COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2018) 1–38

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12634

Facts and Possibilities: A Model-Based Theory of Sentential Reasoning

Sangeet S. Khemlani,[a] Ruth M. J. Byrne,[b] Philip N. Johnson-Laird[c,d]

[a]*Navy Center for Applied Research in Artificial Intelligence, US Naval Research Laboratory*
[b]*School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin*
[c]*Department of Psychology, Princeton University*
[d]*Department of Psychology, New York University*

Received 8 April 2017; received in revised form 17 April 2018; accepted 3 May 2018

Abstract

This article presents a fundamental advance in the theory of mental models as an explanation of reasoning about facts, possibilities, and probabilities. It postulates that the meanings of compound assertions, such as conditionals (*if*) and disjunctions (*or*), unlike those in logic, refer to conjunctions of epistemic possibilities that hold in default of information to the contrary. Various factors such as general knowledge can modulate these interpretations. New information can always override sentential inferences; that is, reasoning in daily life is defeasible (or nonmonotonic). The theory is a dual process one: It distinguishes between intuitive inferences (based on system 1) and deliberative inferences (based on system 2). The article describes a computer implementation of the theory, including its two systems of reasoning, and it shows how the program simulates crucial predictions that evidence corroborates. It concludes with a discussion of how the theory contrasts with those based on logic or on probabilities.

Keywords: Deduction; Logic; Mental models; Nonmonotonicity; Reasoning; Possibility

1. Introduction

People reason about facts, possibilities, and probabilities. Psychologists have carried out many studies of factual inferences, such as:

1. If the card is an ace then it is a heart.
   The card is an ace.
   Therefore, the card is a heart.

Correspondence should be sent to Sangeet Khemlani, Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory, 4555 Overlook Drive, Washington, DC 20375. E-mail: skhemlani@gmail.com

The conclusion is valid: It holds in all the cases in which the premises are true (Jeffrey, 1981, p. 1). Validity therefore depends on meaning and truth. Since different logics assign different meanings to certain terms, such as "possibly" (see, e.g., Hughes & Cresswell, 1996), what is valid in one logic can be invalid in another. Words such as "if," "or," and "and" are sentential connectives, because they can form "compound" sentences by connecting clauses. Reasoning from compounds is known as "sentential reasoning" after its counterpart of sentential logic. Recent theories treat sentential inferences as a special case of reasoning about probabilities (e.g., Pfeifer, 2013; Over, 2009; for a review, see Johnson-Laird, Khemlani, & Goodwin, 2015a,b; and for a reply, Baratgin et al., 2015). So the previous inference is said to be akin to this one:

2. Probably, if the card is an ace then it is a heart.
   The card is an ace.
   Therefore, probably the card is a heart.

Almost all psychological theories, however, neglect a more rudimentary form of uncertainty—inferences about possibilities. And possibilities diverge from probabilities. Reasoners draw conclusions about possibilities from premises that make no mention of them, for example:

3. The card is an ace or it is a heart, or both.
   Therefore, it is possible that the card is an ace.

They are unlikely to infer from the preceding premise that it is probable that the card is an ace. The inference in (3) is intuitive, and individuals tend to accept it. But, as we will explain, it is invalid in all logics. No psychological theory other than the one we describe below accounts for it either, and other theories are silent about how reasoning integrates facts, possibilities, and probabilities.

Our aim is to show how an advance in the theory of mental models explains inferences such as (3). This new theory treats possibilities as fundamental to everyday reasoning. They are the topic of "modal" logic (e.g., Hughes & Cresswell, 1996) and of psychological investigations (e.g., Bell & Johnson-Laird, 1998; Byrnes & Beilin, 1991; Goldvarg & Johnson-Laird, 2000; Inhelder & Piaget, 1958; Piéraut-Le Bonniec, 1980; Sophian & Somerville, 1988). Yet no comprehensive theory of modal reasoning exists. One reason is the ambiguity of the concept of possibility. Philosophers distinguish the "alethic" modality in which a possibility is anything that is not self-contradictory, for example, it is logically possible that the moon was made of cheese; the "epistemic" modality in which it is anything consistent with knowledge, for example, it is not possible that the moon was made of cheese; and the "deontic" modality in which it is anything that is permissible, for example, it is possible to smoke outside London pubs (e.g., Bucciarelli & Johnson-Laird, 2005; Bucciarelli, Khemlani, & Johnson-Laird, 2008).

The theory we describe—the "model" theory, for short—concerns reasoning about epistemic possibilities. And it is a radical departure from those based on logic or on

probability. It rejects logical formulations of the meanings of sentential connectives such as "if" and "or" (pace Braine & O'Brien, 1998; Rips, 1994). Conditionals of the sort, *if A then B*, for instance, do not correspond to the so-called material conditional of logic or to conditional probabilities (see also Byrne & Johnson-Laird, 2009; Schroyens, 2010). Probabilistic approaches to reasoning postulate that the probability calculus should replace logic, especially for conditionals, and that what matters is probabilistic validity (Adams, 1998)—*p*-validity, for short—which holds whenever a conclusion is no less probable than its premises given any consistent assignment of probabilities (e.g., Chater & Oaksford, 2009; Cruz, Baratgin, Oaksford, & Over, 2015; Evans, 2012; Oaksford & Chater, 2007; Over, 2009; Pfeifer, 2013; Pfeifer & Kleiter, 2009). A consequence of the model theory, however, is that probabilities are not part of the fundamental meanings of compound assertions. Probabilities enter into everyday reasoning only if they are invoked by contents or tasks (see Goodwin, 2014, for corroboration of this claim).

The present paper begins with an outline of the new theory of sentential reasoning, focusing on facts and possibilities, and then describes its implementation in a computer program, showing how the program also generates extensional probabilities (for the theory's account of the probabilities of unique events, see Khemlani, Lotstein, & Johnson-Laird, 2012, 2015). It reviews evidence for the theory in passing, and it shows how the program's outputs fit experimental results. Finally, it discusses the primary consequences of the theory, potential objections to it, and its principal alternatives.

2. The model theory

The original model theory is due to Craik (1943), who argued that the mind constructs small-scale models of the world in order to make decisions. But he argued that reasoning depends on verbal rules. So the modern model theory began with the assumption that reasoning depends on models too (Johnson-Laird, 1975). It has a long-standing account of what the mind computes in deduction: It aims to infer conclusions that are new, are parsimonious, and maintain the semantic information in the premises, which guarantees validity (Johnson-Laird & Byrne, 1991, Ch. 2). Inferences that meet these constraints are rational deductions. When no conclusion meets them, nothing follows—an evaluation contrary to orthodox logic, which allows infinitely many different conclusions to follow from any set of premises.

The theory also explains how the computations are carried out, and its fundamental idea is that reasoning uses the meanings of assertions to build models that simulate the situations that the premises describe, and then draws conclusions from these models. Models are iconic in that they have a structure corresponding, insofar as possible, to the structure of what they represent. Visual images are iconic, but they cannot represent abstract entities; mental models underlie images and can represent abstract entities. Likewise, models of a process can be kinematic that is, they can simulate the separate steps of a sequence of events (Khemlani, Mackiewicz, Bucciarelli, & Johnson-Laird, 2013). We now turn to the key advances of the new theory, and we explain how they build on

the theory's previous version of sentential reasoning (summarized in Johnson-Laird & Byrne, 2002).

2.1. Sentential connectives refer to conjunctions of default possibilities

In logic, the meaning of sentential connectives is "truth functional", that is, the semantics of each connective yields a truth value of the compound in which it occurs as a function only of the truth or falsity of the clauses it connects. For readers unfamiliar with this idea, Khemlani (2018, table 11.2) presents the truth-functional treatment of various compounds. But this treatment of conditionals yields bizarre inferences in daily life. Such conditionals are known as "material conditionals," and they are true in every case except when their *if*-clause is true but their *then*-clause is false. A conditional whose *if*-clause is false is therefore true. This consequence yields the following "paradox":

4. Pam is well.
   Therefore, if Pam is not well then she has the flu.

Most psychologists therefore reject the material conditional as the basis of everyday conditionals (e.g., Evans & Over, 2004; Johnson-Laird & Byrne, 2002; Oaksford & Chater, 2007).

The previous version of the model theory postulated that people understand compound assertions by constructing models representing the disjunctive alternatives to which they refer. The new theory postulates instead that compounds refer to conjunctions of possibilities that each hold in default of information to the contrary—we refer to them as "default" possibilities. The conjunction of a factual claim, *A*, and its negation:

A & not A

is contradictory. But a conjunction of the corresponding possibilities, where "possible" is used in its everyday sense rather than in a sense from modal logic:

A is possible & not A is possible

is consistent in daily life, and so compound assertions refer to conjunctions of default possibilities. The meaning of these possibilities is epistemic; that is, "it is possible that A" is akin to saying, "as far individuals know, it is possible that A." People build finite models of possibilities, and finite models are likewise a way of embodying modal logic in artificial intelligence (cf. Meyer & van der Hoek, 1995). But, as we show below, the present theory is not compatible with modal logic.

A conditional, *if A then B*, refers to the following exhaustive conjunction of default possibilities:

possible(A & B)
& possible(not A & B)
& possible(not A & not B).

So, each of these cases is possible. The fourth case, *A & not B*, is impossible by default, because its impossibility can be inferred from the exhaustive conjunction. The four cases, which are the "partition" of the assertion, occur in the truth table for a material conditional (see Table 2, column 3, below), which captures its meaning in logic. The truth table for the material conditional yields *true* for just those cases that are possible. So it is easy to overlook a crucial difference. Entries in a truth table are mutually exclusive alternatives: Their conjunction is self-contradictory because *A & B* is not consistent with *not-A & B*, and so on. In contrast, the three possibilities above and the impossibility that they imply can all occur in the same conjunction without contradiction in daily life. In the earlier version of the model theory, the possibilities to which an assertion referred were in a disjunction; in the present version, the possibilities hold in default of information to the contrary, and they are in a conjunction. The difference is subtle, but important. We illustrate the contrast below in example (6).

Counterfactual conditionals, such as:

5. If he had pulled the trigger then the gun would have fired

refer by default to situations that were once possible but that did not happen—counterfactual possibilities (Byrne, 2005). In example (5), *not A & not B* is a fact: He did not pull the trigger and the gun did not fire, but *A & B* is a counterfactual possibility: It was once possible that he pulled the trigger and the gun fired. Likewise, *not-A & B* is another counterfactual possibility: It was once possible that he did not pull the trigger but the gun fired—someone else perhaps pulled the trigger. And *A & not B* was never a possibility according to (5): He pulled the trigger and the gun did not fire.

An inclusive disjunction, *A or B or both*, refers by default to the exhaustive conjunction of these default possibilities that each holds if no information is to the contrary:

possible(A & not B)
& possible(Not A & B)
& possible(A & B).

Only one case is missing from the exhaustive possibilities, and so it is impossible:

impossible(not A & not B).

Critics have argued: "But these [first] three cases are always possible for jointly contingent statements: that is why they are rows of the truth table for [the disjunction *A or B or both*]. This new definition makes almost every disjunction true!" (Baratgin et al., 2015). The critics are mistaken: two jointly contingent assertions also allow the possibility of *not-A and not-B*, but the disjunction rules this case out as impossible. And the inference from the disjunction to a conclusion about a possibility (e.g., *possibly A and not B*) is invalid in infinitely many epistemic modal logics (see below). Zimmermann (2000) proposed that deontic disjunctions refer to lists of alternatives in a possible-world semantics, and Geurts (2005) extended this analysis to factual disjunctions. In contrast, the model theory applies it to all compounds, including conditionals.

Conjunctions, *A and B*, and other assertions referring to only one possibility, are special cases in which all other possibilities are eliminated:

possible(A & B)
& impossible(A & not B)
& impossible(not A & B)
& impossible(not A & not B).

An assertion of a single possibility is equivalent to an assertion of a matter of fact.

A recent experiment examined several sorts of inference for which, as the model theory predicted, participants inferred conclusions about possibilities from premises that made no mention of them (Hinterecker, Knauff, & Johnson-Laird, 2016). A typical trial was:

6. The U.S. will ratify the Kyoto protocol and commit to reducing CO_2 emissions, or global temperatures will reach a theoretical point of no return in the next 100 years, or both.
   Therefore, possibly, the U.S. will ratify the Kyoto protocol.

The participants drew three sorts of inference on separate trials, and their percentages of acceptance are shown in parentheses:

A or B or both.
Therefore, possibly A. (91%)
Therefore, possibly B. (94%)
Therefore, possibly A and B. (88%)

and they rejected one sort:

Therefore, possibly not A and not B. (18%)

Yet the three inferences that people accept are invalid if the possibilities are in a disjunction (as in the previous version of the model theory). A disjunction implies only that at least one of them is true, not all of them. They are also invalid in any modal logic. Example (6) would be invalid in case *B* is true: Global temperatures will reach a theoretical point of no return in the next 100 years, but *A* is false: the U.S. will not ratify the Kyoto protocol. In this case, the premise in (6) is true in logic, but the conclusion is false, and so the inference is invalid.

The way to save the logical analysis of inference (6) is to treat it as an enthymeme, that is, as an inference missing a premise. Whatever the premise may be, it needs to guarantee that *A* is not impossible, that is, it is not impossible that the U.S. signs the Kyoto protocol. A premise that appears to do the trick is:

7. It is not impossible that the U.S. signs the Kyoto protocol.

Alas, this premise is equivalent to the conclusion to be proved, and so the inference is valid but circular. Without such a premise, or one that implies it, the modal inference of *possibly A* from a non-modal disjunction, *A or B or both*, is invalid in all modal logics. So, the new theory differs from modal logics.

2.2. Validity depends on modal semantics and difficulty depends on models

Because its semantics for compounds is not truth functional, the model theory does not allow the paradoxes of the material conditional, for example, (4). It has a similar consequence for disjunctive inferences such as:

8. Eva read a newspaper.
   Therefore, Eva read *Don Quixote* or a newspaper or both.

The inference is valid in logic. It is also p-valid in probabilistic logic, which we described earlier. But, according to the model theory, for a conclusion to be a necessary consequence of an inference, its premises must refer to those possibilities to which the conclusion also refers. The premise in (8) refers to the possibility that Eva read a newspaper, but not to the possibility that she read *Don Quixote*. It's possible that she read *Don Quixote*, but not necessary. The conclusion refers to the following possibility:

9. Eva read *Don Quixote* and she did not read a newspaper.

and it is not one to which the premise in (8) refers. Most of the participants (71%) in an experiment rejected the inference in (8), just as they rejected paradoxes of the material conditional (Orenes & Johnson-Laird, 2012).

The inference in (10) raises a problem of its own:

10. Scientists will discover a cure for Parkinson's disease in 10 years OR the number of patients who suffer from Parkinson's disease will triple by 2050, but NOT both.
    Therefore, scientists will discover a cure for Parkinson's disease in 10 years OR the number of patients who suffer from Parkinson's disease will triple by 2050, OR both.

The inference is from an exclusive to an inclusive disjunction:

A or B but not both.
Therefore, A or B or both.

It is valid in logic, because each case in which the premise is true is also true for the conclusion. It is also p-valid, because its conclusion is at least as probable as its premise in any consistent assignment of probabilities. But it is not valid in the model theory, because its premise does not support the possibility of both *A* and *B*. As the model theory

predicts, almost no participants (3%) accepted such inferences from an exclusive to an inclusive disjunction (Hinterecker et al., 2016, Experiment 1). Skeptics might argue that the participants based their decision on the clash between "but not both" and "or both." But this argument fails to explain why more participants (24%) accepted the inference from an inclusive to an exclusive disjunction:

A or B or both.
Therefore, A or B but not both.

The inference is invalid in logic because the premise refers to a possibility to which the conclusion does not. But the model theory permits a weak sense of necessity in which a conclusion is evaluated as necessary because it refers only to possibilities to which the premises refer, though they refer to other possibilities. In the example above, the premises refer to three possibilities, and the conclusion refers to only two of them.

Some conclusions are easy to draw while others are difficult. Models are iconic, and each one represents a distinct possibility, that is, what is common to the different ways in which the possibility can occur. Hence, inferences that require reasoners to consider multiple distinct possibilities call for multiple distinct models. The model theory's long-standing prediction is that the more models that are needed to make an inference, the more difficult the inference should be—it should take longer and be more prone to error. Many studies have corroborated this prediction (e.g., Khemlani, Orenes, & Johnson-Laird, 2014). The new theory maintains and extends the prediction.

To reduce the load on working memory, reasoners tend to build mental models, which represent only what is possible in each model and omit what is impossible. Consider a disjunction such as:

11. There is a circle or there is a triangle, or both.

It refers to a conjunction of three exhaustive possibilities, and iconicity demands a model of each of them. The mental models of (11) can be depicted in the following sort of diagram in which each row denotes a mental model of a different possibility:

○ [circle & not triangle]
△ [not circle & triangle]
○ △ [circle & triangle]

The models represent only possibilities, not impossibilities (cf. Johnson-Laird & Savary, 1999). And within each possibility, the model does not represent what is impossible, and so it does not represent the absence of the triangle in the first row. Likewise, the conditional:

12. If there is a circle then there is a triangle

has only mental models of what is possible:

○ △
...

As the ellipsis suggests, the second model does not make explicit the possibilities in which the subordinate *if*-clause is false; that is, there is not a circle. *Mental* models underlie intuitive reasoning, but for simple tasks, they can be fleshed out into *fully explicit* models, which use negation to represent clauses in the premises that are false. So the fully explicit models of the conditional (12) are:

○ △
¬○ ¬△
¬○ △

They are listed in the order in which individuals tend to think of them. These models correspond to the exhaustive set of default possibilities to which the assertions refer. Table 1 presents the mental models and the fully explicit models for compound assertions based on the main connectives. The models in the table are the same as those in the previous version of the theory, but their interpretation is different. Compound assertions with only a single model refer to putative facts, whereas compounds with multiple models refer to conjunctions of default possibilities.

Reasoning depends on conjoining the sets of models for the different premises. The process is simple in principle (see, e.g., Johnson-Laird, 2006, Ch. 8). Two sets of models are conjoined pairwise to form their product. So, if two models—one from one set and one from the other set—are consistent with one another, the result is a model of all the propositions represented in both models. If the two models are inconsistent with one

Table 1
The mental models and fully explicit models for compound assertions based on the principal sentential connectives

Connective	Compound Assertion	Mental Models	Fully Explicit Models
Conjunction	A and B	A B	A B
Joint denial	Neither A nor B	¬A ¬B	¬A ¬B
Exclusive disjunction	A or else B but not both	A	A ¬B
		B	¬A B
Inclusive disjunction	A or B or both	A	A ¬B
		B	¬A B
		A B	A B
Conditional	If A then B	A B	A B
		...	¬A ¬B
			¬A B
Biconditional	If and only if A then B	A B	A B
		...	¬A ¬B

Notes. "¬" denotes negation, and "..." represents additional but implicit possibilities. Multiple models (e.g., for "A or B or both") represent conjunctions of default possibilities, and a single model (e.g., for "A and B") represents a fact.

another, because one represents a proposition and the other its negation, the result is the null model; that is, the empty model akin to the empty set, which is a subset of all sets. The null model represents contradictions. As an example, consider the product of the fully explicit models for *if A then B* and for *A*, where "¬" denotes negation, and the output of the computer program is in a different font from the main text:

If A then B.			*A.*		The products of the pairs of models
A	B	&	A	→	A B
¬A	¬B	&	A	→	null model
¬A	B	&	A	→	null model

Just as the empty set is a subset of all sets, so the null model is a member of all sets of models, and therefore it can be ignored, provided that there are models in the set that are not null. Hence, the result of the product is:

```
A   B
```

from which a factual claim can be inferred: B follows.

Baratgin et al. (2015) argued that the model theory cannot treat a disjunction, *A or not A*, as a tautology because the conjunction of *A and not-A* is impossible. Here they overlooked the theory's mechanism for combining models (see, e.g., Johnson-Laird, 2006, Ch. 8). The disjunction, *A or B or both*, has the following fully explicit models (see Table 1):

```
 A  ¬B
¬A   B
 A   B
```

If we substitute *not A* for *B* in these models, the result for the disjunction *A or not A* is:

```
 A  ¬¬A
¬A   ¬A
 A   ¬A
```

The double negation in the first conjunct cancels out to yield *A*, the second conjunct yields ¬*A*, and the third conjunct is a contradiction, so the product of these models is:

```
 A
¬A
```

The product shows that each case in the partition of the disjunction is possible, and so they represent a tautological conjunction of possibilities: *A is possible and not-A is possible*.

The model theory postulates two systems for reasoning. One uses mental models and the other uses fully explicit models. The original dual-system of deductive reasoning is

due to the late Peter Wason (Johnson-Laird & Wason, 1970; see also Ragni, Kola, & Johnson-Laird, 2017; Wason & Johnson-Laird, 1970), and Wason's co-authors are among the many who have developed dual systems (e.g., Evans, 2008; Johnson-Laird & Steedman, 1978; Stanovich, 1999; Wason & Evans, 1975). The two systems in the model theory compute different functions from the same domain of premises. System 1 emulates intuitive reasoning: It focuses on inferences yielding a single mental model, which it can hold in a memory buffer. The buffer has a small finite capacity, and so system 1 is equivalent to a finite-state automaton (Aho & Ullman, 1972). System 2 emulates deliberation: It relies on fully explicit models. It has access to a working memory of limited processing capacity, and so it can count, consider alternative models, and carry out other recursive processes. The theory therefore treats the two systems as distinct, though they interact: System 1 can err, and system 2 can often correct these errors—its reasoning is correct unless an inference demands more working memory than is available to the system. Unlike some other dual process theories, the model theory's two systems appear to be unique in that they share many components in common, and their two algorithms are close to one another. The two systems differ in the conclusions that they yield from certain premises (for examples, see Table 4 below). Indeed, that divergence is an essential property of a dual system of reasoning. Sentential reasoning is computationally intractable (Cook, 1971), and it is intractable in the model theory too, and so it will fail for inferences that call for more working memory capacity than is available—a factor that varies from one person to another.

We illustrate the way in which System 1 and System 2 differ by considering two different inferences. The following sort of inference (known as modus ponens) is from a conditional premise and a categorical premise:

13. If it is cloudy then it is raining.
    It is cloudy.
    Therefore, it is raining.

The inference is simple, and most reasoners are able to make it (Johnson-Laird, Byrne, & Schaeken, 1992). Earlier, we showed how individuals can combine the models of the premises to infer the conclusion, *It is raining*. A contrasting inference (known as modus tollens) is as follows:

14. If it is cloudy then it is raining.
    It is not raining.
    Therefore, it is not cloudy.

The product of the mental models of the two premises is:

Conditional		Categorical		The products of the pairs of models
cloudy raining	&	¬raining	→	null model
...	&	¬raining	→	¬raining

The products yield only a model of the categorical premise, and so it seems that nothing follows—an erroneous response that reasoners often make (Johnson-Laird et al., 1992). They can reach the valid conclusion only if they flesh out their mental models of the conditional into fully explicit models (see Table 1). They can then grasp that the conditional refers to a possibility consistent with the categorical premise:

```
¬cloudy ¬raining
```

This model yields a novel conclusion that is necessary given the premises: *It is not cloudy*. The inference should be more difficult than modus ponens, and it is. The model theory predicts that modus tollens should be easier from a biconditional premise, *If and only if A then B*, because it has only two fully explicit models of possibilities whereas a conditional has three such models (see Table 1). This prediction has been corroborated in experiments (Johnson-Laird et al., 1992), and it is difficult to explain for theories based on formal rules of inference.

As the model theory predicts, reasoners tended to accept inferences of the following sort (Espino & Byrne, 2013; Ormerod & Richardson, 2003):

15. A or B or both.
    Therefore, if not A then B.

They do so more often than they accept inferences of the converse sort:

16. If not A then B.
    Therefore, A or B or both.

The inference in (15) can be drawn from the mental model of the premises and the conclusion:

```
¬A  B.
```

But (16) cannot be drawn from a mental model of the conclusion. It depends on fully explicit models. The inference (15) is p-invalid, whereas (16) is p-valid (see Adams, 1998, p. 120). Hence, the results corroborate the model theory but contravene probabilistic logic.

2.3. Modulation from knowledge, content, and context

The model theory posits that reasoners rely on knowledge to "modulate" the interpretation of sentential connectives (Johnson-Laird & Byrne, 2002). The meanings of words, knowledge, and the conversational context can block the construction of models of possibilities, and they can add causal, spatiotemporal, and other relations between elements in

models. Experiments have corroborated these effects. A simple example occurs with a conditional such as:

17. If Fred is on his bike then he'll fall off it.

It refers to just two possibilities, because it is impossible for Fred to fall off his bike unless he is on it. Modulation can therefore block the construction of models, and thereby yield various interpretations of conditionals (e.g., Johnson-Laird & Byrne, 2002; Juhos, Quelhas, & Johnson-Laird, 2012; Quelhas & Johnson-Laird, 2016; Quelhas, Johnson-Laird, & Juhos, 2010). A corollary affects inferences of the sort:

B
Therefore, A or B or both.

As we saw earlier, reasoners tend to reject this sort of inference: The premise, *B*, fails to support the possibility of *A and not B*. Such inferences, however, should be more acceptable if *A* implies *B*. Here's an example:

18. Eva read a novel.
Therefore, Eva read *Don Quixote* or a novel.

It is impossible that Eva read *Don Quixote* and not a novel, because *Don Quixote* is a novel. And so there are just two possibilities for what Eva read:

```
 Don Quixote     a novel
¬Don Quixote     a novel
```

Both of them refer to the same possibility as the premise does, and reasoners tended to accept the inference (Orenes & Johnson-Laird, 2012).

Johnson-Laird and Byrne (2002) proposed the process of modulation, but they did not implement it computationally, and they did not anticipate its role in yielding a priori truths (see Section 4.3). Below, we describe how the computational implementation of the updated theory carries out the process. An important corollary of modulation is contrary to formal rules of inference: Because content and context can block the construction of possibilities, inferences in daily life have to be evaluated on their own merits. Hence, apart from those "core" inferences on which knowledge and context have no effect, inferences depend, not just on the meanings of their clauses, but on the relations between them.

2.4. Verification and modal semantics

The conditions in which assertions are true should not be confused with the process of verification, which checks whether these conditions hold given observations or facts at

hand—a process that can be difficult and even impossible. To illustrate what is at stake, suppose that a doctor advises a patient to have an advanced treatment for an illness. The doctor predicts:

19. If you don't have the advanced treatment then you won't get better.

The patient gets a second opinion from another doctor, who makes a contrary prediction:

20. If you don't have the advanced treatment then you will get better anyway.

The doctors disagree, and so the patient gets an opinion from a third doctor, who predicts:

21. If you have the advanced treatment then you will get better.

This opinion supports the first doctor's advice, and so as a matter of fact:

22. The patient has the advanced treatment and gets better.

So, which of the three doctors made the correct prediction? Readers may care to jot down their answer.

In truth-functional logic, the fact that the patient has the treatment and gets better verifies all three of the doctors' prognoses (19–21). They are all true. The example illustrates a nasty secret about specific assertions in logic: No single fact can corroborate one and only one conditional or disjunctive prediction. It always corroborates multiple predictions. Indeed, the fact above implies only one certainty—the falsity of a fourth conditional:

23. If you have the advanced treatment then you won't get better.

Table 2 presents the truth tables of the four preceding conditional predictions. It shows each possible sort of evidence, that is, the partition of *A & B*, *A & not-B*, *Not-A & B*, and *Not-A & not-B*, and the truth value of each conditional in each of these cases. The first row of the table confirms our analysis above: All three doctors made true predictions.

The model theory differs from logic. The fact that the patient had the treatment and got better does not corroborate all three of the conditionals. System 1 yields the intuition that a fact that matches a mental model of a conditional corroborates it. Hence, the fact in (22) corroborates only the conditional in (21), that is, column 3 in Table 2. In system 1, the fact is irrelevant to the other two predictions, because it does not correspond to their mental models. Because system 2 can consider multiple possibilities, it recognizes that definitive truth values for specific conditionals also depend on the counterfactual possibilities that evidence creates. We describe these counterfactuals below in our outline of the program implementing the theory.

Table 2
Four material conditionals and their truth tables in logic, which show their truth values for each case in the partition

	The Four Sorts of Material Conditional Based on *A*, *B*, and Their Negations, With Their Truth Values for Each of the Four Cases in the Partition			
The Four Cases in the Partition	1. If not A then not B	2. If not A then B	3. If A then B	4. If A then not B
A & B	True	True	**True**	
A & not B	True	True		**True**
Not A & B		**True**	True	True
Not A & not B	**True**		True	True

Notes. "True" denotes that the conditional is true; an empty cell denotes that it is false. The model theory's system 1 yields the intuition that each case in the partition verifies only one conditional (in bold) whose explicit mental model matches it.

Analogous truth conditions hold for specific disjunctions, such as:

24. There is soda or beer in the fridge, or both.

It is true provided that the conjunction of possibilities to which it refers is true. In logic, however, it suffices for soda to be in the fridge for the disjunction to be true. That is the case for a truth-functional interpretation; and system 1 allows intuition to yield the same result by default. But suppose that observers discover that it is impossible for beer to be in the fridge. In these circumstances, they might be inclined to dismiss the disjunction, or even to answer that it is only partly true. For the same reason, as we saw earlier, they reject the following inference:

25. There is soda in the fridge.
    Therefore, there is soda or beer in the fridge, or both.

In short, the empirical verification of specific assertions is subtle. If system 1 matches evidence to a mental model of an assertion, its intuitive response is that assertion is true. If it matches evidence to a mental model of the falsity of an assertion, its intuitive response is that the assertion is false. Otherwise, its intuitive response is that the evidence is irrelevant. Johnson-Laird and Tagart (1969) observed such judgments, which they attributed to a "defective" truth table. This explanation is contrary to the new theory's semantics, which a recent study of verification corroborates (Goodwin & Johnson-Laird, under review). System 2 can examine the alternative models corresponding to counterfactual cases that evidence creates, and it may be able to determine that an assertion is true for certain. Verification can depend on background knowledge and even on experimental investigation, but these matters take us beyond specific assertions to general ones. The model theory's central prediction is that accurate verification depends on considering the facts from evidence and the status of counterfactual possibilities that the evidence creates.

Many studies bear out the main predictions of the model theory. Table 3 lists the principles and predictions of the new model theory, and it provides goals for any theory of sentential reasoning to meet.

3. A computational implementation of the model theory

We developed a program, mSentential, that implements the principles of sentential reasoning in Table 3. Its source code in Common Lisp is downloadable from http://mentalmodels.princeton.edu. It enabled us to discover some recondite predictions of the theory and to model experimental results. The implementation shows how reasoning is nondeterministic and stochastic. This section describes it in sufficient detail for programmers to understand and to implement its basic principles, and illustrates its stochastic role in modeling data.

The program emulates sentential reasoning. It can carry out the following seven tasks given a set of premises:

- It can infer its own simple conclusion.
- It can establish whether a given conclusion follows necessarily, possibly, or not at all.

Table 3
Five primary principles of the new model theory, the empirical predictions they yield, and examples of studies that corroborate the predictions

Principle of the Model Theory	Prediction	Exemplary Data
Representation: Reasoners interpret compound assertions as conjunctive sets of possibilities	1. Reasoners should draw modal conclusions from non-modal premises, e.g., *A or B or both. Therefore, possibly A*	Hinterecker et al. (2016)
Inference: Necessary inferences are those in which the models of the premises support all and only the models of the conclusion	2. Reasoners should reject inferences in which the premises do not support one of the models of the conclusion, e.g., *A or B but not both. Therefore, A or B or both*	Hinterecker et al. (2016)
Dual systems: Intuitive inferences depend on mental models and deliberative inferences depend on fully explicit models	3. Mental models should lead to fallacies in certain cases, e.g., *One of these assertions is true and one of them is false: A and B. B or else C. Therefore, it is possible that A and B*	Khemlani and Johnson-Laird (2009)
Modulation: Background knowledge blocks the construction of possibilities and can add relations	4. Reasoners should interpret ambiguous disjunctive constructions, e.g., *A or B*, as exclusive disjunctions when the contents block the model of *A and B*	Orenes and Johnson-Laird (2012); Quelhas and Johnson-Laird (2016)
Verification: It depends on relations between the evidence and models of assertions	5. Intuitions should evaluate some evidence as irrelevant to the truth or falsify of conditionals	Goodwin and Johnson-Laird (in press)

- It can assess the consistency of the premises, that is, whether or not they could all be true.
- It can establish the extensional probability of a conclusion.
- It can establish whether any premise is true or false a priori or else is contingent.
- It can establish whether given evidence verifies a premise.
- It can construct explanations from its knowledge base to resolve inconsistent premises.

Fig. 1 presents a diagram of the program's overall structure, showing its two systems and their main components, many of which they share. The intuitive system carries out the main functions for system 1. It bases its conclusions on a single mental model. The deliberative component is the heart of system 2 for reasoning with fully explicit models. Modulation consults a small illustrative knowledge base to block the construction of models. The verification component assesses whether or not a compound assertion could be true in the light of evidence and formulates the counterfactuals that need to hold for it to be true for certain. All inferences in daily life are defeasible. And so when a premise is

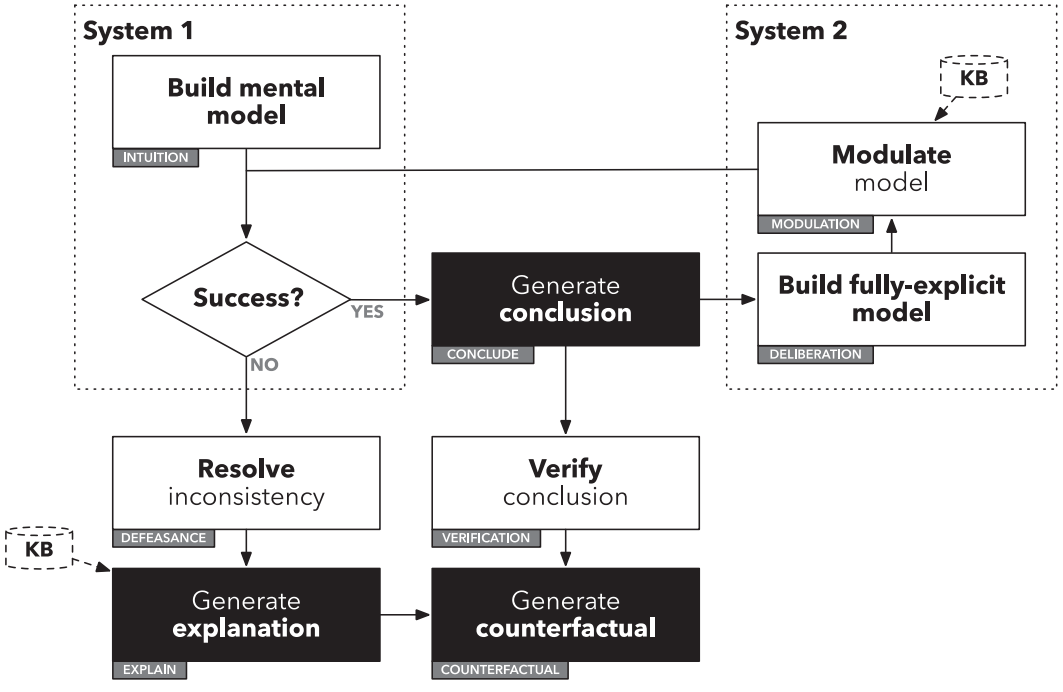

Fig. 1. A diagram of the reasoning program implementing the new model theory. The white boxes are its principal components, the black boxes are outputs that the program generates, and the arrows denote the flow of control from one component to another. The diamond labeled "Success?" denotes a test for whether the system constructed a non-null model. Cylinders labeled "KB" depict stages at which the program looks up information from a small knowledge base of models representing facts. Gray boxes denote the names of the specific functions in the code that compute the task depicted.

inconsistent with an earlier inference, both systems 1 and 2 call a defeasance component, which rejects premises (see Marek & Truszczynski, 2013). It decides which premise to abandon, and it consults the program's rudimentary knowledge base to create, if possible, an explanation that resolves the inconsistency (see Johnson-Laird, Girotto, & Legrenzi, 2004; Khemlani & Johnson-Laird, 2011, 2012, 2013a). It also formulates a counterfactual conditional to describe an idealized version of the premise it has rejected. We now outline how each of the program's components work, and how its stochastic parameters allow it to model data.

3.1. Lexicon, grammar, and parsing

The program has a simple language component, which is not shown in Fig. 1. It contains a lexicon consisting of the principal sentential connectives, negation, punctuation, and sentential variables. Each connective has its part of speech and a meaning consisting of a modal semantics used to construct sets of models of possibilities. Both systems 1 and 2 rely on the parser to provide information for building models—mental models for system 1 and fully explicit models for system 2. A shift-and-reduce parser (Aho & Ullman, 1972) uses an unambiguous but recursive context-free grammar, and it captures well-formed compound assertions, such as:

26. It is raining and it is windy ore it is not sunny.

where "ore" is an exclusive disjunction. The parser closes off structures as soon as it can, unless it encounters a comma, which functions as a left parenthesis. So assertion (26) has the grammatical structure: *(A and B) or not C*, where *A*, *B*, and *C* constitute the "atoms" of the sentence, that is, constituents that the parser cannot reduce any further. Each grammatical rule captures a different sort of compound and has attached to it an appropriate semantic function that directs the building of models. The parser calls the semantic functions from the lexicon and grammar to construct models. Given the premise in (26), the output of the parser is a conjunction of mental models of possibilities in system 1:

```
raining windy
               ¬sunny
```

and fully explicit models in system 2:

```
 raining  windy  sunny
¬raining  windy ¬sunny
 raining ¬windy ¬sunny
¬raining ¬windy ¬sunny
```

We use abbreviations (e.g., raining), where necessary, to represent atomic sentences.

3.2. System 1 bases intuitions on single mental models

System 1 has a procedure for forming the product of two sets of mental models, which relies on forming pairwise combinations of them (see the principles described in Section 2.2). If one model in a pair is inconsistent with another, the result is the null model. Likewise, if one model in the pair is a null model, the result is null too. Otherwise, the result is a model of the possibility that the pair represents, taking into account the role of implicit models. System 1 attempts to form a conclusion from premises that yield a single mental model (as in examples 13 and 14 above). But, if an atom is affirmed (or negated) in all of the models, system 1 will assess the atom (or its negation) as necessary.

If a given conclusion refers to a possibility and is affirmed in at least one model of the premise models, system 1 assesses it as possible. Consider the test case (3) in the introduction:

3. The card is an ace or it is a heart, or both.
   Therefore, it is possible that the card is an ace.

The program establishes that an ace occurs in at least one model of the premises, and so it can infer that the conclusion is possible (see Table 3, prediction 1). It draws this conclusion, and the program can also call system 2 to check the inference. As the next section shows, the program can draw other sorts of conclusion.

3.3. Conclusions depend on the relations between models

The procedure for drawing conclusions checks the relation between the models of the premises and the models of the conclusion. The premises of an inference "support" a model of the conclusion if that model occurs as part or whole of one of the premises' models, for example, the premise model: A B C, supports each of the following conclusion models:

```
A
  B
    C
A B
A   C
  B C
A B C
```

Granted that the premises and conclusion have at least one atom in common, there are five set-theoretic relations between their respective models. First, if the premise models support none of the conclusion models, the conclusion is impossible—the premises and conclusion are inconsistent with one another. Second, if the premises support all and only the conclusion models, the conclusion is necessary. Third, if the premises support all the

conclusion models but the conclusion has at least one other model too, as in the inference: *A*, therefore *A or B*, the conclusion is necessary in logic, but it is only possible in the model theory. Fourth, if the premises support all the conclusion models but the premises have at least one other model, the conclusion is possible. Reasoners may occasionally infer that the conclusion follows of necessity, because the premises support all of its possibilities. The program has a parameter to allow it to infer that that the conclusion is of "weak necessity" in this case; otherwise, it is only possible. Fifth, in any other case, the premises support at least one of the conclusion models, and the conclusion is possible. Table 4 below presents examples of these relations for both systems. We mentioned that the premises and conclusion represent atoms in common. If they don't, then they are independent of one another unless knowledge establishes relations between them.

3.4. System 2 bases deliberations on fully explicit models

System 2 forms products of two sets of fully explicit models. Consider again the premise in (3): *The card is an ace or it is a heart, or both*. System 2 builds the fully explicit models of the premise:

```
 ace ¬heart
¬ace  heart
 ace  heart
```

And it, too, can infer that it's possible that the card is an ace. On the assumption that the three possibilities are equiprobable, a separate component of the program (not depicted in Fig. 1) infers from these models an extensional probability of 2/3 that the card is an ace, because an ace occurs in two out of the three models of possibilities. The program assumes that each model is equiprobable (see Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999, for corroboratory evidence). System 2 can therefore make inferences about possibilities and probabilities from premises that make no reference to them.

Table 4
Nine pairs of outputs from systems 1 and 2

The Premises	System 1 Inference	System 2 Inference
1. A. If A then B	B is necessary	B is necessary
2. B. If A then B	A is weakly necessary	A is possible
3. Not A. If A then B	Not B is possible	Not B is possible
4. Not B. If A then B	Not A is possible	Not A is necessary
5. A. A or B, or both	Not B is weakly necessary	Not B is possible
6. Not A. A or B, or both	B is necessary	B is necessary
7. A. A ore B ore C	Not B and not C is necessary	Not B and not C is possible
8. Not B and not C A ore B ore C	A is necessary	A is necessary
9. A ore B Not-A ore B	B is necessary	B is impossible: inconsistent premises

The inference of *or*-introduction, which is:

A.
Therefore, A or B or both.

follows as a necessity in truth-functional logic, but only as a possibility in the model theory (contrary to a claim in Cruz, Over, & Oaksford, 2017). Its premise fails to support the possibility of *not-A and B*, to which the conclusion refers. In a revealing contrast, the inference:

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

yields a necessary conclusion. The procedure for building models yields these fully explicit ones for the two premises:

```
A ¬B C
A  B C
```

The function that evaluates conclusions determines that *C* is common to both models, and so it follows of necessity.

System 1 is less powerful than system 2: The intuitive system has the power of only a finite-state automaton, and so any loop of operations repeats only a small finite number of times in system 1, with a maximum that the program's users can set. Table 4 presents some illustrative differences between the program's deductions using the two systems. As the table shows, system 1 can err in several ways; for example, it can infer that a conclusion is necessary when it is only possible (e.g., inference 7), or even impossible (e.g., inference 9); it can infer that a conclusion is only possible when it is necessary (e.g., inference 4); and it can infer cases of weak necessity in which the premises refer to all the possibilities to which the conclusion refers but the conclusion refers to other possibilities too (e.g., inference 5). It likewise errs in assessments of the consistency of assertions. In principle, system 2 does not err; in practice it suffers from the limited capacity of working memory.

3.5. The use of the program to model data

We used the program to simulate sets of data that test critical predictions of the model theory. The program uses two parameters that control its performance. One parameter (σ) sets the probability of engaging system 2, and the other parameter (γ) sets the probability of using weak necessity in assessing a conclusion. The theory's first prediction in Table 3 is its strongest: Reasoners should make modal inferences from compound sentences in accordance with the mental models of the compound. The program makes the same predictions regardless of the values of its two parameters for the data from Hinterecker et al. (2016; Experiment 3). Fig. 2 (top left panel) presents the data for the default parameter

settings ($\sigma = 0.0$, $\gamma = 0.0$), and it shows the program's predictions, which it generated from carrying out the four inferences 1,000 times each. The program's output for these data and subsequent simulations is available at: https://osf.io/ftje8/. The results matched the participants' performance in the experiment well ($R^2 = .99$, RMSE = .12). Hence, the program makes the first prediction of the model theory (in Table 3): Compounds refer to conjunctive possibilities, and so reasoners infer possibilities from compounds making no mention of them.

Hinterecker et al. (2016) ran another study (Experiment 1) that tested the theory's second prediction (in Table 3): Reasoners should reject deductions in which the conclusion refers to possibilities that the premises do not support. The program modeled the results, allowing that

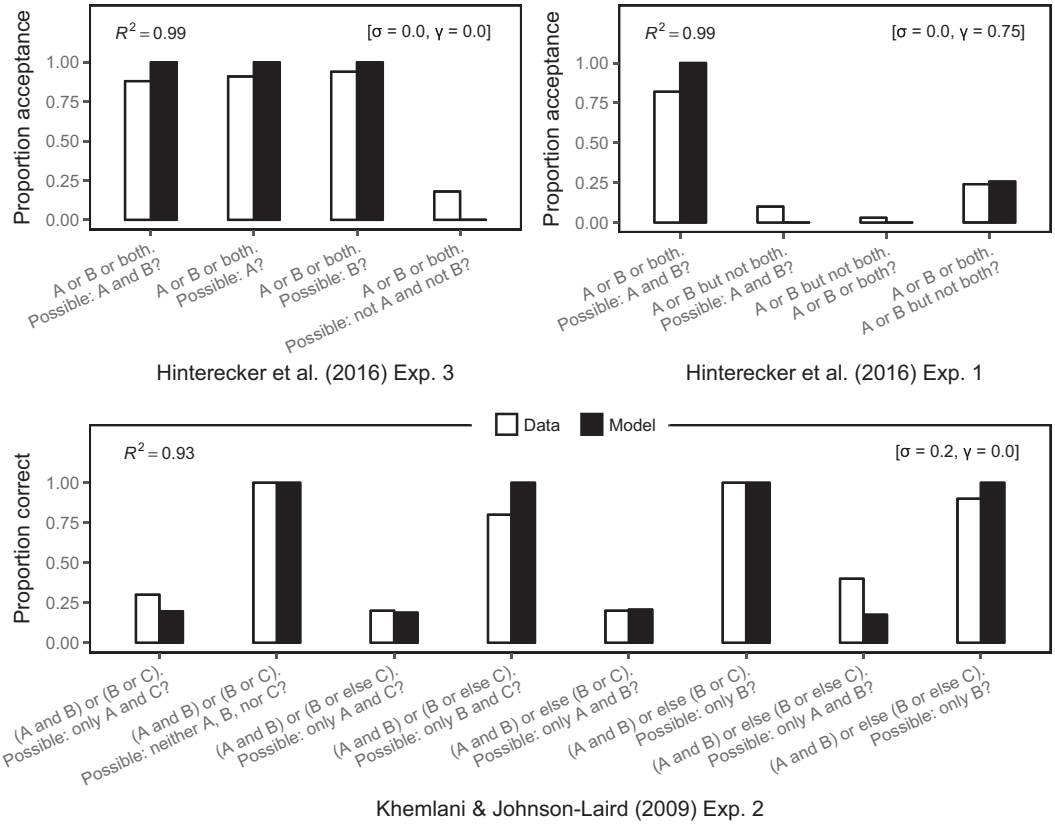

Fig. 2. The percentages of participants' inferences in studies testing three of the model theory's predictions. The white bars plot the data from the experiments, and the black bars show the computer model's predictions using the parameters that govern (a) the probability that system 2 is engaged in making inferences (σ); and (b) the probability that system 2 uses a weak notion of necessity (γ). *Top left panel:* Data from Hinterecker et al. (2016), which tests the theory's first prediction (see Table 3). *Top right panel:* Data from Hinterecker et al. (2016), which tests the theory's second prediction. *Bottom panel*: Data from Khemlani and Johnson-Laird (2009, Experiment 2), which tests the theory's third prediction. For brevity, the bottom panel uses "or" to indicate inclusive disjunctions and "or else" to indicate exclusive disjunctions.

reasoners might rely on weak necessity; that is, they should accept inferences in which the conclusion refers only to possibilities that the premises refer to, though the premises also refer to other possibilities (see Section 3.3). The probability of relying on weak necessity depends on the γ parameter. The program simulated reasoners' performance on these inferences. As before, it ran 1,000 simulations per inference, and its results matched the human data (R^2 = .99, RMSE = .10; see Fig. 2, top right panel). Hence, it modeled the theory's second prediction.

Khemlani and Johnson-Laird (2009) carried out a study that tests the theory's third prediction that certain inferences should yield different conclusions depending on whether they rely only on system 1 or on system 2 instead (see Table 3). For example, consider an inference of this sort:

(A and B) or (B or C).
Is it possible that only A and C?

The mental models of the first premise do not include one in which only *A* and *C* occur, and so as they predict, most people responded that their conjunction is impossible. But the fully explicit models of the premise include:

```
A ¬B  C.
```

Hence, the correct conclusion is that the conjunction of *A and C* alone is possible. A control problem changed the conclusion to: *Is it possible that neither A, B, nor C*. The mental models do not include such a model, and most people respond that the conclusion is not possible. The fully explicit models show that this response is correct. The program simulated reasoners' performance on these and other inferences. The probability of engaging system 2 and building fully explicit models depends on the σ parameter. As before, the program ran 1,000 simulations per inference, and its results matched human data (R^2 = .93, RMSE = .12; see Fig. 2, bottom panel).

3.6. Modulation of the interpretation of compounds

Modulation checks whether a current premise matches any knowledge in the knowledge base. This mechanism is more akin to a conventional search through a lookup table rather than a realistic retrieval from semantic memory. Other systems, such as ACT-R (Anderson, 2007), deal with memory retrieval, but the process we describe finesses the issue, because modulation appears to depend on what is retrieved rather than on how it is retrieved. As a result, we chose not to fit the computational model to data on experiments that test predictions based on modulation from background knowledge. We instead describe validations of the program's qualitative predictions.

The program searches for atoms in the premise that match those in a knowledge base. If there are none, or the models in knowledge contain atoms that are not in the premise, then the program reports that there is no prior relevant knowledge and that the assertion is contingent. But when the knowledge base contains models with atoms matching those in a premise, the program evaluates these models. However, it diverges from the model

theory in one respect: The theory permits that knowledge can modulate both intuitions and deliberations, but the program's method of retrieving knowledge works well with only fully explicit models. It does not successfully retrieve knowledge based on mental models, and so it arbitrarily calls on modulation only from system 2. Future versions of the program should resolve the divergence.

To illustrate the process, consider the following conditional assertion:

27. If it is raining then it is pouring.

In principle, the conditional has three fully explicit models (see Table 1):

```
 raining  pouring
¬raining ¬pouring
¬raining  pouring
```

But the meaning of *pouring* implies that it is *raining*, and this knowledge is represented in fully explicit models in the program's knowledge base:

```
 pouring  raining
¬pouring  raining
¬pouring ¬raining
```

The program computes the product of this set of models and the set for assertion (27). The result blocks the model in which it is pouring but not raining, which is not consistent with any of the models in knowledge, and so (27) has only two models:

```
 raining  pouring
¬raining ¬pouring
```

They yield a biconditional interpretation: If and only if it is raining then it is pouring.

A qualitative prediction of the modulation algorithm, which became evident after the initial implementation of the program, is its affect on assertions such as:

28. If it is pouring then it is raining.

Its models match the models in knowledge, and so the program responds that the assertion is true a priori. A major consequence of the modulation algorithm is therefore that knowledge should yield a priori truth values for certain assertions. Steinberg (1970, 1975) showed that participants sort assertions about categories into sets that reflect a priori truth values; for example, *the tulip is a flower* is true, and *the infant is an adult* is false. They also treat nonsensical assertions, such as *the moon is a newspaper*, as false, and their negations as true (Steinberg, 1972). But, as far as we know, no studies had examined compound assertions until Quelhas, Rasga, and Johnson-Laird (2017) presented participants with problems of this sort:

29. If Sonia has pneumonia then she is ill.
    Do you consider that this sentence is:
    True ☐ False ☐ Could be true or could be false ☐

Almost all participants judged such assertions to be true, and they judged the following sort to be false:

30. If Sonia has pneumonia then she is healthy.

A second qualitative prediction of the modulation algorithm shows that when knowledge provides a fact that affirms or denies a clause in a premise, the fact functions as a premise too, and it may yield an inference. From its knowledge that the Louvre is in Paris, the program concludes that Pat is married from the assertion:

31. If the Louvre is in Paris then Pat is married.

Table 5 illustrates various effects of modulation.

Table 5
Eight illustrations of modulation: Assertions, the effects of modulation on their fully explicit models, and descriptions of these outcomes

Assertions	Modulated Models	Their Descriptions (and Rationale)
If raining then hot	raining hot ¬raining hot ¬raining ¬hot	Unmodulated (models in KB do not interrelate raining with heat)
If raining then pouring	raining pouring ¬raining ¬pouring	Biconditional (models in KB prohibit pouring without raining)
Not raining or pouring	¬raining ¬pouring raining pouring	Exclusive disjunction (ditto)
If pouring then raining	pouring raining ¬pouring raining ¬pouring ¬raining	A priori true (ditto)
If God exists then atheism is wrong	exists wrong ¬exists ¬wrong	A priori true (models in KB prohibit joint affirmation or joint denial of God's existence and atheism)
God exists or atheism is right	exists ¬right ¬exists right	A priori true (ditto)
If and only if God exists then atheism is right	null model	A priori false (ditto)
If it is not raining then the Louvre is not in Paris	raining Louvre-in-Paris	It follows that it is raining (models in KB establish that the Louvre is in Paris)

Note. KB, knowledge base.

3.7. Verification and the provenance of counterfactuals

The empirical verification of specific assertions depends on relevant evidence, which the program treats as an additional assertion or as information from its knowledge base. If the models of the evidence match all the models of the compound, the compound is true. Observations, however, tend to take the form of categorical assertions or conjunctions. In system 1, if such a conjunction matches a mental model of a compound, the system judges that the compound is true; if matches the mental model of the falsity of the compound, the system judges that the compound is false; and in any other case it judges that the evidence is irrelevant.

For example, the assertion:

32. If the cause occurred then the effect occurred.

has the mental model:

```
cause effect
```

Evidence corresponding to:

```
cause effect
```

elicits a judgment of truth in system 1. Evidence corresponding to:

```
cause ¬effect
```

elicits the evaluation of falsity. Any other evidence, such as:

```
¬cause ¬effect
```

neither matches the mental models of the conditional nor the case in which it is false, and so it elicits the evaluation that the evidence is irrelevant.

In contrast, system 2 uses fully explicit models, such as these for the conditional in (32):

```
 cause  effect
¬cause ¬effect
¬cause  effect
```

So, the system can find a match between evidence and all three of these cases. It takes the definitive truth to depend on both the evidence and the counterfactual status of the remaining possibilities. Given, say, the evidence:

```
cause effect
```

system 2 judges that the conditional (32) is possibly true (see the example of the three doctors' predictions in Section 2.4). Decisive evidence has to corroborate in addition the counterfactuals created by the evidence above:

```
 cause   effect  – the evidence establishes this case as a fact.
¬cause  ¬effect  – a counterfactual possibility
¬cause   effect  – a counterfactual possibility
```

The program therefore frames a counterfactual conditional describing the counterfactual possibilities:

33. If the cause had not occurred then effect might not have occurred.

The *if*-clause has to be negative to describe the two counterfactual possibilities, but the *then*-clause could have been affirmative: *The effect might have occurred*. However, because the evidence shows that the effect did occur, the program opts for a *then*-clause of the opposite polarity to the facts, as (33) illustrates.

Table 6 summarizes the output of the program illustrating the various counterfactuals that can occur with conditionals, biconditionals, inclusive disjunctions, and exclusive disjunctions, in light of different sorts of evidence. There are alternative ways of expressing a counterfactual, for example, "could" can often be substituted for "might," "should" can often be substituted for "would," the order of clauses can be changed, and so on. The program predicts one way of expressing a counterfactual that, if true, yields definite truth for the compound that the evidence addresses. The first row in Table 6 illustrates the generation of the counterfactual in (33).

One detail remains. Suppose an indicative conditional asserts that if the cause occurred then the effect occurred, and the evidence is that, in fact, neither of them occurred. The evidence corresponds to one of the possibilities to which the conditional refers (¬ cause ¬ effect), and the other two cases therefore become counterfactual possibilities:

```
 cause effect
¬cause effect
```

The program uses them to infer that if the cause had occurred then the effect would have occurred, again using a *then*-clause of the opposite polarity to the facts. But it also uses the counterfactual possibilities to infer that the effect might have occurred even though the cause did not occur. The second clause of this assertion is true according to the evidence, and so the claim is known as a *semifactual.* The counterfactual and semifactual claims need to hold for compound assertions to be true for certain. In many cases, background knowledge establishes them. An observation that a patient had an anesthetic and became unconscious corroborates the claim that if the

Table 6
Counterfactual descriptions that the computer model generates: Compounds, factual evidence, the counterfactual possibilities, and their descriptions

Compound	The Factual Evidence	Counterfactual Possibilities	Program's Descriptions of the Counterfactual Possibilities
1. If A then B	A and B	Not-A and B Not-A and not-B	If A had not occurred then B might not have occurred
	A and not-B	–	(Assertion is false.)
	Not-A and B	A and B Not-A and not-B	B might not have occurred even though A did not occur, and if A had occurred then B would have occurred
	Not-A and not-B	A and B Not-A and B	B might have occurred even though A did not occur, and if A had occurred then B would have occurred
2. If and only if A then B	A and B	Not-A and not-B	If A had not occurred then B would not have occurred
	A and not-B	–	(Assertion is false.)
	Not-A and B	–	(Assertion is false.)
	Not-A and not-B	A and B	If A had occurred then B would have occurred
3. A or B, or both	A and B	A and not-B Not-A and B	B might not have occurred even though A occurred, and if A had not occurred then B would have occurred
	A and not-B	A and B Not-A and B	B might have occurred even though A occurred, and if A had not occurred then B would have occurred
	Not-A and B	A and B A and not B	If A had occurred then B might not have occurred
	Not-A and not-B	–	(Assertion is false.)
4. A or else B, but not both	A and B	–	(Assertion is false.)
	A and not-B	Not-A and B	If A had not occurred then B would have occurred
	Not-A and B	A and not-B	If A had occurred then B would not have occurred
	Not-A and not-B	–	(Assertion is false.)

patient is given an anesthetic then he will become unconscious, provided that one knows that without its administration he would not have become unconscious.

3.8. Summary

We have now outlined how the program implementing the model theory works. It can draw its own conclusions, evaluate a given conclusion as necessary or possible, assess the consistency of assertions, establish a priori truth values, verify a compound assertion and use it and the evidence to formulate counterfactuals that need to hold for definitive truth, and explain an inconsistency. These processes reflect the modulation of premises by knowledge, which can also lead to a priori truth values. Table 7 summarizes the main functions in the program's two systems and the effects of its two free parameters.

Table 7
The two systems in mSentential, their principal functions, and the two free parameters that regulate its performance

System 1	System 2	Free Parameters
• Constructs mental models of the premises • Derives, if possible, conclusions or evaluates given conclusions, focusing on single mental models • Its evaluations may rely on weak necessity depending on parameter γ • Whether or not it draws a conclusion, engages system 2 depending on parameter σ	• Fleshes out mental models into fully explicit models • Uses them to corroborate inferences or evaluations made in system 1, and makes its own inferences • Its evaluations may rely on weak necessity (γ) • Modulates assertions using background knowledge, including establishing a priori truth values • Given factual evidence, verifies the truth or falsity of an assertion, and formulates relevant counterfactuals • Attempts to reason by formulating an explanation based on knowledge to resolve an inconsistency, rejecting a premise, and reframing an original premise as a counterfactual	• σ = The probability that system 2 is called to check or to make an inference by using fully explicit models • γ = The probability that systems 1 and 2 rely on weak necessity in which a conclusion is evaluated as necessary because it refers only to possibilities to which the premises refer

4. General discussion

Mental models were first proposed as a basis for deductive reasoning over forty years ago. Since then the theory has expanded to cover most areas of reasoning including induction and abduction (Johnson-Laird et al., 2015a). The present article has outlined a recent advance in the theory's account of sentential reasoning. We discuss the primary consequences of this account, potential objections to it, and its principal alternatives.

4.1. People draw modal conclusions from non-modal premises

The model theory differs from both logic and probabilistic logic. What underlies it are possibilities—not necessities, not truth values, and not probabilities. If a compound assertion has multiple models, they each represent epistemic possibilities that have the force of a conjunction. Hence, a premise such as:

34. The fault is in the software or in the connection, or both.

is true in the same circumstances as an exhaustive conjunction of the default possibilities:

35. Possibly the fault is in the software, possibly it is in the connection, and possibly it is in both.

The two assertions seem equivalent, and almost everyone infers each of the possibilities in (35) from the disjunction (34), which does not mention possibilities (Hinterecker et al., 2016). Yet the inferences of the three conjuncts in (35) are invalid in all of the infinitely many modal logics (Hughes & Cresswell, 2012) and in probabilistic logic too (Adams, 1998). They are invalid in modal logic because, for instance, it may be impossible that the fault is in the software. In this case, the conclusion (35) is false, but the premise (34) can still be true. This inference and the others can be proved in logic only with additional premises that guarantee that each conjunct is possible, and that one conjunct does not contradict the other. Hence, in logic, the inferences call for additional information to rule out impossibilities, whereas in the model theory, the inferences are made by default, and they call for additional information to block them. So, if one of the conjuncts is known to be impossible, or to contradict the other, modulation blocks the construction of the corresponding models. No other existing theory predicts these inferences.

4.2. Intuitions and deliberations emerge from two systems of reasoning

System 1 makes intuitive inferences, focusing on a single mental model at a time. In deliberative inferences, however, system 2 considers alternative models that are fully explicit. Many other theories depend on dual processes, and they have their critics (e.g., Keren & Schul, 2009). What is unique to the model theory is that the two systems are implemented in a computer model. Moreover, its two systems share many components in common. Hence, system 1's intuitive answers are based on some of the processes that system 2 also relies on. This sort of architecture differs from many other dual-system accounts.

4.3. Modulation predicts a priori truth and falsity

The model theory makes no use of logical form or formal rules of inference. Its computational implementation includes a parser that uses lexical meanings and grammatical rules to compose meanings and thereby to construct models. Knowledge can modulate the interpretation of assertions and in some cases recover a priori truth values, a most controversial matter in philosophy. Philosophers tend to accept that a logical tautology, such as:

36. No unmarried man is married

is true a priori. But a source of controversy is whether, as philosophers from Kant (1934) to Carnap (1947) have supposed, a claim, such as:

37. No bachelor is married

is true a priori on the basis of its meaning. Quine (1953, p. 23) argued to the contrary that example (37) is not true a priori, and that the distinction between such assertions that

seem true a priori and those that are contingent is "an unempirical dogma of empiricism." Not anymore. The empirical studies we have described show that individuals innocent of philosophical niceties judged that assertions can be true (or false) a priori as a result of their meaning.

In logic, if a material conditional is false then its *if*-clause is true. So a very short proof for the existence of God is sound in logic:

38. It is not the case that if God exists then atheism is correct.
    Therefore, God exists.

Its premise is true, and it implies both that God exists and that atheism is not correct. It therefore follows from this conjunction that God exists. In the model theory, a conditional's meaning is not a material implication, not a conditional probability, not a set of possible worlds, and not an inferential relation. It is instead a conjunction of possibilities, each of which is assumed in default of information to the contrary. And so the falsity of a conditional does not imply that its *if*-clause is true, which renders the "proof" in (38) invalid. Individuals judge that the following assertion is false:

39. If Sonia has pneumonia then she is healthy.

But its falsity does not imply that Sonia has pneumonia, and indeed individuals judge that it is possible that Sonia does not have pneumonia (Quelhas et al., 2016). Only one case is impossible:

```
Sonia has pneumonia  Sonia is healthy
```

That is why (39) is false. The modulation algorithm we described mirrors these evaluations.

Yet a complex sort of modulation is at present beyond the program. As Byrne (1989) showed, individuals draw their own conclusion from premises, such as:

42. If she meets her friend then she will go to a play.
    She meets her friend.

They infer that she will go to a play. But when the premises have a further conditional of the following sort added to them:

41. If she has enough money then she will go to a play.

reasoners tend not to make the inference (see also Byrne, Espino, & Santamaria, 1999). The additional premise reminds them of a necessary condition for going to a play: One needs money to pay for the tickets. But no premise has established this condition, and so they balk at the inference. The inference is complex, and the modulation algorithm has yet to capture it.

4.4. Accurate verification depends on counterfactual possibilities

A match between evidence and a specific compound's mental models suffices for the intuition that the compound is true. A definitive verification, however, depends on the counterfactual status of the other possibilities to which the compound refers. Their status often can be evaluated only from existing knowledge. But when they are true, the compound is true.

4.5. Objections to the model theory

The subtlest and most powerful objection to the model theory is implicit in Grice's (1989) defense of the material conditional and other logical meanings for compounds in daily life. He argued that conversation follows certain principles (or "maxims"), and so speakers can communicate more than they say; that is, they can create "conversational implicatures." ance, they say:

42. If she had the treatment then she'll get better.

The conversational implicature is that they don't believe that the *if*-clause is false. If they had believed it to be false, then they would not have used a conditional, because it would have been misleading. And if it is false, Grice claimed, then the conditional is true, because it is a material conditional. Conversational implicatures are not valid deductions, but are defeasible, and so speakers can cancel them without inconsistency, for example:

43. I'm not allowed to tell you whether or not she had the treatment, but if she did, she'll get better.

Grice proposed another sort of implicature, a "conventional" implicature, which derives from the meanings of words. In his example:

44. He is an Englishman, and therefore brave

the meaning of "therefore" implicates that the speaker takes bravery to follow from being an Englishman (Grice, 1989, p. 25). Because these conventional implicatures depend on meanings, they cannot be cancelled without inconsistency.

Grice's insight was that conversation creates defeasible inferences. It has led to many systems of pragmatics (e.g., Goodman & Frank, 2016; Sperber & Wilson, 1986). Yet his defense of material conditionals fails. Consider this conditional:

45. If the Government didn't cut interest rates then inflation will increase.

According to Grice, its utterance should imply that the Government is cutting interest rates. But the implicature can be cancelled if, for example, an economist tells you:

45'. It's good that the Government is cutting interest rates. If it didn't cut them then inflation will increase.

Her opening remark implies that she does not believe the *if*-clause of her subsequent conditional, and so her disbelief cancels the implicature. According to Grice, the conditional assertion in (45') is therefore a material conditional. Suppose, then, that the Government cuts interest rates. The *if*-clause is false, and so the material conditional is true whether or not inflation increases. (If only economic predictions were that easy!) The example shows that the paradoxes of material conditionals still occur even in the context of conversational implicatures. So Grice's defense of truth-functional semantics requires endorsement of the paradoxes and the counterintuitive proofs, such as the one for the existence of God, that follow from material conditionals.

A second objection to the model theory is analogous to the first. Some theorists have argued that the test-case inferences of the sort:

A or B or both.
Therefore, possibly A.
Therefore, possibly B.
Therefore, possibly A and B.

are conversational implicatures (e.g., Sauerland, 2004). Other linguists, as we mentioned earlier, treat them as consequences of the semantics of compounds (e.g., Geurts, 2005; Zimmermann, 2000). Neither approach is quite right. Conversational implicatures fail for the reason we have just described. And the possibilities cannot be part of a definitive semantics of compounds, because the inferences are invalid (as we described earlier in this section). The model theory has the best of both worlds: The meaning of "or" yields possibilities as defaults, just as the meaning of "bird" entails that it flies as a default. Perhaps this analysis is analogous to treating the inferences as implicatures, but it depends on a different semantics, and one that is not truth functional.

A third objection to the model theory is that compounds refer, not to conjunctions of possibilities, but to disjunctions of them. Conjunctions imply disjunctions, but the converse does not hold. And disjunctions of possibilities cannot explain the test-case inferences. If an inclusive disjunction, *A or B*, implied only:

Possibly A or possibly B or possibly both.

then reasoners should never draw the inference that *A* is possible. Modal logics are correct to categorize such inferences as invalid. What, then, is the mechanism yielding these inferences? At present only the model theory has an answer, and it treats them as following from a conjunction of possibilities that each holds in default of information to the contrary.

The model theory generalizes beyond sentential reasoning to other domains, such as causal, deontic, relational, and quantified reasoning (Bucciarelli et al., 2008; Goodwin & Johnson-Laird, 2005; Khemlani, Barbey, & Johnson-Laird, 2014; Khemlani & Johnson-Laird, 2013b). A description of relations, such as:

46. Ann is better than Beth, and Beth is worse than Cath

is indeterminate, and so it refers to a conjunction of possibilities, which yield conclusions, such as:

47. Possibly Ann is better than Cath, and possibly Ann is worse than Cath.

Likewise quantified premises in a syllogism such as:

48. All the drivers are parents.
    All the drivers are beekeepers.

have a model of a conjunction of possible individuals that yields such conclusions as:

49. Possibly all the parents are beekeepers, and possibly all the beekeepers are parents.

In fact, a previous study has corroborated the occurrence of such inferences (e.g., Evans, Handley, Harper, & Johnson-Laird, 1999).

4.6. Alternative accounts of sentential reasoning

Could some alternative to the model theory of reasoning turn out to be correct? The answer has to be "yes," because there are infinitely many ways to compute any computable function. There are two major candidates: sentential logic and probability logic. Sentential logic with modal operators such as "possibly" runs into severe difficulties. It fails to predict inferences that people consider obvious, such as the test cases from nonmodal premises to modal conclusions (6). And it predicts inferences that people do not make, such as inferences from exclusive to inclusive disjunctions (10). It is monotonic; and its semantics for conditionals yields bizarre paradoxes (8).

Probability logic fares better. But, it does predict inferences at which most people balk, such as those of the following sort:

A
Therefore, A or B, or both.

It also fails to predict inferences that they do make, such as the greater preponderance of inferences from inclusive to exclusive disjunctions than vice versa, and inferences from *A or B or both* to *if not A then B*. And its principal formulation is monotonic (Baratgin et al., 2015). In fact, quite what semantics the probabilistic approach assigns to compound assertions other than conditionals is unknown. How, for instance, does it account for inferences from disjunctions to conjunctions of possibilities? It needs to rule out their invalidity before it can assign p-validity to their conclusions.

Any alternative to the model theory would need to predict that inferences from multiple possibilities are harder than those from single possibilities. It would need to account for the other phenomena summarized in Table 3. And it would need to represent

possibilities, allow meaning and knowledge to modify the interpretation of compounds, deliver a priori truth values, and deal with defeasible inferences. It is not impossible for an alternative to meet these goals—indeed, Koralus and Mascarenhas (2013) developed an "erotetic" theory of reasoning, which integrates model-based reasoning with the sentential calculus—but the outcome might resemble the model theory more than any modal or probabilistic logic.

4.7. Conclusion

Everyday reasoning concerns what is possible, even when premises make no mention of the matter. Possibilities are the simplest uncertainties, and they accommodate probabilities. The previous neglect of possibility allowed psychologists to overlook the corresponding gap in their theories based on logic, probabilistic logic, and mental models. The present efforts to fill this gap have bolstered the discovery that reasoning in daily life depends on models of possibilities.

Acknowledgments

We presented some aspects of the present theory to London Reasoning Workshops at Birkbeck College London in 2015 and 2017; at the conference in memory of Vittorio Girotto, at University College London in 2016; at the CONCATS seminar in New York University in 2017; and at the Cognitive Science Conference in 2017. We are grateful to many colleagues for help and advice. They include fellow researchers into mental models, critics of the theory, and some innocent bystanders—the list is too long to cite. But, we offer a special thanks to the late Vittorio Girotto for his brilliant contributions to the model theory. We dedicate this paper to his memory.

References

Adams, E. W. (1998). *A primer of probability logic*. Stanford, CA: Center for the Study of Language and Information.

Aho, A. V., & Ullman, J. D. (1972). *The theory of parsing, translation, and compiling. Vol I: Parsing*. Englewood Cliffs, NJ: Prentice-Hall.

Anderson, J. R. (2007). *How can the mind occur in the physical universe?* New York: Oxford University Press.

Baratgin, J., Douven, I., Evans, J. St. B. T., Oaksford, M., Over, D., Politzer, G., & Thompson, V. (2015). The new paradigm and mental models. *Trends in Cognitive Sciences*, *19*, 547–548.

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

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

Bucciarelli, M., & Johnson-Laird, P. N. (2005). I deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology*, *50*, 159–193.

Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision Making*, *3*, 121–139.

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

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT.

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

Byrne, R. M. J., & Johnson-Laird, P. N. (2009). 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences*, *13*, 282–286.

Byrnes, J. P., & Beilin, H. (1991). The cognitive basis of uncertainty. *Human Development*, *34*, 189–203.

Carnap, R. (1947). *Meaning and necessity: A study in semantics and modal logic*. Chicago, IL: University of Chicago Press.

Chater, N., & Oaksford, M. (2009). Local and global inferential relations: Response to Over (2009). *Thinking & Reasoning*, *15*, 439–446.

Cook, S. A. (1971). The complexity of theorem proving procedures. *Proceedings of the Third Annual Association of Computing Machinery Symposium on the Theory of Computing*, 151–158.

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

Cresswell, M. J., & Hughes, G. E. (2012). *A new introduction to modal logic*. Routledge.

Cruz, N., Baratgin, J., Oaksford, M., & Over, D. E. (2015). Bayesian reasoning with ifs and ands and ors. *Frontiers in Psychology*, *6*, 109.

Cruz, N., Over, D., & Oaksford, M. (2017). The elusive oddness of or-introduction. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 259–264). Austin, TX: Cognitive Science Society.

Espino, O., & Byrne, R. M. J. (2013). The compatibility heuristic in non-categorical hypothetical reasoning: Inferences between conditionals and disjunctions. *Cognitive Psychology*, *67*, 98–129.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Evans, J. St. B. T. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, *18*, 5–31.

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

Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford, UK: Oxford University Press.

Geurts, B. (2005). Entertaining alternatives: Disjunctions as modals. *Natural Language Semantics*, *13*, 383–410. https://doi.org/10.1007/s11050-005-2052-7.

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

Goodman, N., & Frank, M. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*, 818–829.

Goodwin, G. P. (2014). Is the basic conditional probabilistic? *Journal of Experimental Psychology: General*, *143*, 1214–1241.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, *112*, 468–493.

Goodwin, G. P., & Johnson-Laird, P. N. (in press). The truth of conditional assertions. Cognitive Science.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2016). Modality, probability, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Perception*, *42*, 1606–1620.

Hughes, G. E., & Cresswell, M. J. (1996). *A new introduction to modal logic*. London: Routledge.

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

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

Johnson-Laird, P. N. (1975). Models of deduction. In R. Falmagne (Ed.), *Reasoning: Representation and Process* (pp. 7–54). Springdale, NJ: Erlbaum.

Johnson-Laird, P. N. (2006). *How we reason*. New York: Oxford University Press.
Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678.
Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. S. (1992). Propositional reasoning by model. *Psychological Review*, *99*, 418–439.
Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, *111*, 640–661.
Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015a). Logic, probability, and human reasoning. *Trends in Cognitive Sciences.*, *19*, 201–214.
Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015b). Response to Baratgin et al.: Mental models integrate probability and deduction. *Trends in Cognitive Sciences*, *19*, 548–549.
Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review*, *106*, 62–88.
Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, *71*, 191–229.
Johnson-Laird, P. N., & Steedman, M. J. (1978). The psychology of syllogisms. *Cognitive Psychology*, *10*, 64–99.
Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology*, *82*, 367–373.
Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, *1*, 134–148.
Juhos, C., Quelhas, A. C., & Johnson-Laird, P. N. (2012). Temporal and spatial relations in sentential reasoning. *Cognition*, *122*, 393–404.
Kant, I. (1934). *Critique of pure reason*. Trans. J. M. D. Meiklejohn, New York: Dutton. (Originally published 1781).
Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*, 533–550.
Khemlani, S. (2018). Reasoning. In S. Thompson-Schill (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*. Wiley & Sons.
Khemlani, S., Barbey, A. K., & Johnson-Laird, P. N. (2014). Causal reasoning: Mental computations, and brain mechanisms. *Frontiers in Human Neuroscience*, *8*, 1–15. https://doi.org/10.3389/fnhum.2014.00849.
Khemlani, S., & Johnson-Laird, P. N. (2009). Disjunctive illusory inferences and how to eliminate them. *Memory & Cognition*, *37*, 615–623.
Khemlani, S., & Johnson-Laird, P. N. (2011). The need to explain. *Quarterly Journal of Experimental Psychology*, *64*, 276–288.
Khemlani, S., & Johnson-Laird, P. N. (2012). Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta Psychologica*, *139*, 486–491.
Khemlani, S., & Johnson-Laird, P. N. (2013a). Cognitive changes from explanations. *Journal of Cognitive Psychology*, *25*, 139–146.
Khemlani, S., & Johnson-Laird, P. N. (2013b). The processes of inference. *Argument and Computation*, *4*, 4–20.
Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2012). The probability of unique events. *Plos-ONE*, *7*, 1–9. Online version.
Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2015). Naive probability: Model-based estimates of unique events. *Cognitive Science*, *39*, 1216–1258.
Khemlani, S. S., Mackiewicz, R., Bucciarelli, M., & Johnson-Laird, P. N. (2013). Kinematic mental simulations in abduction and deduction. *Proceedings of the National Academy of Sciences*, *110*, 16766–16771.
Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2014). The negations of conjunctions, conditionals, and disjunctions. *Acta Psychologica*, *151*, 1–7.

Koralus, P., & Mascarenhas, S. (2013). The erotetic theory of reasoning: Bridges between formal semantics and the psychology of deductive inference. *Philosophical Perspectives*, *27*, 312–365.

Marek, V. W., & Truszczynski, M. (2013). *Nonmonotonic logic: Context-dependent reasoning*. New York: Springer Science & Business Media.

Meyer, J.-J. C., & van der Hoek, W. (1995). *Epistemic logic for AI and computer science*. Cambridge, UK: Cambridge University Press.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality*. Oxford, UK: Oxford University Press.

Orenes, I., & Johnson-Laird, P. N. (2012). Logic, models, and paradoxical inferences. *Mind & Language*, *27*, 357–377.

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

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, *15*, 431–438.

Pfeifer, N. (2013). The new psychology of reasoning: A mental probability logical perspective. *Thinking & Reasoning*, *19*, 329–345.

Pfeifer, N., & Kleiter, G. D. (2009). Framing human inference by coherence based probability logic. *Journal of Applied Logic*, *7*, 206–217.

Piéraut-Le Bonniec, G. (1980). *The development of modal reasoning: Genesis of necessity and possibility notions*. New York: Academic Press.

Quelhas, A. C., & Johnson-Laird, P. N. (2016). The modulation of disjunctive assertions. *Quarterly Journal of Experimental Psychology*, *70*, 703–717.

Quelhas, A. C., Johnson-Laird, P. N., & Juhos, C. (2010). The modulation of conditional assertions and its effects on reasoning. *Quarterly Journal of Experimental Psychology*, *63*, 1716–1739.

Quelhas, A. C., Rasga, C., & Johnson-Laird, P. N. (2017). A prior true and false conditionals. *Cognitive Science*, *41*, 1003–1030.

Quine, W. V. O. (1953). Two dogmas of empiricism. In W. V. O. Quine (Ed.), *From a logical point of view* (pp. 20–46). Cambridge, MA: Harvard University Press. (Originally published, 1951)

Ragni, M., Kola, I., & Johnson-Laird, P. N. (2017). The Wason selection task: A meta-analysis. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 980–985). Austin, TX: Cognitive Science Society.

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

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, *27*, 367–391.

Schroyens, W. (2010). Mistaking the instance for the rule: A critical analysis of the truth-table evaluation paradigm. *Quarterly Journal of Experimental Psychology*, *63*, 246–259.

Sophian, C., & Somerville, S. C. (1988). Early developments in logical reasoning: Considering alternative possibilities. *Cognitive Development*, *3*, 183–222.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford, UK: Basil Blackwell.

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

Steinberg, D. D. (1970). Analyticity, amphigory, and the semantic interpretation of sentences. *Journal of Verbal Learning and Verbal Behavior*, *9*, 37–51.

Steinberg, D. D. (1972). Truth, amphigory, and the semantic interpretation of sentences. *Journal of Experimental Psychology*, *93*, 217–218.

Steinberg, D. D. (1975). Semantic universals in sentence processing and interpretation: A study of Chinese, Finnish, Japanese, and Slovenian speakers. *Journal of Psycholinguistic Research*, *4*, 169–193.

Wason, P. C., & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition*, *3*, 141–154.

Wason, P. C., & Johnson-Laird, P. N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, *61*, 509–515.

Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics*, *8*, 255–290.