some instances of modus ponens are so easy to make that psychologists have argued that their validity must depend on a formal rule of inference (e.g., Macnamara, 1986; Falmagne & Gonsalves, 1995). Formal rules, by definition, are blind to content. They make no distinction between basic conditionals and conditionals of other sorts. Thus, as J. Macnamara (personal communication, 1989) wrote:

> By a formal logical rule, I take it, we mean a rule that applies to a string in virtue of its form. That is, the rule can apply whenever a string is described as having a certain form . . . The question of whether there is a psychological version of this rule in the minds of normal people (not trained in logic) turns on whether they have a secure intuition, applying equally to any content, that [the rule applies]. I take it that they have. And for me, that's an end of it.

Is modus ponens applicable in this universal way? Several skeptics have claimed that it is not. Thus, Lycan (1991) argued that "no interesting 'rules' of inference are even normatively valid in the first place" (p. 5). He reviewed four sorts of counterexample to modus ponens, which we examine in turn. First, a speaker asserts "I'll be polite if you insult me, but I won't be polite if you insult my wife." The hearer insults both the speaker and his wife. The example, which is due to Allan Gibbard (as cited in Lycan, 1991), illustrates how contradictions can arise from modus ponens. Individuals are likely to treat this case like our earlier example of a boxer punching both above and below the belt (see the section The Principle of Pragmatic Modulation). They know that insults to spouses take precedence, like blows below the belt, and so they infer that the speaker will not be polite. Second, consider the following sequence of premises:

> If Albert comes to the party, it will be great.
> If Albert and Betty come to the party, it will be awful.
> If Albert and Betty and Carl come to the party, it will be great.
> . . . and so on, with alternating consequents.

Each assertion can be treated as true. Yet, given that both Albert and Betty come to the party, modus ponens from the first premise yields the conclusion that the party will be great, and modus ponens from the second premise yields the conclusion that it will be awful (see Lewis, 1973, for the provenance of the example). It hinges on the fact that antecedents often establish only a partial context (see the principle of core meanings). Hence, the mutual context of the set of conditionals as a whole modifies the interpretation of their individual antecedents:

> If Albert comes to the party *without Betty,* it will be great.
> If Albert and Betty come to the party *without Carl,* it will be awful.
> . . . and so on.

Third, consider the following immediate inference:

> If my good friend Smedley finishes his book, I'll be happy.
> Therefore, if my good friend Smedley finishes his book and concludes it with a vicious attack on me, I'll be happy.

This inference is of the form known as *strengthening the antecedent:*

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

Such inferences are valid deductions with material implications, but not, as in the present case, with everyday conditionals. Material implications are true if their consequents are true and if the antecedents are false (see Table 2), and so in strengthening the antecedent of a material implication the conclusion is true if the premise is true. Our elucidation of this example rests once again on the partial context that the antecedent describes (cf. our earlier analysis of striking a soaking wet match). Fourth, there is McGee's (1985) example based on the 1980 U.S. presidential election. The contest was between Ronald Reagan and Jimmy Carter, with the Republican John Anderson a distant third. The following premises asserted prior to the election are true:

> If a Republican wins the election, then if it's not Reagan who wins, it will be Anderson.
> A Republican will win.

The following conclusion is valid according to modus ponens:

> Therefore, if it's not Reagan who wins, it will be Anderson.

From a formal standpoint, the inference is valid; compare

> If there's a letter on the board, then if it's not an A it's a B.
> There's a letter on the board.
> Therefore, if it's not an A then it's a B.

However, many people judge that the conclusion to the Reagan inference is false, because if Reagan had not won, then Carter would have won. The elucidation of their view depends once again on background knowledge. They know that Reagan and Anderson cannot both win. Hence, the fully explicit models of the first premise are as follows (where *Republican* represents a Republican as winning):

Factual possibilities:			
	Republican	¬ Reagan	Anderson
	Republican	Reagan	¬ Anderson
	¬ Republican	¬ Reagan	¬ Anderson

They also know that either Reagan or his Democratic opponent, Carter, will win. When models of this knowledge are unified with the preceding possibilities, the result is as follows:

Factual possibilities:	Republican	Reagan	¬ Anderson	¬ Carter
	Democrat	¬ Reagan	¬ Anderson	Carter

The premise that a Republican will win is combined with these models, and the result is that Reagan will win, and so the conditional conclusion in the example does not follow from the premises.

Lycan (1991) drew the following moral: In certain contexts and with certain interpretations, modus ponens is valid. He denied only that such inferences are valid in virtue of their form. We agree. Problems arise only if one envisages a formal rule of modus ponens that is automatically triggered by any sentences that match its syntactic form, regardless of their content or context.

The Four Inferences

The model theory predicts that the fewer the models that have to be constructed for an inference, the easier the inference should be: Reasoners should be faster to reach a conclusion and more likely to be correct. A special case of this prediction is that those inferences that can be drawn from mental models should be easier than those that can be drawn only from fully explicit models. This prediction has many ramifications for reasoning with conditionals, and we examine the important cases.

The first consequence is that modus ponens should be easier than modus tollens, because modus ponens can be drawn from the mental models of a conditional interpretation, whereas modus tollens can be drawn only from fully explicit models. Indeed, the most robust phenomenon in conditional reasoning is that modus ponens is easier than modus tollens—participants often respond that nothing follows from the modus tollens premises (see J. Evans et al., 1993, for a review). The model theory readily accounts for this response. The mental models of a conditional, *If A then C,* are as follows:

Factual possibilities:	a	c
	. . .	

At the most primitive level of performance (Level 1 in the program described earlier), which does not take footnotes into account, these models do not distinguish between a conditional and a biconditional interpretation. At Level 2, the conditional interpretation has a footnote indicating that the implicit model represents possibilities in which *A* is false, and the biconditional interpretation has a footnote indicating that the implicit model represents possibilities in which both *A* is false and *C* is false. The categorical premise for modus ponens, *A,* eliminates the implicit model, and the categorical conclusion, *C,* follows from the remaining model. When the same models are combined with the categorical premise for modus tollens, *Not C,* the only result is the model ¬ c, and so it seems that nothing follows.

The theory predicts that performance on modus tollens should be improved by any manipulation that helps reasoners to cease focusing on the explicit mental model of the conditional and to convert the footnote on the implicit model into an explicit model of the possibility in which the antecedent is false. Girotto, Mazzocco, and Tasso (1997) corroborated this prediction. They manipulated the order of the two premises. When the conditional occurs first, working memory is already preoccupied during the interpretation of the categorical premise, and so reasoners are unlikely to construct an additional fully explicit model. However, when the categorical premise occurs first, it enables reasoners to reject the explicit mental model immediately. This step frees up the processing capacity of working memory and allows reasoners to use the categorical premise to construct the fully explicit model, ¬ a ¬ c, which yields the conclusion, *not A.*

Another way in which to enhance modus tollens should be to use materials that help the participants to construct fully explicit models. As the theory predicts, modus tollens is easier with a biconditional interpretation, which has just two fully explicit models, than with a conditional interpretation, which has three fully explicit models (Johnson-Laird et al., 1992). Likewise, the difference between modus ponens and modus tollens tends to disappear with conditionals of the form *"A only if C,* because the negative force of *only* highlights the possibility of *not C* (J. Evans, 1977; J. Evans and Beck, 1981). Still another way to enhance modus tollens is to use subjunctive conditionals, such as "If Linda were in Galway then Cathy would be in Dublin." The counterfactual interpretation makes explicit the possibility required for modus tollens:

Fact:	¬ Linda in Galway	¬ Cathy in Dublin
Counterfactual possibility:	Linda in Galway	Cathy in Dublin

As the theory predicts, reasoners made more modus tollens inferences from subjunctive conditionals than from indicative conditionals (Byrne & Tasso, 1999). Some reasoners, as Thompson and Byrne (2000) showed, make the counterfactual interpretation of subjunctive conditionals, but others make the factual interpretation or focus on only one of the two models above. These differences predict the inferences that they are likely to draw, for example, the counterfactual interpretation increases modus tollens and denial of the antecedent in comparison with the factual interpretation.

A second consequence of the model theory, as J. Evans (1993) pointed out, is that affirmation of the consequent should be more frequent than denial of the antecedent. If individuals reason at a primitive level, ignoring footnotes, they construct the following mental models for a basic conditional, *If A then C:*

a	c
. . .	

They will therefore affirm the consequent, but they will not deny the antecedent, because when these models are combined with the categorical premise, *not A,* the result is the model ¬ a, from which nothing follows. If reasoners use mental footnotes to construct fully explicit models, then they should refrain from both inferences unless they make the biconditional interpretation:

a	c
¬ a	¬ c

The theory therefore predicts that affirmation of the consequent should occur more often, and more rapidly, than denial of the antecedent. Although some authors have claimed that the literature is equivocal (e.g., J. Evans, 1993; O'Brien et al., 1998), a recent systematic study showed these predicted effects on both latencies and percentages of responses (Barrouillet, Grosset, & Lecas,

2000). Earlier studies have also shown significant biases in favor of affirmation of the consequent over denial of the antecedent (e.g., Wason, 1964). Likewise, Schroyens, Schaeken, and d'Ydewalle's (2000) meta-analysis of the literature showed that affirmation of the consequent is more frequent than denial of the antecedent. Instead of a general process of fleshing out models explicitly, Schroyens, Schaeken, and d'Ydewalle proposed that reasoners engage in a goal-directed process of searching for explicit models that refute either a putative conclusion or the failure to draw a conclusion—for example, to refute affirmation of the consequent, they search for a model in which the consequent holds and the antecedent is false:

Factual possibility: ¬ a c

Several other factors should affect basic conditional inferences. When individuals carry out a series of inferences, they develop different *strategies* for coping with them (Johnson-Laird, Savary, & Bucciarelli, 1999). Some reasoners consider the possibilities compatible with the premises; others follow up the consequences of a categorical premise or supposition; still others convert all the premises into a chain of conditionals. Hence, different individuals are likely to develop different strategies for coping with conditional inferences. Another factor that should affect conditional inferences is the so-called *figural effect*, which is the tendency for reasoners to draw conclusions interrelating items in the same order in which information about those items entered working memory (see Bauer & Johnson-Laird, 1993; Johnson-Laird, 1975). The effect, as J. Evans (1993) emphasized, should increase modus ponens and denial of the antecedent and decrease modus tollens and affirmation of the consequent.

Modulation and Inferences

Semantic and pragmatic modulation can yield models that undermine the sufficiency or the necessity of the antecedent for the occurrence of the consequent (Thompson, 1994, 1995). In the case of the soaked match, for instance, the antecedent is no longer sufficient for the consequent. Byrne (1989) established that individuals made modus ponens from such premises as the following:

If he went fishing then he had a fish supper.
He went fishing.

They inferred that he had a fish supper. However, they tended not to draw this conclusion with the addition of certain conditional premises, such as the following:

If he went fishing, then he had fish for supper.
If he caught some fish, then he had fish for supper.
He went fishing.

This suppression counts against Macnamara's (personal communication, 1989) litmus test for a formal rule of modus ponens, which should apply regardless of context (see the beginning of the present section). Politzer and Braine (1991) argued that the additional conditional falsifies the first conditional. The claim is, in effect, that conditionals with incomplete descriptions of their antecedent contexts are false. However, as our earlier account of the soaked match showed (see the section The Principle of Pragmatic Modulation), this analysis of pragmatic effects is not viable. Moreover, Byrne, Espino, and Santamaría (1999) demonstrated suppression without affecting the believability of the original conditional (see also Byrne, 1991). According to the principle of pragmatic modulation, the second conditional makes salient to reasoners that to go fishing in itself is not sufficient for a fish supper. One also has to catch fish. Hence, reasoners balk at making modus ponens from a premise that merely asserts that a person went fishing.

Modulation can also affect inferences when reasoners themselves generate their own instances that render the antecedent unnecessary (Markovits, 1984; Markovits & Vachon, 1990) or insufficient (Cummins, Lubart, Alksnis, & Rist, 1991). Stevenson and Over (1995) argued that the sufficiency of an antecedent for a consequent is represented by the proportion of equiprobable models in which the antecedent holds and in which the consequent also holds. This idea fits the model-based theory of extensional probabilities (see Johnson-Laird, Legrenzi, et al., 1999). Indeed, Stevenson and Over showed that lowering the credibility of a conditional suppresses inferences from it in exactly the way that modulation predicts.

In the absence of semantic cues, certain inferences become more difficult (see Byrne & Handley, 1997; Byrne, Handley, & Johnson-Laird, 1995). Similarly, reasoners are likely to reject unbelievable conclusions (see the large literature on effects of belief, e.g., J. Evans et al., 1993). Participants in a study carried out by Santamaría, García-Madruga, and Johnson-Laird (1998) were given the following sort of premises:

If Ann is hungry then she has a snack.
If she has a snack then she eats a light supper.

A conditional interpretation yields the transitive conclusion:

∴ If Ann is hungry then she eats a light supper.

The participants, however, tended to declare that nothing followed from the premises. The conclusion seems unbelievable. It asserts, contrary to the normal causal relation, that hunger leads to a light meal. It lacks the causal intermediary—the snack—that explains the unusual relation. Once again, people's beliefs modulate the models that they construct, and they construct the null model in the case of conflict.

Experiment 4: Modulating Modus Tollens

Modulation should have a major impact on modus tollens. Some contents should make the inference easier, and some should make it harder. For example, consider the following premises:

If Bill is in Rio de Janeiro then he is in Brazil.
Bill is not in Brazil.

These premises should readily yield the following conclusion:

Bill is not in Rio de Janeiro.

Reasoners know that Rio de Janeiro is in Brazil, and so if Bill is not in Brazil then he cannot be in Rio. Their familiarity with the spatial inclusion in the conditional should yield an explicit model of the possibility referred to in the categorical premise:

Bill ¬ In Rio ¬ In Brazil

Premises of this sort should yield more modus tollens conclusions than premises based on unfamiliar spatial inclusions, such as:

If Ann is in the Champagne Suite then she is in the Hotel LaBlanc.
Ann is not in the Hotel LaBlanc.

Spatial exclusions should inhibit modus tollens, for example:

If Bill is in Brazil then he is not in Rio de Janeiro.
Bill is in Rio de Janeiro.

Reasoners should tend balk at the following conclusion:

Bill is not in Brazil.

They know that if Bill is in Rio, then he must be in Brazil. Premises of this sort should yield fewer modus tollens conclusions than premises based on unfamiliar spatial exclusions, such as the following:

If Ann is in the Hotel LaBlanc then she is not in the Champagne Suite.
Ann is in the Champagne Suite.

Semantic modulation accordingly predicts that reasoners should make modus tollens inferences more often from spatial inclusions than from spatial exclusions, and pragmatic modulation predicts that this effect should be larger for the familiar relations than for the unfamiliar relations.

Method. The participants were their own controls and drew conclusions to two familiar and to two unfamiliar inclusions, and to two familiar and to two unfamiliar exclusions. In addition, there were two filler inferences in the form of modus ponens. The resulting 10 sets of premises were presented in a different random order to each of the participants. Their task was to write down what, if anything, followed necessarily from the premises. The instructions explained that a conclusion follows necessarily if it must be the case given that the premises are true. The materials for the modus tollens inferences concerned geographical locations and locations in hotels and were assigned twice at random to the forms of inference. We tested 50 undergraduates at the University of Dublin, who participated voluntarily, but we rejected the results from the 9 participants who failed to make responses to all the premises.

Results and discussion. Table 6 presents the percentages of correct conclusions to the four sorts of problem. The results corroborated the model theory's predictions. The participants drew twice as many modus tollens conclusions to the spatial inclusions as to the spatial exclusions (Wilcoxon test, $z = 4.33$, $p < .0001$). No reliable difference occurred between the familiar and the unfamiliar premises (Wilcoxon test, $z = 1.42$, $p < .16$). However, as predicted, the interaction was reliable: The difference between the inclusions and exclusions was greater for the familiar relations than for the unfamiliar relations (Wilcoxon test, $z = 2.72$, $p < .01$). For the inclusions, the participants drew more modus tollens conclusions from the familiar relations than from the unfamiliar relations (Wilcoxon test, $z = 2.14$, $p < .05$), whereas for the exclusions they drew fewer modus tollens conclusions from the familiar relations than from the unfamiliar relations (Wilcoxon test, $z = 2.17$, $p < .05$).

Table 6
The Percentages of Modus Tollens Conclusions for the Four Sorts of Inference in Experiment 4: The Spatial Inclusion and Exclusion Problems With Familiar and Unfamiliar Relations

Contents	Inclusions	Exclusions	Overall
Familiar	92	34	63
Unfamiliar	82	54	68
Overall	87	44	

This study shows an effect of content on modus tollens. Reasoners' grasp of spatial inclusions makes them more likely to make the inference, whereas their knowledge of spatial exclusions makes them less likely to make the inference. The phenomena reflect both semantic and pragmatic modulation. The model theory predicts both effects. In contrast, if reasoning is a formal process based on rules, it is difficult to explain why content has such a large effect on performance.

The Effects of Negation on Conditional Reasoning

In a major series of studies, Evans and his colleagues have investigated the effects of negation on reasoning with basic conditionals (see, e.g., J. Evans, 1989). These studies have established the occurrence of systematic biases. Evans emphasized that he uses the term *bias* as a description of a phenomenon, not as an explanation. One such bias occurring with the four main conditional inferences is that reasoners are more likely to draw negative conclusions than affirmative conclusions (J. Evans, 1977; cf. Dugan & Revlin, 1990; Wildman & Fletcher, 1977). At one time, Evans suggested that reasoners may merely have been cautious about drawing affirmative conclusions (J. Evans, 1977, 1982). However, he later conceded that this explanation is unsatisfactory (J. Evans, 1993). The bias does not occur with modus ponens, with premises of the form *A only if C* (J. Evans, 1977), or with other sorts of deduction. The model theory suggests an alternative explanation. It is relatively hard to draw an affirmative conclusion from a denial of the antecedent in the following case:

If A then not C.
Not A.
Therefore, C.

Mental models of the conditional do not allow any conclusion to be drawn:

Factual possibilities: a ¬ c
. . .

The conclusion follows only from the fully explicit models of a biconditional interpretation. To construct such an interpretation, it is necessary to work out the possibility in which *A* is false, and in which *not C* is false. The latter is in effect a double negation: If *not C* is false, then *C* is true. The fully explicit model in the biconditional interpretation is, accordingly, as follows:

Factual possibility: ¬ a c

The categorical premise, *Not A*, now yields the double negative conclusion: *C*. The theory explains the bias as a result of the difficulty of inferring affirmative conclusions that depend on a double negation. Johnson-Laird suggested this explanation in a review of J. Evans (1993), and Evans and Handley (1999) en-

dorsed it. The explanation has the advantage that it does not predict any bias for modus ponens, and no exceptions to it are in the data that Evans (1993) reported. Likewise, according to the model theory, *A only if C* has two explicit mental models, and so they do not need to be made explicit. Hence, the explanation does not predict the bias for these assertions. Evans has proposed a similar hypothesis. He has observed that individuals are less inclined to make a valid inference if it calls for the falsification of a negative component, that is, the negation of a negation. The phenomenon occurs with conditional and disjunctive reasoning (e.g. J. Evans, 1972; J. Evans & Newstead, 1977; Roberge, 1971, 1974). In short, the model theory explains the bias: It is difficult to draw affirmative conclusions from double negations.

Evans and his colleagues have proposed an important simplification to the model theory, which we have incorporated in the present account (see J. Evans, 1993; J. Evans, Clibbens, & Rood, 1996; J. Evans & Handley, 1999). The original theory had postulated that a negative assertion was likely to elicit models of both the assertion and its corresponding unnegated proposition (Johnson-Laird & Byrne, 1991). Thus, a conditional of the form *If not A then C* was supposed to elicit the following mental models:

¬ a c
a

The point of the assumption was to account for "matching" bias. This bias is the tendency for reasoners to ignore negatives in conditionals in matching them to categorical premises or to cards in the selection task. Evans and his colleagues, however, have discovered the key to the phenomenon. It depends on the relative difficulty of grasping that one assertion refutes another. This task is easier with explicit negations than with implicit negations. It is thus easier to understand that the assertion "The number is not 4" refutes "The number is 4" than to understand that "The number is 9" refutes "The number is 4" (cf. J. Evans & Handley, 1999; J. Evans, Legrenzi, & Girotto, 1999). The difficulty of implicit negation, in turn, appears to depend on background knowledge about the size of the contrast class (Oaksford & Stenning, 1992; Schroyens, Verschueren, Schaeken, & d'Ydewalle, 2000). The model theory can accordingly rely on the simple unadorned principle of truth. Mental models represent true assertions, whether they are affirmative or negative, but not false assertions.

The Selection Task

The selection task, which we described in the Introduction, has launched more studies than any other form of conditional reasoning, but the literature has outpaced knowledge. Because the selection task is complex, the phenomena are diverse. The model theory appears to explain them, and it yields some new predictions. The selection task calls for a meta-linguistic grasp of *truth* and *falsity* and for the ability to think about potential evidence—such as numbers and letters on the other side of cards—bearing on the truth or falsity of assertions. Reasoners need to represent the possible counterexamples to conditionals, that is, they need to overcome the principle of truth and to envisage what falsifies a factual conditional or what violates a deontic conditional. For example, the conditional *If there is an A then there is a 2* has the following counterexample:

Impossibility: a ¬ 2

Reasoners then need to match each of the cards to the counterexample and to choose only those cards that could be instances of it. The outcomes in this case are as follows:

A: Select as a potential counterexample.
B: Do not select, because it cannot be a counterexample.
2: Do not select, because it cannot be a counterexample.
3: Select as potential counterexample, because 3 is a case of ¬ 2.

According to the model theory, there are three sources of difficulty in the selection task. First, reasoners may fail to realize the need to consider counterexamples to the conditional. Second, they may have difficulty in working out the counterexamples, because they have to construct them from their mental models of the assertion. This failure to construct counterexamples is critical, and it explains the differences between the selection task and the main conditional inferences (Markovits & Savary, 1992). Semantic and pragmatic modulation have a major influence on the ease of envisaging counterexamples. Third, reasoners may have difficulty in understanding negation, for example, that *3* is an instance of the model ¬ 2. With a basic conditional of the form "*If A, then 2,*" naive individuals tend to construct the following mental models and to base their selections on them:

Factual possibilities: a 2
. . .

They select *A*, and they select *2* in addition unless they construct the following explicit model:

Factual possibility: ¬ a 2

F. Cara and Broadbent (personal communication, 1992) demonstrated the correlation between the conditional interpretation and the selection of the *A* card, and the biconditional interpretation and the selection of the *A* and the *2* cards.

Griggs and Jackson (1990) established a related phenomenon in studies testing a prediction from Margolis (1987). When participants had to circle just two cards that violate a basic conditional of the form "If *A*, then *2*," then a greater proportion of them selected *not A* and *not 2* than in the standard task. According to the model theory, this response is a result of a biconditional interpretation, which can yield the fully explicit models:

Factual possibilities: a 2
¬ a ¬ 2

Reasoners in the standard task are reluctant to select all four cards corresponding to these models, because the "demand characteristics" of the experiment suggest that this response would be inappropriate. Those that make the biconditional interpretation accordingly select the *A* and the *2* cards, which are more salient. When reasoners have merely to select two cards, however, then those who made the biconditional interpretation can select those cards that match its second explicit model.

When conditionals contain negated antecedents or consequents, the participants select the card satisfying the antecedent, but they show the "matching" bias for the consequent, that is, they ignore negation in the consequent and choose the card that it mentions (J. Evans, 1989; J. Evans & Lynch, 1973). Hence, given the condi-

tional "If there is an *A* then there is not a *2*," they make the *correct* selection of the *A* and *2* cards, unlike their error with affirmative conditionals. The model theory, following J. Evans and Handley (1999), explains this choice in terms of the comparison between the mental models of the conditional and the four cards. The participants do not have a genuine insight into the task. They merely fail to grasp that the *3* card is a true instance of the case represented by ¬ 2. The participants' apparent insight into the task disappears when the four cards are explicitly labeled as follows (J. Evans & Handley, 1999):

A not A 2 not 2

They select the *A* and *not 2* cards to test the conditional "If *A*, then not *2*."

According to the model theory, the main stumbling block to correct selections is the failure to construct counterexamples. The theory accordingly makes a key prediction:

> Any experimental manipulation that fosters explicit models of counterexamples to conditionals should enhance performance in the selection task.

Semantic or pragmatic modulation can lead individuals to construct counterexamples directly or to flesh out explicit models of the conditional. Participants in another early study had to select those instances that would violate a deontic conditional that was familiar to them, for example, "If an envelope is sealed, then it has a 5 penny stamp on it." They already knew what was permissible:

Factual possibilities:	sealed	5 penny	:Deontic possibilities
	¬ sealed	5 penny	
	¬ sealed	4 penny	

The complement of these models represents the violation of the rule:

Factual possibility: sealed 4 penny :Deontic impossibility

Hence, the participants performed much better with these conditionals than with basic conditionals, and there was no transfer from them to the basic conditionals (Johnson-Laird, Legrenzi, & Legrenzi, 1972). Individuals who were not familiar with the postal regulation that inspired these materials did not show any enhanced performance with them (e.g., Cheng & Holyoak, 1985; Golding, 1981; Griggs & Cox, 1982). As modulation predicts, improvement occurs with familiar relations (Wason & Shapiro, 1971), but not with arbitrary or unfamiliar relations (Griggs, 1983; Griggs & Cox, 1983; Griggs & Newstead, 1982; Manktelow & Evans, 1979).

Deontic conditionals have often led to improved performance. Thus, as Griggs and Cox (1982) showed, the conditional "If a person is drinking beer then the person must be over 18" tends to elicit the selection of the correct potential violations (the card representing a beer drinker and the card representing an individual less than 18 years old). Cheng and Holyoak (1985) advanced a pragmatic theory in order to explain this phenomenon. They argued that reasoners map conditionals onto *pragmatic reasoning schemas*, such as the following:

> If the precondition is not satisfied (e.g., person is not over 18 years), then the action (e.g., drinking beer) must not be taken.

And the schema, in turn, elicits the correct selection of cards (Cheng & Holyoak, 1985; Kroger, Cheng, & Holyoak, 1993). As the example illustrates, however, pragmatic reasoning schemas contain the modal auxiliaries *may* and *must*, which are systematically ambiguous between what is possible and what is permissible. This ambiguity shows that the schemas are high level rather than foundational. Pragmatic modulation provides an alternative explanation. Reasoners use their general knowledge to construct fully explicit models of the conditional:

Deontic possibilities:	drinking beer	over 18	:Factual possibilities
	¬ drinking beer	over 18	
	¬ drinking beer	¬ over 18	

The complement of these models is a violation of the deontic principle, which general knowledge may provide directly:

Deontic impossibility: drinking beer ¬ over 18 :Factual possibility

Reasoners can use this model to evaluate the cards, selecting only those that are potential violations.

Cosmides (1989) proposed an evolutionary explanation of the deontic selection task. She argued that human evolution has led to a specific inferential module concerned with violations of social contracts (see also Gigerenzer & Hug, 1992). People are thus innately equipped to check for cheaters. Cosmides showed that a background story eliciting the idea of cheating led participants to make a surprising selection. To test a conditional rule of the form If *A* then *C*, they selected instances corresponding to *Not A* and *C*. In the context of the story, the rule "If a man has a tattoo on his face then he eats cassava root" tended to elicit selections of the following cards: no tattoo and eats cassava root. The result is remarkable, but there is no need to invoke an innate inferential module to explain it (see also Griggs, 1984; Holyoak & Cheng, 1995). In the context of the story, the rule signifies that men without the tattoo are not allowed to eat cassava root. Hence, pragmatic modulation yields an *enabling* deontic interpretation (see Table 4):

Factual possibility:	tattoo	eats cassava	:Deontic possibility
	tattoo	¬ eats cassava	
	¬ tattoo	¬ eats cassava	

The violation of the conditional is, accordingly, as follows:

Factual possibility: ¬ tattoo eats cassava :Deontic impossibility

And this model controls the participants' selections.

Certain contents lead to effects of the participants' point of view. For example, Manktelow and Over (1991) used the deontic conditional "If you tidy your room then you may go out to play." The pragmatic modulation of this conditional should lead to a

biconditional interpretation (Fillenbaum, 1977) and hence to the following representation:

Factual possibilities:	tidy	play	:Deontic possibilities
	¬ tidy	¬ play	

There are therefore two sorts of violation:

Factual possibilities:	tidy	¬ play	:Deontic impossibilities
	¬ tidy	play	

It follows that reasoners should select all four cards. The first of these violations, however, is likely to concern the child to whom the conditional applies; the second of them is likely to concern the parent who grants the conditional permission. As Manktelow and Over (1991) showed, participants who were asked to take the point of view of one of the protagonists tended to base their selections on the violation of concern to that individual (see also Manktelow & Over, 1995). Even children are sensitive to point of view as Light, Girotto, and Legrenzi (1990) showed in the first study of the effects of this variable on the selection task (cf. Girotto, Gilly, Blaye, & Light, 1989). As the model theory predicts, however, adults with a neutral point of view do tend to select all four cards (Politzer & Nguyen-Xuan, 1992).

If the model theory is correct, then effects of point of view should occur with conditionals that are not deontic. Consider, for example, the following assertion: "If the Greeks disarmed, then the Turks disarmed," which is likely to yield a biconditional interpretation. As Johnson-Laird and Byrne (1996) argued, reasoners who take the Greek point of view should tend to select cards corresponding to the following counterexample:

Greeks disarmed ¬ Turks disarmed

However, reasoners who take the Turkish point of view should tend to select cards corresponding to the following counterexample:

¬ Greeks disarmed Turks disarmed

Two recent studies have corroborated such effects of point of view with conditionals that are not deontic (see Fairley, Manktelow, & Over, 1999; Staller, Sloman, & Ben-Zeev, 1999).

The model theory's key prediction is that any manipulation that yields explicit models of counterexamples should enhance performance in the selection task. This prediction applies to any materials whatsoever, and it has been corroborated in three separate lines of research. First, as Griggs and his colleagues have shown, instructions to check for violations improved performance with basic conditionals (Chrostowski & Griggs, 1985; Dominowski, 1995; Griggs, 1995; Griggs & Cox, 1983; Platt & Griggs, 1993). Likewise, Green (1995; see also Green & Larking, 1995) showed that instructions to envisage counterexamples to factual conditionals also improved performance. Green, Over, and Pyne (1997) showed that reasoners' assessments of how likely they were to encounter a counterexample (in four stacks of cards) also predicted their selections (see also Green, 1997).

Second, Sperber, Cara, and Girotto (1995) used a more indirect procedure to render counterexamples more *relevant*—in the sense of Sperber and Wilson (1995)—and thereby improved performance. For example, they told their participants that a certain machine generated cards according to the following rule:

If a card has an *A* on one side, then it has a *2* on the other side.

The machine went wrong, but it has been repaired, and the participants now have to check that the job has been done properly. They were thus likely to represent the machine's potential error explicitly:

A ¬ 2

This and other conditionals used in Sperber and Girotto's experiments elicited neither a permission schema (*pace* Cheng & Holyoak, 1985) nor a check for cheaters (*pace* Cosmides, 1989); yet the participants' performance improved significantly.

Third, Love and Kessler (1995) used a content and context that suggested the possibility of counterexamples. For example, they used the conditional rule "If there are Xow then there must be a force field," where the Xow are strange crystallike living organisms who depend for their existence on a force field. In a context that suggested the possibility of counterexamples—mutant Xows who can survive without a force field—the participants carried out the selection task more accurately than in a control condition that did not suggest such counterexamples. Likewise, Liberman and Klar (1996) demonstrated that apparent effects of "checking for cheaters" are better explained in terms of the participants' grasp of appropriate counterexamples and of the relevance of looking for instances of them.

Reasoners can be sensitive to the likelihood of encountering a potential counterexample. This phenomenon has led some theorists to introduce Bayesian considerations into their analysis of the selection task (Kirby, 1994; Nickerson, 1996). Oaksford and Chater (e.g., 1994, 1996) have defended a normative approach inspired by Anderson's (1993) *rational analysis.* They argued that deductive logic is the wrong normative model to apply to the selection task, and they relied instead on an existing normative theory for the selection of data relevant to hypotheses. They proposed that participants rationally seek to maximize the expected gain in information from selecting a card. Nickerson (1996) and others have defended similar ideas. Thus, if a person is testing, say, the conditional "If it is a raven then it is black," it makes sense to look for counterexamples among ravens and black entities in the world, because the frequency of nonblack entities is so much greater than the frequency of black entities.

Are people rational in basing their choices in the selection task on expected gain in information? Not necessarily. Santamaría and Johnson-Laird (2001) paid their participants 1,000 pesetas (about $7) at the start of a selection task using a basic conditional and then charged them 250 pesetas for each card that they chose to select. The participants were told that they would keep whatever money they had not spent, provided that their evaluation of the conditional was correct. The instructions made clear that the basic conditional applied only to the four cards, but the monetary incentive did not improve performance in comparison with an unpaid control group. The participants' selections would not have allowed them to evaluate the conditional correctly. It is hard to justify their performance as rational when it cost them money. Stanovich and his colleagues have similarly shown that participants' Scholastic Achievement Test scores—a measure of higher cognitive ability—correlate with

accuracy in the selection task with basic conditionals (Stanovich, 1999). The more competent individuals made the logically correct selections.

If the Bayesian analyses of the selection task are correct, then manipulating the probabilities of the antecedent, *A*, and the consequent, *C*, of the conditional If A then C should affect performance. The logically correct selections should be more frequent when the probabilities of *A* and *C* are high than when they are low. Indeed, when these two probabilities are high, the *not C* card may be more informative than the *A* card. Such effects do occur (e.g., Kirby, 1994; Oaksford, Chater, & Grainger, 1999), but there have also been failures to detect them (Oberauer, Wilhelm, & Diaz, 1999; Santamaría & Johnson-Laird, 2001). Likewise, when people carried out the selection task with an inclusive disjunction, such as "Every card has a number which is even on one side, or it has a letter which is a vowel on the other side," the participants tended to make the correct selections (Wason & Johnson-Laird, 1969). However, the most frequent error was to select the cards that match the mental models of the disjunction. This error cannot be defended as an attempt to maximize the expected gain of information. A card with an even number, for example, is a true instance of the rule, whatever is on its other side, and so the participants gained no information whatsoever by turning it over. We conclude that naive individuals are not always sensitive to the expected gain in information and that they make genuine logical errors in the selection task.

The crux of the selection task is that content and context can recruit knowledge that modulates the set of models of the conditional. Without such knowledge, reasoners are likely to base their selections on their mental models of conditionals. Modulation, however, can lead them to construct counterexamples to conditionals. It can lead to a biconditional interpretation in which the participants' point of view picks out the potential counterexamples. It can lead to an enabling interpretation. It may even lead to other interpretations (see Table 4), with their concomitant selections.

Illusory Inferences With Conditionals

The model theory entails that certain inferences with conditionals should yield compelling, but invalid, conclusions. These illusory inferences arise from the failure of mental models to represent what is false according to the premises. The illusions have been demonstrated in a variety of domains including quantified, modal, and probabilistic reasoning (see, e.g., Goldvarg & Johnson-Laird, 2000, 2001; Johnson-Laird et al., 2000; Yang & Johnson-Laird, 2000). The most compelling illusions, however, occur with basic conditionals, for example:

> Suppose that the following assertions apply to a specific hand of cards:
>
> If there is a king in the hand then there is an ace in the hand, or else if there is not a king in the hand then there is an ace in the hand.
>
> There is a king in the hand.
>
> What, if anything, follows?

Everyone in an experiment carried out by Johnson-Laird and Savary (1999) drew the conclusion:

> Therefore, there is an ace.

Few expert reasoners resist this conclusion. However, as we explain presently, it is a fallacy, whether the disjunction is interpreted as inclusive or exclusive. The exclusive interpretation can be expressed unequivocally by the rubric used in the following problem:

> Suppose you are playing cards with Billy and you get two clues about the cards in his hand. You know that one of the clues is true and that one of them is false, but unfortunately you don't know which one is true and which one is false:
>
> If there is a king in his hand then there is an ace in his hand.
>
> If there is not a king in his hand then there is an ace in his hand.
>
> Please select the correct answer:
>
> a) There is an ace in Billy's hand.
>
> b) There is not an ace in Billy's hand.
>
> c) There may, or may not, be an ace in Billy's hand.

Most people inferred that there is an ace in Billy's hand. They make analogous errors in estimating the probabilities of cards (Johnson-Laird & Savary, 1996). The model theory predicts that reasoners construct the following mental models of the disjunction of conditionals:

Factual possibilities:	king	ace
	¬ king	ace

From these models, it follows that there is an ace. However, this conclusion is not valid. An exclusive disjunction of the conditionals yields the fully explicit models:

Factual possibilities:	king	¬ ace
	¬ king	¬ ace

Hence, granted a disjunction of the two conditionals, one of them could be false. With an exclusive disjunction, one of them must be false. If, say, the first conditional is false, then there is no guarantee that there is an ace in the hand even when, as in the first problem above, there is definitely a king in the hand. According to the model theory, the illusions arise from the principle of truth. The failure to represent what is false according to the premises leads to erroneous models. Hence, any manipulation that emphasizes falsity should alleviate the illusions. Several studies have corroborated this prediction (Goldvarg & Johnson-Laird, 2000, 2001; Tabossi, Bell, & Johnson-Laird, 1999).

Illusory inferences strike some critics as artificial, but they do occur in everyday life. A search on the World Wide Web for the sequence *or else if* yielded several illusions, including the following warning from a professor to students taking his class:

> Either a grade of zero will be recorded if your absence is not excused, or else if your absence is excused other work you do in the course will count.

The mental models of this assertion yield the two possibilities that presumably the professor and his students had in mind:

¬ excused zero-grade
excused other-work-counts

In other words, both the professor and his students succumbed to an illusion. What the professor should have asserted to convey the two intended possibilities was a conjunction of the conditionals:

> A grade of zero will be recorded if your absence is not excused, but if your absence is excused, then other work you do in the course will count.

Neither the illusions nor their alleviation can be explained by current rule theories. These theories rely solely on valid principles of inference, which cannot account for systematic invalidity.

General Discussion

The model theory of conditionals rests on five assumptions in addition to the original theory of mental models, which includes the principle of truth. These assumptions are as follows:

1. The principle of *core meanings:* The antecedent of a basic conditional describes a possibility, at least in part, and the consequence can occur in this possibility. The core meaning of If A then C is the *conditional* interpretation, and the core meaning of If A then possibly C is the *tautological* interpretation.

2. The principle of *subjunctive meanings:* A subjunctive conditional refers to the same set of possibilities as the corresponding indicative conditional, but the set consists either of factual possibilities or of a fact in which the antecedent and consequent did not occur and counterfactual possibilities in which they did occur.

3. The principle of *implicit models:* Basic conditionals have mental models representing the possibilities in which their antecedents are satisfied, but only implicit mental models for the possibilities in which their antecedents are not satisfied. A mental "footnote" on the implicit model can be used to make fully explicit models, but individuals are liable to forget the footnote, and even to forget the implicit model itself for complex compound assertions.

4. The principle of *semantic modulation:* The meanings of the antecedent and consequent, and coreferential links between these two clauses, can add information to models, prevent the construction of otherwise feasible models, and aid the process of constructing fully explicit models.

5. The principle of *pragmatic modulation:* The context of a conditional depends on general knowledge in long-term memory and knowledge of the specific circumstances of its utterance. This context is normally represented in explicit models. These models can modulate the mental models of a conditional, taking precedence over contradictory models, and they can add information to models, prevent the construction of otherwise feasible models, and aid the process of constructing fully explicit models.

What is the real nature of conditionals and conditional reasoning? Readers may have the impression that for some unknown reason conditionals have many meanings and that they accordingly give rise to many diverse results in reasoning. Readers may even wonder why, given such diverse interpretations, languages contain conditionals. In our view, conditionals have core meanings, and the varied phenomena arise from their mental models. Speakers think of a possibility and assert that something holds or may hold in that context. The assertion is accordingly a conditional. Listeners understand the conditional by envisaging the antecedent possibility and that the consequent holds or may hold in it (cf. Ramsey, 1929/1990). Both parties appreciate that there are other possibilities, but they do not normally bother to envisage them. The result is a rudimentary representation—a set of mental models in which not all possibilities are represented explicitly. The focus on possibilities that satisfy antecedents gives rise to the apparent paradoxes that occur when individuals are forced to think about truth instead of possibilities. However, comparable paradoxes occur with disjunctions. This same focus also gives rise to systematic misinterpretations. Individuals overlook what is false, particularly the possibilities when the antecedent is false. As Experiments 1 and 2 showed, this oversight yields the same interpretation for both conjunctions of conditionals and disjunctions of conditionals. The conjunctive-like interpretation of conditionals is akin to a regression to a child's interpretation.

The meanings of antecedents and consequents and their referential relations can modulate the core meanings of conditionals. Likewise, general knowledge and knowledge of context—represented in explicit models of what is possible—can also modulate the core meanings. One consequence is a semantic relation between the antecedent and the consequent, such as temporal or spatial relation between them. However, modulation can also prevent the construction of models in the sets corresponding to core interpretations—a process that yields 10 distinct sets of possibilities (see Table 4). Experiment 3 corroborated the occurrence of such modulations. They are also likely to occur with other sentential connectives. The following inference, for example, is valid in form:

> Eva's in Rio or she's in Brazil.
> She's not in Brazil.
> Therefore, she's in Rio.

However, no sensible person other than a logician is likely to draw this conclusion. It is impossible for Eva to be in Rio and not in Brazil, because Rio is in Brazil.

If a connective such as a conditional is truth functional, then it takes two truth values as input, one for the antecedent and one for the consequent, and it delivers as output a truth value that depends solely on these input truth values. In other words, whether a conditional is true or false depends not on the particular contents of its antecedent and consequent but only on their truth values. This principle is unworkable for conditionals in natural language. If the interpretative system has access only to the truth values of the antecedent and consequent, then it is unable to take into account either temporal or spatial relations or to determine which of the 10 different sets of possibilities is applicable. Conditionals are not truth functional. Nor, in our view, are any other sentential connectives in natural language.

Because individuals focus on the antecedent possibilities of conditionals, certain inferences are difficult. Thus, modus tollens (If A then B: Not B, therefore, Not A) can be difficult both in life and in the laboratory. Similarly, when individuals seek evidence in the selection task to test the truth of a hypothesis, they tend not to use the falsifying case of a basic conditional. And, just as the focus on truth leads to confusion between disjunctions and conjunctions of conditionals, it also underlies compelling illusory inferences, such as the following one:

One of these assertions is true and one of them is false:
If there is a king in the hand then there is an ace.
If there isn't a king in the hand then there is an ace.
Therefore, there is an ace in the hand.

Such errors arise according to the model theory because individuals neglect what is false.

The theory makes a key prediction that any manipulation that reduces the focus on antecedent possibilities should improve performance with the difficult inferences. This prediction has been confirmed for many sorts of conditional reasoning. For example, the difficulty of modus tollens is reduced with biconditionals of the form A if and only if C because *only if* tends to elicit a representation of the possibility in which neither the antecedent nor the consequent is satisfied:

Factual possibilities:	a	c
	¬ a	¬ c

Similar phenomena occur with counterfactual conditionals of the form If A had happened then C would have happened because the counterfactual interpretation also represents the facts of the matter:

Fact:	¬ a	¬ c
Counterfactual possibilities:	a	c

Content and context can modulate the representation of conditionals, affecting the antecedent's necessity or sufficiency for the consequent. These effects, in turn, enhance or suppress inferences. Modus ponens is suppressed in contexts that make clear that the antecedent provides only a partial condition for the consequent to occur or leads to it only with a certain probability. As Experiment 4 showed, modulation also affects modus tollens. The inference was enhanced by knowledge of a spatial inclusion and the use of familiar relations; it was suppressed by knowledge of spatial exclusion and the use of unfamiliar relations. Similarly, in the selection task, modulation can increase the likelihood that reasoners represent counterexamples and, in consequence, make correct selections. In deontic domains, the premises often elicit a biconditional interpretation and, accordingly, have two distinct sorts of counterexample. The participants' point of view determines which of these counterexamples they use to control their selections. The same effects occur in domains that are not deontic, contrary to other theories of the selection task.

What are the alternative accounts of conditionals? By far the most important of them are those based on formal rules. The late Martin Braine and his colleagues (see, e.g., Brain & O'Brien, 1991) proposed that the meaning of a conditional is equivalent to two rules of inference, one for modus ponens and another for conditional proof (see also Ryle, 1949, chapter 5, for a precursor to this idea). Rips (1994) proposed a similar formal system that he implemented computationally, though the full program is not in the public domain. Rule theories and the model theory concur that individuals can reason from suppositions (see, e.g., Byrne et al., 1995). However, our evidence is otherwise incompatible with rule theories. They cannot explain why individuals list the same possibilities for conjunctions and disjunctions of conditionals (Experiments 1 and 2). They cannot explain the modulation of core meanings into 10 distinct sets of possibilities (including those corroborated in Experiment 3). They cannot explain effects of both semantic and pragmatic modulation on conditional inferences, such as the enhancement and inhibition of modus tollens (Experiment 4). They cannot explain the comparable effects of content on the selection task. They cannot explain the developmental sequence of interpretations (conjunction, biconditional, conditional). And they cannot explain the occurrence of systematic errors in reasoning, such as the illusory inferences based on conditionals. The model theory predicts all of these phenomena.

Conditionals have an indefinite number of meanings—10 sets of possibilities and a variety of relations between antecedent and consequent. This diversity is inimical to formal rules. Yet, advanced thinkers who reflect on their own reasoning may construct formal rules for themselves for the core meanings of basic conditionals. This process leads ultimately to the discipline of formal logic. Rules and models are not incompatible for core meanings. However, the meaning of a term is not the same as the rules of inference for it (Johnson-Laird, 1983, p. 41; Osherson, 1974–1976, Vol. 3, p. 253; Prior, 1960). Rules of inference enable reasoners to pass from premises to conclusions; meanings relate assertions to possibilities in the world.

Ultimately, the complexity of conditionals has simple causes. Their chameleon-like characteristics arise from interactions among a set of elementary components: Their core meanings referring to possibilities, their representation in mental models, and their semantic and pragmatic modulation.

References

Adams, E. W. (1970). Subjunctive and indicative conditionals. *Foundations of Language, 6,* 89–94.

Adams, E. W. (1975). *The logic of conditionals: An application of probability to deductive logic.* Dordrecht, the Netherlands: Reidel.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Baron, J. (1994). *Thinking and deciding* (2nd ed.). Cambridge, England: Cambridge University Press.

Barres, P. E., & Johnson-Laird, P. N. (1997). Why is it hard to imagine what is false? In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (p. 859). Mahwah, NJ: Erlbaum.

Barrouillet, P. (1997). Modifying the representation of if . . . then sentences in adolescents by inducing a structure mapping strategy. *Current Psychology of Cognition, 16,* 609–637.

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

Barrouillet, P., & Lecas, J.-F. (1998). How can mental models theory account for content effects in conditional reasoning? A developmental perspective. *Cognition, 67,* 209–253.

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

Barwise, J. (1986). Conditionals and conditional information. In E. C. Traugott, A. ter Meulen, J. S. Reilly, & C. A. Ferguson (Eds.), *On conditionals* (pp. 21–54). Cambridge, England: Cambridge University Press.

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.

Bonatti, L. (1994a). Propositional reasoning by model? *Psychology Review, 101,* 725–733.

Bonatti, L. (1994b). Why should we abandon the mental logic hypothesis? *Cognition, 50,* 17–39.

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

Braine, M. D. S., & O'Brien, D. P. (1991). A theory of If: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98,* 182–203.

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

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

Brewka, G., Dix, J., & Konolige, K. (1997). *Nonmonotonic reasoning: An overview*. Stanford, CA: Center for Language Science and Information Publications, Stanford University.

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

Byrne, R. M. J. (1991). Can valid inferences be suppressed? *Cognition, 39,* 71–78.

Byrne, R. M. J. (1997). Cognitive processes in counterfactual thinking about what might have been. In D. K. Medin (Ed.), *The psychology of learning and motivation, advances in research and theory* (Vol. 37, pp. 105–154). San Diego, CA: Academic Press.

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. (1992). Reasoning strategies. *Irish Journal of Psychology, 13,* 111–124.

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

Byrne, R. M. J., Handley, S. J., & Johnson-Laird, P. N. (1995). Reasoning from suppositions. *Quarterly Journal of Experimental Psychology, 48A,* 915–944.

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., Segura, S., Culhane, R., Tasso, A., & Berrocal, P. (2000). The temporality effect in counterfactual thinking about what might have been. *Memory & Cognition, 28,* 264–281.

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

Case, R. (1985). *Intellectual development: Birth to adulthood.* New York: Academic Press.

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

Chrostowksi, J. J., & Griggs, R. A. (1985). The effects of problem content, instructions and verbalization procedure on Wason's selection task. *Current Psychological Research and Reviews, 4,* 99–107.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition, 31,* 187–276.

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

Delval, J. A., & Riviere, A. (1975). "Si llueve, Elisa lleva sombrero": Una investigación psicologia sobre la tabla de verdad del condicional. [A psychological investigation of the conditional truth table]. *Revista De Psicologia General y Aplicada, 136,* 825–850.

Dominowski, R. L. (1995). Content effects in Wason's selection task. In S. E. Newstead & J. S. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 41–65). Hillsdale, NJ: Erlbaum.

Dudman, V. H. (1988). Indicative and subjunctive conditionals. *Analysis, 48,* 113–122.

Dugan, C. M., & Revlin, R. (1990). Response options and presentation format as contributors to conditional reasoning. *Quarterly Journal of Experimental Psychology, 42A,* 829–848.

Evans, G. (1980). Pronouns. *Linguistic Inquiry, 11,* 337–362.

Evans, J. S. B. T. (1972). Interpretation and "matching bias" in a reasoning task. *Quarterly Journal of Experimental Psychology, 24,* 193–199.

Evans, J. S. B. T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology, 29,* 297–306.

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

Evans, J. S. B. T. (1989). *Bias in human reasoning: Causes and consequences.* Hillsdale, NJ: Erlbaum.

Evans, J. S. B. T. (1993). The mental model of conditional reasoning: Critical appraisal and revision. *Cognition, 48,* 1–20.

Evans, J. S. B. T., & Beck, M. A. (1981). Directionality and temporal factors in conditional reasoning. *Current Psychological Research, 1,* 111–120.

Evans, J. S. B. T., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory and Language, 35,* 392–404.

Evans, J. S. B. T., & Handley, S. J. (1999). The role of negation in conditional inference. *Quarterly Journal of Experimental Psychology, 52A,* 739–769.

Evans, J. S. B. T., Legrenzi, P., & Girotto, V. (1999). The influence of linguistic form on reasoning: The case of matching bias. *Quarterly Journal of Experimental Psychology, 52A,* 185–214.

Evans, J. S. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64,* 391–397.

Evans, J. S. B. T., & Newstead, S. E. (1977). Language and reasoning: A study of temporal factors. *Cognition, 8,* 265–283.

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

Fairley, N., Manktelow, K. I., & Over, D. E. (1999). Necessity, sufficiency and perspective effects in causal conditional reasoning. *Quarterly Journal of Experimental Psychology, 52A,* 771–790.

Falmagne, R. J., & Gonsalves, J. (1995). Deductive inference. *Annual Review of Psychology, 46,* 525–559.

Fillenbaum, S. (1977). Mind your p's and q's: The role of content and context in some uses of and, or, and if. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 11, pp. 41–100). New York: Academic Press.

Fillenbaum, S. (1993). Deductive reasoning: What are taken to be the premises and how are they interpreted? *Behavioral and Brain Sciences, 16,* 348–349.

Fodor, J. A., Garrett, M. F., Walker, E. C. T., & Parkes, C. H. (1980). Against definitions. *Cognition, 8,* 263–367.

Garnham, A., & Oakhill, J. V. (1994). *Thinking and reasoning.* Oxford, England: Basil Blackwell.

Gazdar, G. (1979). *Pragmatics: Implications, presupposition and logical form.* New York: Academic Press.

Geach, P. (1962). *Reference and generality*. Ithaca, NY: Cornell University Press.

Geis, M. C., & Zwicky, A. M. (1971). On invited inferences. *Linguistic Inquiry, 2,* 561–566.

George, C. (1999). Evaluation of the plausibility of a conclusion derivable from several arguments with uncertain premises. *Thinking & Reasoning, 5,* 245–281.

Gigerenzer, G., & Hug, K. (1992). Domain specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43,* 127–171.

Girotto, V., Gilly, M., Blaye, A., & Light, P. (1989). Children's performance in the selection task: Plausibility and familiarity. *British Journal of Psychology, 80,* 79–95.

Girotto, V., & Johnson-Laird, P. N. (2002). [Results on the probability of conditionals and conditional probabilities]. Unpublished raw data.

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.

Golding, E. (1981). *The effect of past experience on problem solving.* Paper presented at the meeting of the British Psychological Society, Guildford, England.

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. (1995). Externalization, counter-examples and the abstract selection task. *Quarterly Journal of Experimental Psychology, 48A,* 424–446.

Green, D. W. (1997). Hypothetical thinking in the selection task: Amplifying a model-based approach. *Current Psychology of Cognition, 16,* 93–102.

Green, D. W., & Larking, R. (1995). The locus of facilitation in the abstract selection task. *Thinking & Reasoning, 1,* 183–199.

Green, D. W., Over, D. E., & Pyne, R. (1997). Probability and choice in the selection task. *Thinking & Reasoning, 3,* 209–235.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–48). New York: Academic Press.

Griggs, R. A. (1983). The role of problem content in the selection task and in the THOG problem. In J. S. B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches* (pp. 16–43). London: Routledge & Kegan Paul.

Griggs, R. A. (1984). Memory cueing and instructional effects on Wason's selectional task. *Current Psychological Research and Reviews, 3,* 3–10.

Griggs, R. A. (1995). The effects of rule clarification, decision justification, and selection instruction on Wason's abstract selection task. In S. E. Newstead & J. S. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 17–39). Hillsdale, NJ: Erlbaum.

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

Griggs, R. A., & Cox, J. R. (1983). The effects of problem content and negation on Wason's selection task. *Quarterly Journal of Experimental Psychology, 35A,* 519–533.

Griggs, R. A., & Jackson, S. L. (1990). Instructional effects on responses in Wason's selection task. *British Journal of Psychology, 81,* 197–204.

Griggs, R. A., & Newstead, S. E. (1982). The role of problem structure in a deductive reasoning task. *Journal of Experimental Psychology: Language, Memory, and Cognition, 8,* 297–307.

Harper, W. L., Stalnaker, R., & Pearce, G. (Eds.). (1981). *Ifs: Conditionals, belief, decision, chance, and time.* Dordrecht, the Netherlands: Reidel.

Harr, J. (1995). *A civil action.* New York: Random House.

Holyoak, K. J., & Cheng, P. (1995). Pragmatic reasoning with a point of view: A response. *Thinking & Reasoning, 1,* 289–313, 373–388.

Jackson, F. (1987). *Conditionals.* Oxford, England: Basil Blackwell.

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

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

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

Johnson-Laird, P. N. (1986). Conditionals and mental models. In E. C. Traugott, A. ter Meulen, J. S. Reilly, & C. A. Ferguson (Eds.), *On conditionals* (pp. 55–75). Cambridge, England: Cambridge University Press.

Johnson-Laird, P. N. (1990). The development of reasoning. In P. Bryant & G. Butterworth (Eds.), *Causes of development* (pp. 121–131). Hemel Hempstead, Hertfordshire, England: Harvester-Wheatsheaf.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (1996). A model point of view: A comment on Holyoak and Cheng. *Thinking & Reasoning, 1,* 339–350.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. S. (1992). Propositional reasoning by model. *Psychological Review, 99,* 418–439.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. (2000, April 21). 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., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63,* 395–400.

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., Savary, F., & Bucciarelli, M. (1999). Strategies and tactics in reasoning. In W. S. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 209–240). Mahwah, NJ: Erlbaum.

Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology, 82,* 367–373.

Kahneman, D., & Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Karttunen, L. (1972). Possible and must. In J. P. Kimball (Ed.), *Syntax and semantics* (Vol. 1, pp. 1–20). New York: Seminar Press.

Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition, 51,* 1–28.

Kratzer, A. (1989). An investigation of the lumps of thought. *Linguistics and Philosophy, 12,* 607–653.

Kripke, S. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16,* 83–94.

Kroger, J. K., Cheng, P. W., & Holyoak, K. J. (1993). Evoking the permission schema: The impact of explicit negation and a violation-checking context. *Quarterly Journal of Experimental Psychology, 46A,* 615–635.

Kuhn, D. (1977). Conditional reasoning in children. *Developmental Psychology, 13,* 342–353.

Langford, P. E. (1992). Evaluation strategies for some nonstandard conditionals during adolescence. *Psychological Reports, 70,* 643–664.

Lecas, J.-F., & Barrouillet, P. (1999). Understanding of conditional rules in childhood and adolescence: A mental models approach. *Current Psychology of Cognition, 18,* 363–396.

Legrenzi, M. S., Girotto, V., Legrenzi, P., & Johnson-Laird, P. N. (2000). *Reasoning to consistency: A theory of naïve nonmonotonic reasoning.* Manuscript in preparation.

Legrenzi, P. (1970). Relations between language and reasoning about deductive rules. In G. B. Flores D'Arcais, & W. J. M. Levelt (Eds.), *Advances in psycholinguistics* (pp. 322–333). Amsterdam: North-Holland.

Levinson, S. (1983). *Pragmatics.* Cambridge, England: Cambridge University Press.

Lewis, D. (1973). *Counterfactuals.* Cambridge, MA: Harvard University Press.

Lewis, D. K. (1976). Probabilities of conditionals and conditional probabilities, II. *Philosophical Review, 95,* 581–589.

Liberman, N., & Klar, Y. (1996). Hypothesis testing in Wason's selection task: Social exchange cheating detection or task understanding. *Cognition, 58,* 127–156.

Light, P. H., Girotto, V., & Legrenzi, P. (1990). Children's reasoning on

conditional promises and permissions. *Cognitive Development, 5,* 369–383.

Love, R., & Kessler, C. (1995). Focussing in Wason's selection task: Content and instruction effects. *Thinking & Reasoning, 1,* 153–182.

Lycan, W. G. (1991). *MPP, RIP*. Unpublished manuscript, University of North Carolina.

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

Manktelow, K. I., & Evans, J. S. B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology, 70,* 477–488.

Manktelow, K. I., & Over, D. E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition, 39,* 85–105.

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

Margolis, L. (1987). *Patterns, thinking and cognition: A theory of judgement.* Chicago: University of Chicago Press.

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

Markovits, H. (1993). The development of conditional reasoning: A Piagetian reformulation of mental models theory. *Merrill-Palmer Quarterly, 39,* 133–160.

Markovits, H., & Savary, F. (1992). Pragmatic schemas and the selection task: To reason or not to reason. *Quarterly Journal of Experimental Psychology, 45A,* 133–148.

Markovits, H., & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. *Developmental Psychology, 26,* 942–951.

McGee, V. (1985). A counterexample to modus ponens. *Journal of Philosophy, 82,* 462–471.

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

Minsky, M. L. (1975). Frame-system theory. In R. C. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing.* Preprints of a conference at Massachusetts Institute of Technology, Cambridge, MA.

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

Nickerson, R. (1996). Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking & Reasoning, 2,* 1–31.

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

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review, 103,* 381–391.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking & Reasoning, 5,* 193–243.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 835–854.

Oberauer, K., Wilhelm, O., & Diaz, R. R. (1999). Bayesian rationality for the Wason selection task? A test of optimal data selection theory. *Thinking & Reasoning, 5,* 115–144.

O'Brien, D. P. (1987). The development of conditional reasoning: An iffy proposition. In H. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 61–90). New York: Academic Press.

O'Brien, D. P. (1999). If is neither *and* nor material implication. Commentary on Lecas & Barrouillet (1999). *Current Psychology of Cognition, 18,* 397–407.

O'Brien, D. P., Dias, M. G., & Roazzi, A. (1998). A case study in the mental models and mental-logic debate: Conditional syllogisms. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic* (pp. 385–420). Mahwah, NJ: Erlbaum.

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, 46A,* 653–678.

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

Over, D. E., & Evans, J. S. B. T. (1997). Two cheers for deductive competence. *Current Psychology of Cognition, 16,* 225–278.

Paris, S. G. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology, 16,* 278–291.

Platt, R. D., & Griggs, R. A. (1993). Facilitation in the abstract selection task: The effects of attentional and instructional factors. *Quarterly Journal of Experimental Psychology, 46A,* 591–613.

Politzer, G. (1986). Laws of language use and of formal logic. *Journal of Psycholinguistic Research, 15,* 47–92.

Politzer, G., & Braine, M. D. S. (1991). Responses to inconsistent premises cannot count as suppression of valid inferences. *Cognition, 38,* 103–108.

Politzer, G., & Nguyen-Xuan, A. (1992). Reasoning about conditional promises and warnings: Darwinian algorithms, mental models, relevance judgements or pragmatic schemas? *Quarterly Journal of Experimental Psychology, 44,* 401–412.

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

Prior, A. N. (1960). The runabout inference-ticket. *Analysis, 21,* 38–39.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 216–270). Cambridge, MA: MIT Press.

Quine, W. V. O. (1952). *Methods of logic.* London: Routledge.

Ramsey, F. P. (1990). General propositions and causality. In D. H. Mellor (Ed.), *Foundations: Essays in philosophy, logic, mathematics and economics* (pp. 145–163). London: Humanities Press. (Original work published 1929)

Reinhart, T. (1986). On the interpretation of 'donkey'-sentences. In E. C. Traugott, A. ter Meulen, J. S. Reilly, & C. A. Ferguson (Eds.), *On conditionals* (pp. 103–122). Cambridge, England: Cambridge University Press.

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

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

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

Roberge, J. J. (1971). Some effects of negation on adults' conditional reasoning abilities. *Psychological Reports, 29,* 839–844.

Roberge, J. J. (1974). Effects of negation on adults' comprehension of fallacious conditional and disjunctive arguments. *Journal of General Psychology, 91,* 287–293.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the micro-structure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press.

Russell, J. (1987). Rule-following, mental models, and the developmental view. In M. Chapman & R. A. Dixon (Eds.), *Meaning and the growth of understanding: Wittgenstein's significance for developmental psychology.* New York: Springer.

Ryle, G. (1949). *The concept of mind.* London: Hutchinson.

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. (2001). *Are people rational in the abstract selection task?* Manuscript in preparation.

Schroyens, W., Schaeken, W., & d'Ydewalle, G. (2000). *Conditional reasoning by model and/or rule: A meta-analytic review of the theories and the data.* Manuscript submitted for publication.

Schroyens, W., Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2000). Conditional reasoning with negations: Implicit and explicit affirmation or denial and the role of contrast classes. *Thinking & Reasoning, 6,* 221–251.

Simon, H. A. (1982). *Models of bounded rationality* (Vols. 1 and 2). Cambridge, MA: MIT Press.

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

Sloutsky, V. M., & Morris, B. J. (1999). *How to make something out of nothing: Adaptive constraints on children's information processing.* Manuscript submitted for publication.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 52,* 3–39.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (Rev. ed.). Oxford, England: Blackwell.

Staller, A., Sloman, S. A., & Ben-Zeev, T. (1999). Perspective effects in non-deontic versions of the Wason selection task. *Memory & Cognition, 28,* 396–405.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (American Philosophical Quarterly Monograph No. 2). Oxford, England: Blackwell.

Stalnaker, R. C. (1970). Probability and conditionals. *Philosophy of Science, 37,* 64–80.

Stalnaker, R. C. (1975). Indicative conditionals. *Philosophia, 5,* 269–286.

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

Staudenmayer, H. (1975). Understanding conditional reasoning with meaningful propositions. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults* (pp. 55–79). Hillsdale, NJ: Erlbaum.

Staudenmayer, H., & Bourne, L. E. (1978). The nature of denied propositions in the conditional reasoning task: Interpretation and learning. In R. Revlin & R. E. Mayer (Eds.), *Human reasoning* (pp. 83–99). New York: Wiley.

Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology, 48A,* 613–643.

Tabossi, P., Bell, V. A., & Johnson-Laird, P. N. (1999). Mental models in deductive, modal, and probabilistic reasoning. In C. Habel & G. Rickheit (Eds.), *Mental models in discourse processing and reasoning* (pp. 299–331). Berlin, Germany: John Benjamins.

Taplin, J. E., Staudenmayer, H., & Taddonio, J. L. (1974). Developmental changes in conditional reasoning: Linguistic or logical? *Journal of Experimental Child Psychology, 17,* 360–373.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition, 22,* 742–758.

Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology, 49,* 1–60.

Thompson, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition, 76,* 209–268.

Thompson, V. A., & Byrne, R. M. J. (2000). *Reasoning counterfactually: Making inferences about things that didn't happen.* Manuscript submitted for publication.

Traugott, E. C., ter Meulen, A., Reilly, J. S., & Ferguson, C. A. (Eds.). (1986). *On conditionals.* Cambridge, England: Cambridge University Press.

Veltman, F. (1986). Data semantics and the pragmatics of indicative conditionals. In E. C. Traugott, A. ter Meulen, J. S. Reilly, & C. A. Ferguson (Eds.), *On conditionals* (pp. 147–168). Cambridge, England: Cambridge University Press.

Wason, P. C. (1964). The effect of self-contradiction on fallacious reasoning. *Quarterly Journal of Experimental Psychology, 20,* 273–281.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology.* Harmonsworth, Middlesex, England: Penguin Books.

Wason, P. C., & Johnson-Laird, P. N. (1969). Proving a disjunctive rule. *Quarterly Journal of Experimental Psychology, 21,* 14–20.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23,* 63–71.

Wildman, T. M., & Fletcher, H. J. (1977). Developmental increases and decreases in solutions of conditional syllogism problems. *Developmental Psychology, 13,* 630–636.

Wilson, D. (1975). *Presuppositions and non-truth-conditional semantics.* London: Academic Press.

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

Received April 5, 2000
Revision received July 20, 2001
Accepted July 31, 2001 ■