2. Modal logics

2.1. The fundamentals

Modal logic is the logic of possibility and necessity, and it goes back to the founder of Western logic, Aristotle (see his *Prior Analytics* in Barnes, 1984). Not all logicians accept modal logic. Frege (1967) in formulating what is now known as the "predicate" calculus rejected it, because possibilities depend on knowledge. His contemporary, Peirce, also formulated his version of the predicate calculus, which included a modal logic (Peirce, 1931-1958, see, e.g., Vol. 5, paragraph 510). In the twentieth century, logicians explored various sets of axioms yielding different modal logics (e.g., Lewis & Langford, 1932). They also distinguished between proof theory, which is a formal system for deriving proofs, and semantic theory, which captures the meanings of logical terms and the concept of validity. A valid inference is one for which the conclusion is true in all cases in which the premises are true (Jeffrey, 1981, p. 1). The distinction between proof theory and semantic theory is pertinent to modal logics. They have burgeoned since logicians discovered that different assumptions about "possible worlds" semantics correspond to different axioms for modal logics (Kripke, 1963). Interest in them has spilled over into artificial intelligence (Meyer & van der Hoek, 1995) and into systems of *nonmonotonic* reasoning, which, unlike orthodox logic, allow conclusions to be withdrawn (McDermott, 1982).

Proof theory depends on a set of axioms and at least one formal rule of inference, or on an equivalent set of formal rules of inference. An elegant method uses trees in proofs, as in Jeffrey's (1981) account. Its customary strategy is to conjoin the premises with the negation of the conclusion to be proved, and to explore all of the different alternative consequences in a tree. If they each yield a contradiction, then the conclusion follows from the premises; otherwise, it does not (see, e.g., Girle, 2009). At the heart of logic is the sentential calculus, which concerns inferences based on idealized versions of the connectives: "if", "or", and

"and". Many modal logics likewise focus on modal versions of the sentential calculus. The formal system adds axioms or rules of inference for the modal operators, *necessary* and *possible* to those of the sentential calculus. The semantic system is based on "possible worlds", in which each possible world fixes the truth value of any proposition. The assertion:

5. ◇A in world w

denotes *possibly A* holds in world *w*, and it is true if *A* is true in at least one possible world relevant to *w*. And the assertion:

6. □A in world w

denotes *necessarily A* holds in world *w*, and it is true if *A* is true in all possible worlds relevant to *w*. The real world is a member of the set of possible worlds, and the relation of *relevance* (aka accessibility) has different meanings that correspond to different axiomatizations of modal logics. Hence, a relevant world from *w* can mean a world that is conceivable from *w* given knowledge (for epistemic logics), one that occurs in a future from *w* (for temporal logics), or one that has the same principles for what is permissible or obligatory as *w* (for deontic logics).

The simplest modal logic is system K (for Kripke), which underlies an infinite family of normal modal logics. Like all normal modal logics, system K treats the two modal operators as
interdefinable using negation:

7. A is possible if and only if it isn't necessary that not A.
   A is necessary if and only if it isn't possible that not A.

System K posits the following rule of inference, which can be expressed instead as an axiom:

8. □(A → B)
   Therefore, (□A → □B).

where "→" denotes the material conditional of logic, which is similar to (but not the same as) *if A then B*. We can therefore paraphrase the rule above as *the necessity of if A then B* implies that *if A is necessary then B is necessary*. There are many other equivalent ways for formulating this axiom. System K, however, is an implausible candidate for any sort of human reasoning. That is because an assertion that is necessarily the case *is* the case:

9. Axiom T: $\Box A \rightarrow A$

But, this principle cannot be proved in system K. So, it is added as an axiom to K to create system T. The axiom corresponds to the semantic assumption that each possible world is relevant to itself, i.e., the relation of relevance is reflexive. Hence, what is necessary in world *w* holds in world *w*.

Logicians have formulated deontic logics (e.g., van Fraasen, 1973; Lewis, 2000). A principle that they need is that any obligatory action is permissible:

10. $\Box A \rightarrow \Diamond A$ – *if A is deontically necessary, then A is deontically possible.*

This principle added to system K, but without axiom T, yields system D, a deontic logic.

In colloquial English, an assertion may contain several modal operators, e.g.:

11. Perhaps it is possible that it may rain.

It could be symbolized as:

12. $\Diamond\Diamond\Diamond$(it rains).

But, suppose that what is possibly possible is merely possible. This principle is embodied in System S5, which is based on T, and is therefore in K's family. S5 collapses any string of modal operators into the final operator in the string, e.g.:

13. $(\Diamond\Diamond\Diamond A) \rightarrow (\Diamond A)$

As a consequence, there are just six modalities in system S5: $\Box A$, $\Diamond A$, A, and their respective negations. A modal logic can allow the collapse of three operators into two, but not of two into one. This principle can be extended upwards, so that it is legitimate to reduce n operators to n - 1, but no further, where n is each natural number. It follows that there is a denumerable infinity of distinct modal logics. We pursue our description of modal logics no further, because system T seems the most likely candidate for possibilities, and it underlies the one attempt to base the psychology of reasoning on modal logic.

2.2. Osherson's theory based on system T

In a major pioneering study, Osherson (1976) proposed a psychological theory of modal reasoning based on system T, which includes K, i.e., any inference valid in K is also valid in T, though not vice versa. Osherson constructed a language that expresses the modalities of possibility, factuality, and necessity, but each sentence in the language contains only one modal operator. He formulated 24 rules of inference for this language, which included some rules from the standard sentential calculus, e.g.:

14. A → B, not-B ∴ not-A

A & B ∴ A

where "∴" stands for "therefore". As Osherson intended, the rules are incomplete, i.e., certain inferences valid in T cannot be proved in his theory, including an inference that features later in this article, and that is known as *or*-introduction: *A* ∴ *A or B or both*. The system also includes rules for defining one modal operator in terms of the other, and rules for inferences that system T implies, such as:

15. $\Box$(A & B) ∴ $\Box$A & $\Box$B

$\Box$A & $\Box$B ∴ $\Box$(A & B)

$\Diamond$A ∨ $\Diamond$B ∴ $\Diamond$ (A ∨ B)

$\Diamond$(A ∨ B) ∴ $\Diamond$A ∨ $\Diamond$B

where "∨" denotes an inclusive disjunction of the sort, *A or B or both*.

To assess his theory, Osherson carried out an experiment that used inferences corresponding to each of its 24 rules of inference and 48 arguments that required several steps using these rules. The high-school participants were assigned to different groups to evaluate subsets of the inferences, and they judged each inference as “the conclusion must be true” (an alethic necessity), “the conclusion does not follow”, or “cannot decide”; and they later put into order of difficulty those inferences that they had judged to be necessary. As Osherson (1976, p. 232) himself remarks, the results were not altogether successful. The participants did not accept all the inferences in a robust way. With hindsight, the experiment’s use of only valid inferences in system T also made it impossible to discover invalid inferences that people accept. We have more to say about such inferences later, but we now turn to the new theory of modality.

3. The model theory of modality

The first half of this section of the article outlines the interpretations of modal assertions in daily life, their underlying meanings that they share with assertions about probabilities, and their mental models. The second half of this section describes the theory of how models underlie modal reasoning.

3.1. *Three interpretations of modals*

The semantics of modality in natural languages has been analyzed in various disciplines: linguistics (e.g., Palmer, 2001), psycholinguistics (e.g., Miller & Johnson-Laird, 1976, Sec. 7.3.1), linguistic philosophy (e.g., White, 1975), formal semantics (e.g., Kratzer, 1991), and computational linguistics (e.g., Isard, 1975; Moens & Steedman, 1978). Our account of modals embodies principles that are common to these approaches, but it is independent of them, and it embodies some unique principles. Given the variety of analyses, the terminology is confusing and ambiguous. Rather than neologize, we use common terms, but readers should bear in mind that their meaning differs from ours in other works. Three main interpretations of “possible,” “necessary,” and their cognates occur in daily life, and in this section, we establish that they are distinct, but propose a common underlying semantics for them. That is, we argue, not that modal terms are ambiguous, but rather that their interpretation depends both on the grammar of sentences in which they occur and on the background knowledge that they elicit.

The first interpretation is *deontic*, and refers to what is permissible and what is obligatory, and their respective negations, e.g.:

16. It is possible for you to stay

signifying that you are allowed to stay. Deontic assertions rest on background knowledge of moral codes, laws, rules, and social conventions. A deontic interpretation is feasible for assertions using an infinitival clause, as in (16), to describe an individual’s actions in a social context. But, a modal auxiliary can have the same interpretation:

17. You may stay.

As in modal logic, deontic possibility and necessity are interdefinable using negation, e.g.:

18. It is possible for you to stay: it is not necessary for you not to stay.

and:

19. It is necessary for you to stay: it is not possible for you not to stay.

One key feature of these interpretations is that “permissible” can replace “possible,” and “obligatory” can replace “necessary,” without changing meaning. Another distinctive feature concerns counterexamples. A counterexample to an obligation violates it, but has no effect on the truth of the obligation (Wertheimer, 1972).

The second interpretation of modals is *alethic*, and typically occurs in assertions about inferences, e.g.:

20. It follows as a necessity that the object’s velocity is constant.

21. It follows as a possibility that the object’s velocity is constant.

These claims can be expressed using adverbs (or modal auxiliaries):

22. Necessarily, it follows that the object's velocity is constant.

23. Possibly, it follows that the object's velocity is constant.

They can also be expressed with modal auxiliaries:

24. It follows that it must be that the object's velocity is constant.

25. It follows that it can be that the object's velocity is constant.

Like deontic modals, alethic necessity and possibility are interdefinable, though the negations can make the equivalence hard to grasp (see Osherson, 1976). A counterexample to a claim about an alethic necessity shows that the claim itself is false: (22) is false given a case in which the consequence does not follow, and (23) is false given that it is impossible for the object's velocity to be constant. The body of knowledge relevant to alethic interpretations concerns general principles of inference, argument, and analysis. Likewise, alethic modal logics often concern provability, and allow that any case that is not logically self-contradictory is at least possible. A key test for an alethic necessity is that it implies the truth of its complement.

The third interpretation of modals is *epistemic*. It concerns the use of knowledge to assess events and propositions. An instructive contrast occurs between these two sorts of assertion (e.g., White, 1975):

26. It is possible that Viv will win the same: Viv may win the game.

and:

27. If is possible for Viv to win the game: Viv can win the game.

The first assertion (26) contains a *that*-complement, which refers to a proposition, whereas the second assertion (27) contains an infinitival complement, which does not refer to a proposition. In (27), it refers to Viv's ability to win the game in the relevant circumstances. In other assertions, infinitivals refer to propensities and other abstractions. A proposition is an expressible entity that can be true or false. So, it is acceptable to assert:

28. It is true that Viv will win the game.

but it is not acceptable to assert:

29. It is true for Viv to win the game.

An infinitival complement is akin to a noun phrase, which can be the subject of a sentence:

30. For Viv to win the game is possible.

Assertions that assert the possibility of a proposition (e.g., 26) imply assertions about the possibility of the infinitival abstraction (e.g., 27). But, the converse implication does not hold, and so the following assertion is not a self-contradiction:

31. It's possible for Viv to win the game, but it is not possible that Viv will win the game.

Readers might therefore suppose that two different senses of "possible" are in play, and that the one with an infinitival is an alethic possibility. The key, however, is the role of "necessary". The assertion:

32. It is necessary for Viv to win this game.

does not imply that Viv will win the game. Hence, the assertion is not alethic. Instead, it states a necessary condition for some other situation to hold, as in:

33. It is necessary for Viv to win the game if she is to win the tournament.

This analysis also applies to assertions with *that*-clauses:

34. It is necessary that Viv wins the game if she is to win the tournament.

Hence, we treat both these sorts of assertion – those with *that*-clauses and those with infinitivals – as instances of epistemic modals, *and we explain presently the root of their difference in the next section* (4.1.2).

What is striking about epistemic possibilities is their obvious kinship with subjective probabilities. Epistemic "possible" has a *scalar* interpretation in the following rank order scale:

35. impossible – just possible – possible – very possible – certain

It is comparable to the scale of non-numerical probabilities – a point perhaps first made by von Mises (1928, p. 62), and which we can label as:

36. impossible – nearly impossible – as probable as not – probable – highly probable – certain

Experimental evidence shows that "possible" tends to refer to a probability of around 50% (see, e.g., Juanich, Teigen, & Gourdon, 2013). A corollary is that a denial of an epistemic possibility, as in:

37. It is impossible that it will snow.

is equivalent to a reference to other end of the scale:

38. It is certain that it won't snow.

This usage contrasts with alethic modality, where *impossible* is equivalent to *necessarily not*. Likewise, as the scale in (35) shows, "possible" and "necessary" are not interdefinable in their epistemic senses. An assertion such as:

39. It is necessary that it rains

does not mean that it is bound to rain, but rather that rain is a necessary condition for some unstated situation. We can make it explicit in an infinitival:

40. It is necessary that it rains for the crops to grow.

There are usages in which epistemic necessity suggests a scale. The assertion:

41. It is hardly necessary that he goes to a doctor (to get better).

signifies that for him to get better, it is hardly necessary to go to the doctor. The non-numerical scale is:

42. not necessary – hardly necessary – as necessary as not – almost necessary – necessary

Such a scale concerns the strength of the reasons for an epistemic necessity. So, unlike alethic necessity, epistemic necessity does not make a factual claim: *it must rain* does not imply than *it will rain* (Steedman, 1977).

In system S5, as we saw in the section on modal logic, any string of modal operators collapses to leave only the final one. Everyday modality, however, is different. In the case of epistemic claims, such as:

43. Perhaps, it may be possible that Hillary runs again.

the three modals – "perhaps", "may" and "possible" – suggest tentativeness about the proposition. They move it towards the less probable end of the scale, which can also represent non-numerical probabilities. The assertion:

44. Possibly, it's necessary that Hillary does not run again.

expresses the possibility of an epistemic reason that Hillary does not run again. Unlike, system S5, these modal terms cannot be simplified.

It can be difficult to determine which interpretation of a modal assertion is appropriate. Deontic interpretations can be confirmed by a substitution of "permissible" for "possible", and of "obligatory" for "necessary" without a change in meaning. Alethic interpretations can be confirmed by testing for the interdefinability of "possible" and "necessary". The two terms are not interdefinable for epistemic assertions. But, in many cases, more than one interpretation is feasible. For example, a fire chief who tells the residents of a building:

45. You may go back into the building now

could be asserting both a deontic and an epistemic possibility. Such indeterminacies should occur if, as we now argue, modals share an underlying semantics in common. Table 1 summarizes the three sorts of modal interpretations, describes their main properties, and states some examples of sentences with their interpretations.

Table 1

The three principal modal interpretations, their key properties, and examples of assertions stating possibilities and necessities.

Sort of modal	Key properties	Typical examples of possibilities	Typical examples of necessities
Deontic	Asserts what is permissible, or impermissible. *Permissible* interdefinable with *obligatory*.	It is possible for Pat to leave.	It is necessary for Pat to leave.
Alethic	Asserts what follows from general principles of inference or analysis. *Possibly* interdefinable with *necessarily*.	Possibly, the object's velocity is constant.	Necessarily, the object's velocity is constant.
Epistemic	Asserts status of an event or proposition based on relevant knowledge. *Possible* interdefinable with *certain*. "Necessary" means that the proposition is a condition for some other proposition to hold.	It is possible for Viv to run a mile. It is possible that Viv will run a mile.	It is necessary for Viv to run a mile (in order to get into shape). It is certain that Viv will run a mile.

3.2. The meaning of modals

A crux is whether a modal term such as *possible* has different meanings as, say, *run* has,
or different interpretations of a single underlying semantics (Broome, 2016; Johnson-Laird, 1978; Kratzer, 1981). The model theory postulates that the syntax of sentences and *modulation* by contextual knowledge yield different interpretations of a single underlying semantics. Indeterminate interpretations occur when different bodies of knowledge can modulate the same assertion. That is why it can be difficult to decide which interpretations of a modal assertion are appropriate.

In modal logics, the meanings of modals depend on "possible worlds" (see section 2.1.). Each possible world fixes the truth value of any assertion about a world, and so, as Partee (1979) argued, possible worlds are far too big to fit inside anyone's head. Other sorts of semantics do exist, e.g., modal meanings can be topological spaces (see McKinsey & Tarski, 1944). Yet, even these alternatives are committed to infinite interpretations. Computer implementations of a possible-worlds semantics rely on the feasibility of demonstrating that an inference is invalid by constructing only finite counterexamples (e.g., Meyer & van der Hoek, 1995).

A simple assertion, such as:

46. The coin came down heads.

refers to an infinity of possibilities, depending on the size of the coin, its shape, its substance, its denomination, its different trajectories, its landing point, and so on and on. The model theory postulates a semantics that allows for an indefinite number of details, but that does not depend on them. This semantics underlies possibilities and probabilities. Consider any situation that individuals can think of as having a small finite number of exhaustive and mutually exclusive alternatives. A *possibility* in daily life refers to a subset of these alternatives. Of the infinite possibilities for a coin toss, individuals can conceive of two outcomes: heads or tails. They build one model for each outcome. It represents what all the two infinities of coin tosses have in

common: the coin comes down heads in one model, and it comes down tails in the other model.

It is the same for other events in daily life. Consider, as an example, the Iowa caucuses of February 1st 2016.   Figure 1 shows the three exhaustive and mutually exclusive possibilities.

Clinton wins

Sanders wins

O’Malley wins

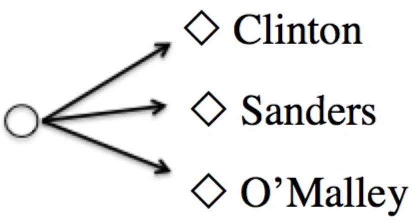

Figure 1. The three alternative possibilities for the Iowa caucuses February 1st 2016.

A possible event refers to a subset of the alternatives, e.g.:

47. Possibly, Clinton or Sanders will win the caucus.

An event that occurs necessarily refers to the complete set of alternatives:

48. Necessarily, Clinton, Sanders, or O’Malley will win the caucus.

On the observation of the actual results of the caucus, the status of the possibilities changes. Only one individual can win the caucus, so on the observation of the winner, one of the three possibilities becomes a fact, and the other two become counterfactual possibilities. There are also cases that are impossible, e.g., Elizabeth Warren could not have won the Iowa caucuses of 2016, because she was not a candidate. Figure 2 shows the change to the possibilities given the outcome of the caucuses.   Speakers can refer to the counterfactual cases using a counterfactual conditional (see Byrne, 2005), as in:

49. If Sanders had won in Iowa, he might have gone on to win the nomination.

Modal logics do not distinguish between counterfactual possibilities and real possibilities.

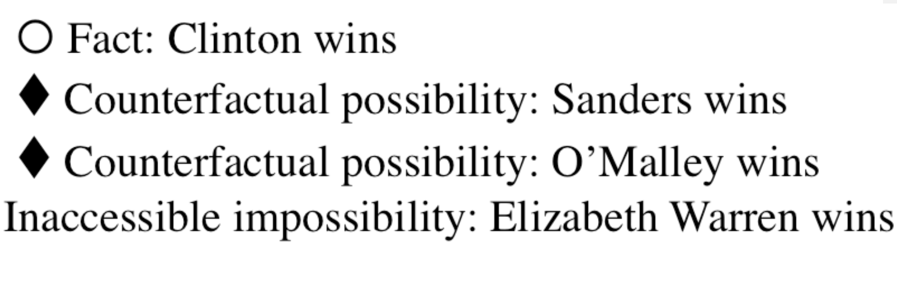

Figure 2. The status of possibilities on the outcome of the Iowa caucuses of February 1st 2016.

Hence, as Karttunen (1972) pointed out, there are discrepancies between logical but infelicitous descriptions, such as:

50. Clinton lost, but she may not have.

and counterfactual and felicitous descriptions:

51. Clinton lost, but she might not have.

The translation of this last assertion into a modal logic such as system T:

52. $C \,\&\, \Diamond \neg C$

where *C* stands for "Clinton lost", yields a self-contradiction, that is, an assertion that is false whatever the value of *C*. Yet, the assertion is not a self-contradiction in everyday discourse, and so such examples are another divergence of everyday discourse from modal logics.

The finite semantics underlies the different interpretations of modals, which result from modulation. The idea that knowledge modulates meanings is part of the original model theory, and it is implemented for sentential connectives in the mSentential program. For deontic interpretations, an assertion elicits a body of deontic knowledge that modulates the semantics. It ensures that alternatives refer to what is permissible, and that actions that violate an obligation do not refute it. The semantics underlies alethic assertions, but its meanings concern relations between one set of assertions (premises or the input to analyses) and another set (conclusions). There are five such relations, which we describe in outlining the theory of alethic reasoning. The semantics underlies epistemic modals. The alternatives concern models based on knowledge. For an epistemic possibility of the sort: *it is possible that A*, knowledge shows that *A* holds in at least one of alternative outcomes in the relevant situation. For an epistemic necessity of the sort: *it is necessary that A*, knowledge shows that for another proposition *B* to hold, *A* is a necessary condition: in all alternative outcomes in which *B* holds, *A* holds.

3.3. The mental models of modals

The theory calls for models to represent possibilities: "what underlies . . . meaning is the ability to envisage states of affairs that may or may not correspond to reality, that is, the ability to construct mental models of possible, hypothetical, and imaginary situations" (Johnson-Laird, 1983, p. 60). They can represent real possibilities and counterfactual possibilities, different sorts of discourse – factual, hypothetical, or fictional, and on occasion the probabilities of propositions. The theory therefore postulates that a system of symbols attached to models can signify their status, just as negation has to be represented in models by a symbol that has an associated semantics (e.g., Johnson-Laird & Byrne, 1991, p. 68-9). For simplicity, when the sort of possibility is clear in this article, we forego the use of labels.

The model theory postulates that compound assertions, such as disjunctions, refer to a conjunction of possibilities that hold in default of information to the contrary. As we showed earlier (in section 1), a disjunction such as:

53. Jared or Ivanka or both of them are in the office.

calls for a representations of its conjunction of its default possibilities for who is in the office:

Jared	
	Ivanka
Jared	Ivanka

These are mental models (in system 1), which represent only what is true in each possibility. Of course, there may be other individuals in the office, but these three mutually exclusive possibilities are exhaustive. There are two precursors to this account. Zimmermann (2000) proposed that deontic disjunctions refer to lists of alternatives in a "possible worlds" semantics, and Geurts (2005) extended this view to factual disjunctions. The model theory, however, applies its default analysis to all compound assertions (Johnson-Laird et al., 2015), and the present theory postulates a finite semantics instead of "possible worlds" (see section 3). To illustrate a modal assertion, consider:

54. It is possible that Donald is in the office.

Its mental models are:

Donald

…

where the ellipsis allows for other possible propositions, so these two cases are also exhaustive. If assertion (54) is a continuation of the description that begins with the disjunction (53), it calls for the conjunction of the two sets of models of possibilities, where possibilities other than Trump being in the office are now explicit:

Donald	Jared
	Jared

Donald		Ivanka
		Ivanka
Donald	Jared	Ivanka
	Jared	Ivanka

Because the rows in such models represent equipossible alternatives, the probability of a particular event, such as Ivanka being in the office, equals the proportion of such cases in the set: 2/3. Hence, the description implies that she is more likely to be in the office than Donald (see Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). Proportions of models can yield scalar possibilities, so the preceding model implies:

55. It is just possible that Donald, Jared, and Ivanka, are in the office.

Moreover, numerical tags can label each possibility in order to infer numerical probabilities (Johnson-Laird, et al., 1999).

A model of what is impossible is the complement of the set of models of possibilities, and so it can be inferred from the set and is not normally represented. An assertion of an impossibility, such as:

56. It is impossible that Melania is in the office

updates the preceding model by adding a representation of Melania as not in the office to each possibility that it contains. Any item common to all the cases in such a set is certain. So, it is certain that Melania is not in the office. A compound assertion, such as a disjunction, that is itself only a possibility:

57. It is possible that Donald or Ivanka or both of them are in the office.

has the following mental models:

Jared	
	Ivanka
Jared	Ivanka

…

The ellipsis denotes other possibilities so that the disjunction itself is only a possibility.

Conditionals of the sort, *If A then B*, unlike disjunctions, have a subordinate clause *if*-clause, and a main *then*-clause. One test for subordination is that a pronoun in the *if*-clause can refer forwards to a referent of the main clause, e.g.:

58. If he served tea, then the butler put on white gloves.

This "cataphoric" interpretation is implausible from one main clause to another, as so the following disjunction seems to refer to two different individuals:

59. He served tea or the butler put on white gloves.

Subordinate clauses have a major semantic consequence. They tend to be presupposed and taken for granted in evaluating a main clause. For conditionals, the subordinate role of the *if*-clause has a striking consequence. A conditional such as:

60. If Donald is in the office then Kellyanne is in the office.

refers to a conjunction of three possibilities that hold by default:

Donald in the office	Kellyanne in the office
¬ Donald in the office	¬ Kellyanne in the office
¬ Donald in the office	Kellyane in the office

They are exhaustive and so there is one impossibility:

Donald in the office	¬ Kellyanne in the office

However, the falsity of the conditional (60) refers to these possibilities:

Donald in the office	¬ Kellyanne in the office
¬ Donald in the office	¬ Kellyanne in the office
¬ Donald in the office	Kellyanne in the office

Hence, cases in which Donald is not in the office have no bearing on the truth or falsity of the conditional, because they are possible whether the conditional is true or false. The model theory accordingly captures the meaning of conditionals, not in a "truth" table, but, as Table 2 shows, in a "possibility" table for true and false conditionals. The table has several consequences, which we now consider.

Table 2

A possibility table for conditionals showing what is possible and what is not possible for each case in the partition for a true conditional and for a false conditional

The partition		Status of cases in the partition	
		If A then B is true	*If A then B* is false
A	B	Possible	Not possible
A	Not-B	Not possible	Possible
Not-A	B	Possible	Possible
Not-A	Not-B	Possible	Possible

One consequence is that the mental models of a conditional, unlike its fully explicit models, do not represent the cases in which its *if*-clause if false, because they are possible whether the conditional is true or false. For the conditional (60), the mental models of the people in the office are:

Donald Kellyanne

… {not Donald}

The ellipsis denoting other possibilities includes other propositions provided that they include the proposition that Donald is not in the office. Likewise, the truth or falsity of a conditional does not depend on these cases in which the *if*-clause of the conditional is false. Hence, the conditional (60) is true given two conditions: it is possible that Donald and Kellyanne are in the office, and it is impossible that Donald is in the office and that Kellyanne is not.

The model theory's account of the meaning of conditionals differs from that of the material conditional in sentential logic, and does not give rise to the latter's well-known paradoxes, which are inferences of these two sorts:

B.

Therefore, If A then B,

and:

Not-A.

Therefore, if A then B.

They are valid in the sentential logic, but most individuals reject them (Orenes & Johnson-Laird, 2012). Neither the premise, *B*, nor the premise, *not-A*, suffices to support the conditions above in which a conditional is true.

Another consequence of the possibility table is that it reconciles two well-established empirical findings. On the one hand, individuals tend to judge that cases in which the *if*-clause of a conditional is false as "irrelevant" to an evaluation of whether the conditional is true or false (e.g., Johnson-Laird & Tagart, 1969). Yet, on the other hand, they tend to include these cases when they are asked to list what is possible given a conditional (e.g., Barrouillet, Grosset, & Lecas, 2000). Both these phenomena are immediate consequences of the possibility table for conditionals.

The possibility table for conditionals makes no reference to probabilities, and so its semantics does not postulate that a conditional expresses a conditional probability (Adams, 1998). But, the probability of an assertion does not depend on those possibilities common to both its truth and falsity. And so Table 2 implies that the probability of a conditional *If A then B* equals the conditional probability of *B* given *A*, because the possibilities of *not-A* are common to the conditional's truth and falsity:

probability of (*if A then B*) = probability of ($B \mid A$).

This equation holds in probability-logic (Adams, 1998). So, epistemic possibilities have a major implication for probabilities.

We have now described the principles governing the models of epistemic possibilities. Table 3 illustrates the mental models and the fully explicit models for various sorts of assertion, including those containing sentential connectives.

4. The model theory of modal reasoning

Psychologists have studied deontic reasoning as embodied in Wason's (1966) well-known "selection" task in which participants select potential violators of a rule (see, e.g., Cheng and Holyoak, 1985; Cosmides, 1989; Griggs & Cox, 1982; Johnson-Laird, Legrenzi, & Legrenzi, 1972). The model theory explains performance in this task (Ragni, Kola, & Johnson-Laird, 2017a, b), and in deontic reasoning in general (Bucciarelli & Johnson-Laird, 2005). However, no previous theory of any sort accounts for inferences with alethic and epistemic modals, and so the model theory focuses on them.

The model theory's assumptions about mental representations drive many of its predictions about reasoning. The difference between mental models and fully explicit models,

Table 3

The mental models and fully explicit models of the possibilities for representative assertions of epistemic possibilities. Each set of models represents an exhaustive conjunction of mutually exclusive possibilities that hold in default of information to the contrary. The ellipsis, …, refers to other possibilities, and the symbol, ¬, refers to sentential negation. The other possibilities for models of conditionals also contain the proposition, ¬ A, as indicated by the parenthetical attached to the ellipsis.

Assertions	Mental models	Fully explicit models
	Possibilities	Possibilities
A	A	A

It is possible that A	A . . .	A . . .
It is not possible that A	¬ A . . .	¬ A . . .
A & B	A B	A B
It is possible that A & B	A B . . .	A B . . .
A or B	A B A B	A ¬ B ¬ A B A B
It is possible that A or B	A B A B . . .	A ¬ B ¬ A B A B . . .
If A then B	A B . . . {¬ A}	A B ¬ A ¬ B ¬ A B
It is possible that if A then B	A B . . . {¬ A} . . .	A B ¬ A ¬ B ¬ A B . . .

for example, predicts the occurrence of illusory inferences. A fundamental principle is that representations are parsimonious, because it is costly to hold them in memory and to process them. Deontic possibilities can be represented in the same way as other possibilities granted that the system mains a record that they are deontic rather than epistemic. We accordingly turn to a theory of reasoning for alethic modals, and then for epistemic modals.

4.1. Reasoning with alethic modals

The alethic evaluation of inferences is more refined than their binary evaluation as valid or invalid. Granted that a semantic relation exists between the premises and conclusion, an inference from one to the other depends on the relation between the conjunction of possibilities to which the premises refer (the premise possibilities) and the conjunction of possibilities to which the conclusion refers (the conclusion possibilities). Five set-theoretic relations can exist between any two sets, and they also hold in the present case, but the premises and conclusion refer to conjunctions of possibilities, and it is helpful to bear in mind that one possibility can imply another. The possibility of:

A B C

implies the possibility of, for example:

A C

The five relations are as follows:

1. The premise and conclusion possibilities are identical, and so the conclusion is *necessary*.
2. If the premise possibilities imply each of the possibilities to which the conclusion refers, but the premise refers to other possibilities too, then the conclusion is *possible*. Inexpert reasoners may infer that the conclusion is necessary because the premises support each of its possibilities, but this evaluation is a mistake, which we refer to as "weak" necessity.
3. If the premise possibilities imply only possibilities to which the conclusion refers, but
   the conclusion refers to other possibilities too, then the conclusion is *possible*.
4. If some of the premise possibilities imply some of the conclusion possibilities, then the conclusion is *possible*.
5. If the premise possibilities support none of the conclusion possibilities, and vice versa, then the conclusion is *impossible*. It contradicts the premises.

Of course, the premises and conclusion may have no semantic relation to one another, e.g.:

61. It is snowing and it is below freezing.
    Therefore, the number is prime.

The two sets of possibilities are disjoint, but in the absence of any further context, the only conclusion is that the conclusion is consistent with the premise, and therefore it is possible. It would be misleading to assert that it follows as a possibility from the premise. The mSentential program implements these principles in a parsimonious way, and Table 4 illustrates its account.

This account makes three main predictions about alethic possibilities. First, naïve individuals should be able to identify putative conclusions as possible, necessary, and impossible. Second, because alethic possibility is interdefinable with alethic necessity, the model theory's assumption that model represent possibilities rather than necessities predicts an interaction: people should be faster to evaluate a conclusion as possible than as necessary, whereas they should be faster to evaluate a conclusion as not necessary (one possibility suffices) than as not possible (all possibilities suffice). Third, because individuals rely on mental models, they should be susceptible to alethic illusions. We examine the evidence for these predictions in section 5.1.

Table 4

Illustrations of the three sorts of alethic conclusion given that there is a semantic relation between the premises and conclusion.

The alethic status of the conclusion given the premises	The relation between the premise and conclusion possibilities	Examples of mental models of premises and conclusion		Examples of inferences (given that modulation has not affected interpretations of the assertions)
		Premise models	Conclusion models	
1. Necessary	Premise possibilities imply each and only the conclusion possibilities	A B A C A B C	A C	*A and B. ∴ A is necessary.* *A. If A or B then C.* *∴ C is necessary.*
2. Impossible	No possibility is common to both the premises and conclusion.	A B A B	Not-A not-B	*A or B, or both.* *∴ not-A and not-B* is *impossible*.
3. Possible	Any other case, e.g., premises and	A	A B	*A.*

	conclusion have at least one possibility in common.	
		
		A
		    B
		A B
			A B
			
			A
			    B
				∴ *A or B or both* is possible.
				
				A or B or both.
				∴ *A or B but not both* is *possible**.

* An inference of weak necessity (see the text).

4.2. Epistemic modals in reasoning

The model theory of epistemic reasoning postulates that compound assertions, such as conditionals and disjunctions, refer to *conjunctions* of possibilities in default of information to the contrary (see 3.2). Only categorical assertions, *A*, and conjunctive assertions, *A and B*, refer to what is common to all possibilities, and therefore make factual claims. As an example of the conjunctive possibilities to which a disjunction refers, consider this assertion:

62. The flaw is in the software or in the cable, or both.

It refers to the following mental models of a conjunction of possible locations of the flaw in default of information to the contrary:

software	
	cable
software	cable

Because these models are exhaustive, they yield a model of what is impossible:

not-software not-cable

Readers might suppose that these models are equivalent to a truth table. But, a point that bears emphasis is that the entries in a truth table are not in a conjunction. They are mutually exclusive alternatives, e.g., the truth of *A* cannot be conjoined with the falsity of *A* without contradiction. In contrast, the three possibilities above are in a single conjunction that is not self-contradictory. It conjoins three default possibilities. System 1 considers the possible locations of the flaw one at a time, and so reasoners should therefore infer each of the following conclusions:

63. It is possible that the flaw is in the software.

It is possible that the flaw is in the cable.

It is possible that the flaw is in the software and in the cable.

These three inferences are invalid in all modal logics – another reason why the partition of possibilities and impossibilities is not equivalent to a truth table. For example, the first of these inferences is invalid in modal logics because it could be impossible that the flaw is in the software. Yet the disjunction would still be true in modal logics if the flaw *is* in the cable. And so the premise is true, but the conclusion is false. The only way to convert the inference into a valid one in modal logics is to treat it as an enthymeme, i.e., as an inference that requires a hitherto missing premise, namely:

64. It is not impossible that the fault is in the software.

But, this premise is equivalent to the conclusion to be proved, and so it makes the inference valid but circular. What the missing premise should be to avoid circularity is a puzzle. Which in itself suggests that a logical treatment of the inference is implausible. In sum, modal logics and mental models differ in a striking way. Modal logics call for additional premises in order to make these inferences: they do not call for additional premises to prevent them. Mental models make them at once by default: they call for additional premises to prevent them.

The denials of defaults have an interesting property. Consider, for example, what you assume by default about dogs:

65. They have four legs, fur, a tail that they can wag, they bark, and so on.

So, if you learn that a friend has acquired a dog, Fido, you can infer such properties by default. But, if you then learn that Fido is three-legged, bald, tailless, and mute, you withdraw these properties without giving up on Fido's doggyhood. Denial treats the default properties as though they were in a disjunction. This principle applies to the default possibilities to which compound assertions apply. The one exception is that there does come a point when a compound ceases to be true. Suppose you believe the disjunction (62) about the flaw being in the software or the cable. You then learn that is impossible that the flaw is in the software, or false as a matter of fact, then this information eliminates all the matching possibilities assumed in default of such evidence to the contrary. Hence, only one possibility is left from the disjunction above for the location of the flaw: it is in the cable, which is a necessary conclusion. In other words, a conjunction of default assumptions has the consequence that when one of default is denied, it is as though the assumptions were in a disjunction. Such behavior may seem bizarre, but it is a consequence of defaults, and therefore commonplace in object-oriented programming languages, such as Python.

Mental models are parsimonious. The modal theory extends this principle to epistemic possibilities: system 1 considers at most a single possibility at a time, and so it condenses a set of consistent possibilities into a single default possibility. Suppose, for example, a premise describes two separate possibilities:

66. It is possible that Tom is single and it is possible that Pat is married.

In default of information to the contrary, system 1 constructs a model of a single possibility:

Tom single Pat married

As a consequence, individuals should infer:

67. Therefore, it is possible that Tom is single and that Pat is married.

The inference is invalid in all modal logics because one event could be inconsistent with the other. The model theory predicts that modulation should block the conjunction of two possibilities into one given inconsistent possibilities, such as:

68. It is possible that Tom is single and it is possible that he is married.

Reasoners know that one person cannot be both single and married, and so they should not infer:

69. It is possible that Tom is single and that he is married.

When a premise asserts an epistemic possibility, such as:

70. It is possible that it is freezing.

a single model of an implicit possibility is added to its mental model (see Table 3), which is shown here as an ellipsis:

freezing

. . .

It follows that the premises:

71. It is possible that it is freezing.

If it is freezing then it is foggy.

yield mental models of freezing and foggy weather, and a model of an additional but implicit possibility. Reasoners should therefore draw the conclusion:

72. It is possible that it is foggy.

This same principle of "epistemic conclusions" applies when a compound assertion itself is asserted as a possibility:

73. It is possible that if it is freezing then it is foggy.

Assertions of the epistemic necessity of propositions, such as:

74. It is necessary that it rains.

are, as we showed earlier, refer to a necessary condition for some other proposition, e.g.:

75. It is necessary that it rains for the crops to grow.

They refer to the following possibilities:

rains	crops grow
rains	not crops grow
not rains	not crops grow

So, "necessity" is an emergent property from models of possibilities. It follows that necessity is not normally represented as such in models. Given premises such as the following ones,

individuals can draw a conclusion:

76. It is necessary that it is freezing.

If it is freezing then it is foggy.

But, because necessity is not normally represented in models, the theory predicts that reasoners should tend to draw a categorical conclusion:

77. Therefore, it is foggy.

In particular, they should be less likely to draw a modal conclusion of necessity than in the case of inferences about possibilities (as in 71).

To summarize: the model three makes three main predictions about epistemic reasoning. First, compound assertions such as conditionals and disjunctions imply conclusions about epistemic possibilities corresponding to those to which they refer by default. But, epistemic conclusions that refer to a possibility that the premises do not imply should be rejected. Second, epistemic possibilities are parsimonious, and so separate consistent possibilities should be condensed into one. Third, an assertion of an epistemic possibility creates a context that leads to inferences of a possibility, whereas an epistemic necessity has no such effect because necessary conditions are not normally represented as such in models.

Table 5 summarizes the six predictions of the model theory: three predictions for alethic reasoning, and three predictions for epistemic reasoning. We now turn to tests of these predictions.

Table 5

The predictions of the model theory for alethic and epistemic modal reasoning

Principles	Illustrative predictions
Alethic modals	
1. Models represent what's possible rather than what's necessary.	The alethic interaction: possible conclusions should be easier to infer than necessary ones; the difference should switch round for denials.
2. Alethic relations: conclusions depend on the relations between premise possibilities and conclusion possibilities.	Reasoners should be able to infer necessary, impossible, and possible conclusions.
3. Truth: mental models represent only what is possible.	Reasoners should infer illusory alethic inferences from certain premises.
Epistemic modals	
4. Conjunctive possibilities: the models that the two systems construct represent conjunctions of possibilities that hold in default of information to the contrary.	Reasoners should infer epistemic possibilities from compound assertions, e.g.: *A or B or both.* ∴ *It's possible that A.* They should reject: *A or B, not both.* ∴ *A or B or both*.
5. Parsimony: consistent possibilities are condensed into models of one possibility unless they are inconsistent.	Reasoners should infer: *It's possible that A and it's possible that B.* ∴ *It's possible that both A and B.*
6. Epistemic conclusions: an assertion of a possibility elicits an additional implicit mental model.	Reasoners should infer: *It's possible that A.* *If A then B.* ∴ *It is possible that B.* more often than their "necessary" analogs.

5. An assessment of the model theory

5.1. Studies of alethic reasoning

Previous studies have tested only inferences in which modal terms occur in conclusions. We begin with these studies of alethic reasoning, and the next section describes new studies of reasoning from modal premises. Alethic reasoning occurs when individuals draw modal conclusions that depend on the relation between models of the premises and models of the conclusion (see 4.1.). Studies have corroborated the theory's three principal predictions in Table 5, which we examine in turn.

The first prediction is that individuals should yield an alethic interaction in making inferences. They should infer the possibility of event more rapidly and more accurately than they infer its necessity, but this difference should switch round for negative assertions. The reason is that individuals need only a single model of the premises to establish a possibility, but all the models of the premises to establish a necessity; in contrast, they need only a single model of the premises to negate a necessity, but all the models of the premises to negate a possibility.

An experiment to test this interaction used descriptions of one-on-one games of basketball in which only two can play (Bell & Johnson-Laird, 1998). There are four putative players, which we abbreviate as *A, B, C, D*, and players can be in the game or out of the game, which we abbreviate as "is in" and "is out". A typical problem was:

78. If A is in then B is in.

If C is in then D is out

Can B be in?

Given that only two can and must play, the possible games to which the premises refer are:

A	B		
	B	C	
	B		D

The mental models represent only the first of these possibilities, and, as the interaction predicts, the participants were faster and more accurate in responding “Yes”, than in doing so when the same premises were paired with:

79. Must B be in?

The “yes” response for the right reasons depends on building fully explicit models of the premises. For inferences in which the correct answer is “No” instead of “Yes”, the experiment used a dual of the previous problem:

80. If A is out then B is out.
    If C is out then D is in.
    Must B be in?

Only three games in which *B* does not play are possible. As the interaction predicts, the participants were faster and more accurate in responding “No”, than in doing so when the same premises were paired with:

81. Can B be in?

The study, which examined various sorts of problem including those based on inclusive disjunctions, corroborated the interaction. It supports the principle that in alethic reasoning models represent possibilities rather than necessities. Only one model is needed to establish a possibility or the denial of a necessity, whereas all models are needed to establish a necessity or the denial of a possibility.

The model theory’s second prediction in Table 5 is that individuals should be able infer necessary, possible, and impossible conclusions. The preceding study showed that they do grasp the different conditions for these various modal conclusions: *must*, *can*, and their denials. Another study also examined such conclusions from syllogistic premises (Evans, Handley, Harper, & Johnson-Laird, 1999). Naive reasoners grasped their differences, and they found it easier to draw possible conclusions than necessary conclusions – one half of the modal interaction. They also found it easier to draw possible conclusions when they were necessary than when they were only possible – a result that the model theory explains: any model in the necessary case bears out the possible conclusion.

The model theory's third prediction in Table 5 is that individuals should be susceptible to alethic illusions. For example, consider this problem:

82. Only one of the following premises is true about a particular hand of cards:

There is a king in the hand or there is an ace, or both.

There is a queen in the hand or there is an ace, or both.

There is a jack in the hand or there is a ten, or both.

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

Mental models, which system 1 constructs, predict a "Yes" response, because reasoners should consider each of the possibilities to which the first premise refers but one at a time. When they get to the possibility of an ace in the hand, they should think, "yes, an ace is possible". The second premise likewise implies that an ace is possible. And the third premise is irrelevant, because it does not mention an ace. So, reasoners should infer that an ace is possible. What system 1 neglects, however, is that when one premise is true, the other two premises are false. So, the first premise or the second premise, or both of them, must be false. When the first premise is false, there is neither a king nor an ace in the hand; when the second premise is false, there is neither a queen nor an ace in the hand. In either case, an ace is not in the hand. Given the premises above, and the question:

83. Is it possible that a king is in the hand?

the model theory predicts that reasoners should also respond, "yes", and this evaluation is correct. An experiment examined such illusions of possibility and their controls, and illusions of impossibility and their controls. The inferences in the four categories were based either on disjunctions or on conditionals (Goldvarg & Johnson-Laird, 2000). The overall results showed that the percentage of correct responses was over 90% for the control inferences, but only 1% for illusions of possibility and 29% correct for illusions of impossibility. Illusions of impossibility should be less compelling because, as the alethic interaction predicts, they demand an examination of all models of possibilities.

Other studies have corroborated alethic illusions (Khemlani & Johnson-Laird, 2009), but perhaps the simplest ones are those in which the task is to evaluate whether it is possible for two assertions both to be true, as in the following problem based on exclusive disjunctions:

84. Either the pie is on the table or else the cake is on the table.

Either the pie isn't on the table or else the cake is on the table.

Could both of these assertions be true at the same time?

A positive answer to the question depends on the two assertions having a possibility in common. Most participants in an experiment responded "yes" (Johnson-Laird, Lotstein, & Byrne, 2012). The mental models of the two premises have in common that a cake is on the table. But, fully explicit models show that this evaluation is an illusion: the two premises refer to no possibility in common. In sum, experiments have corroborated the model theory's three main predictions about alethic reasoning.

5.2. Studies of reasoning from epistemic premises

Epistemic reasoning concerns inferences about the possibility of propositions, and the model theory makes three principle predictions, which we examine in turn. The first prediction, conjunctive possibilities in Table 5, is that compound assertions refer to conjunctions of epistemic possibilities that hold in default of information to the contrary. It predicts that reasoners should accept inferences, such as:

85. The fault is in the software or it is in the connection, or both.

Therefore, it is possible that the fault is in the software.

Hinterecker, Knauff, & Johnson-Laird (2016) carried out such a study, and it corroborated the prediction. They showed that individuals accepted the following sorts of inference, with sensible everyday contents, on separate trials:

86. A or B, or both.

∴ It is possible that A.

∴ It is possible that B

∴ It is possible that A and that B.

The participants also rejected the inference:

87. ∴ It is possible that not A and that not B.

The computer implementation of the theory provided an accurate fit to these and other data from the experiment (Khemlani, Hinterecker, & Johnson-Laird, 2017). For most people, the inferences above are obvious, and some have even argued that their validity is obvious too. But, as we showed earlier, they are invalid in all modal logics. The model theory's other predictions had never been tested, and so we carried out three main experiments to test them.

5.2.1. Experiment 1

The model theory's second prediction, parsimony in Table 5, predicts that individuals should condense two separate possibilities into one unless they are inconsistent with one another. Experiment 1 tested this prediction, and it also tested an aspect of conjunctive possibilities in Table 5, participants they should reject conclusions referring to a possibility that is not supported in the premises' possibilities. Here is a typical condensation inference, where *Yes/No* indicates that the model theory predicts a "Yes" evaluation whereas modal logics predict a "No" evaluation:

A *Yes/No* inference:

88. Premise: It is possible that Ann is in Bath and it is possible that Ben is in Ayr.

Conclusion: It is possible that Ann is in Bath and that Ben is in Ayr.

Does the premise imply that the conclusion is true?

The mental models of the premise are (see Table 3):

Ann in Bath Ben in Ayr

. . .

and so they imply the conclusioin. The experiment included *Yes/Yes* control inferences, where both the model theory and modal logics predict the acceptance of the inference:

A *Yes/Yes* inference:

89. Premise: It is possible that Ann is in Bath and that Ben is in Ayr.

Conclusion: It is possible that Ann is in Bath and it is possible that Ben is in Ayr.

The experiment also included inferences that the model theory predicts reasoners should reject, though they are valid in modal logics, such as system T:

A *No/Yes* inference:

90. Premise: It is possible that Ann is in Bath or it is possible that Tom is in Ayr, but not both.

Conclusion: It is possible that Ann is in Bath or it is possible that Tom is in Ayr, or both.

The model theory predicts that reasoners should reject the inference, because nothing in the premise's models supports one of the possibilities to which the conclusion refers: it is possible that Ann is in Bath and it is possible that Tom is in Ayr. The experiment included control inferences, where both the model theory and modal logics predict a *No* response:

A *No/No* inference:

91. Premise: It is possible that Ann is in Bath or it is possible that Tom is in Ayr, but not both.

Conclusion: It is possible that Ann is in Bath and that Tom is in Ayr.

Method. The 51 logically naive participants (27 men, 24 women; M = 38.9 years) acted as their own controls and carried out four inferences of different sorts in each of the four categories: *Yes/No, Yes/Yes, No/Yes,* and *No/No*, i.e., a total of 16 trials. All problems used only the epistemic modal, *it is possible that,* and varied the connectives *and*, inclusive *or*, and exclusive *or*, in a systematic way. The inferences were presented in one of two different random orders to each participant. Their contents were two syllable names of persons located in cities, as in the preceding examples. These contents were selected at random from a list we had compiled for the inferences. As in our subsequent experiments, the participants carried out the study on an online website (Amazon's Mechanical Turk), and we took the usual precautions for such a procedure, e.g., the program checked that the participants were native speakers of English, and it allowed only one participant per computer. The instructions explained that the task was not a test of intelligence or personality, but concerned the reasoning patterns of people. As the participants were told, each page presented a premise and then a conclusion, followed by the question, "Does the premise imply that the conclusion is true?" We used this format instead of an alethic one, because the premises already contained a modal. The participants responded by pressing one of two keys that were assigned to "Yes" and "No" in a counterbalanced way. They could take as much time as they needed, but they had to try to be correct. Before we carried out any of the studies, we made extensive tests of the mTurk system to ensure that the responses were stored reliably. The raw data for this experiment and the subsequent ones are cached on a web site in the public domain: http://www.cc.uni-freiburg.de/code

Results and discussion.

Table 6 presents the summary results for each of the four sorts of

Table 6

The percentages of evaluations of the inferences in Experiment 1 that fitted the model theory's predictions about the condensation of separate possibilities into one, and the rejection of inferences in which the premises fail to support a possibility to which the conclusion refers. The balance of the percentages did not match its predictions

The predictions of the two accounts	Percentages of the participants' evaluations

		matching the model theory's predictions
The model theory	Modal logic	
Yes	No	89
Yes	Yes	92
No	Yes	76
No	No	75

problem. The r^2 correlation between the model theory's predictions and the results is 0.98, whereas the r^2 correlation between modal logic system T's evaluations and the results is 0.05. Overall, the participants drew 83% of the model theory's predicted conclusions but only 50% of system T's evaluations (Wilcoxon, $z = 6.2$, $p < .0000001$). When theory and logic diverged and the model theory predicted a 'Yes' evaluation, it occurred on 89% of trials, which was reliably greater than system T's predicted 'No' evaluation (11%; Wilcoxon, $z = 6.3$, $p < .0000001$). Likewise, when the model theory predicted a 'No' evaluation, it occurred on 76% of trials, which was reliably greater than system T's predicted 'Yes' evaluation (24%; Wilcoxon, $z = 5.5$, $p < .000001$). Hence, reasoners tend to condense possibilities, and to reject inferences in which the premise possibilities fail to imply one of the possibilities to which the conclusion refers.

Experiment 2

This experiment tested another aspect of model theory's condensation prediction (see Table 5): individuals tend to condense possibilities unless modulation by knowledge blocks the process because the two possibilities are inconsistent. The experiment therefore examined trials with consistent possibilities that should allow condensation into single possibilities, and trials with inconsistent possibilities that should block it, e.g.:

92. Premise: It is possible that Tom is single and it is possible that Tom is married.

Conclusion: It is possible that Tom is single and married.

Does the premise imply that the conclusion is true?

Method. We tested 54 participants (30 men, 24 women; M = 35.1 years) from the same population as before, who acted as their own controls and evaluated 12 inferences based on premises about epistemic possibilities. Half of the inferences should allow condensation into a conclusion referring to a single possibility, and half of them should block it, because the condensation yields an inconsistent possibility. One sort of inference had a conjunctive premise, and one sort of inference had a disjunctive premise, and they were based on three pairs of binary predicates: *right* or *wrong*, *alive* or *dead*, *single* or *married*. The inferences that should allow condensation used two different names in each clause, and the inferences that should block it used the same name in each clause. The problems were presented in a different random order to each participant. The procedure was identical to that in the previous study.

Results and discussion. Table 7 presents the percentages of inferences in which the participants condensed two separate possibilities into one, depending on whether the possibilities were consistent or inconsistent, and on whether the main premise was a conjunction or an inclusive disjunction.

Table 7

The percentages of conclusions of the sort, It is possible that A and B, *in which the participants condensed two possibilities into one, depending on whether* A *and* B *referred to two different individuals and so were consistent possibilities or referred to one individual and so were inconsistent possibilities, and on whether the premise was a conjunction or an inclusive disjunction*

The sorts of premise	The consistency or inconsistency of *A* and *B* in the premise, i.e., whether they referred to two different individuals or the same individual	
	Consistent possibilities	Inconsistent possibilities
Conjunctive premises *It is possible that A and it is possible that B.*	89	28
Disjunctive premises *It is possible that A or it is possible that B, or both.*	58	15

Overall, they made 74% of condensations from premises with different names that allowed condensation, whereas they made only 22% of condensations from premises with the same names, which modulation should block (Wilcoxon test, $z = 5.74$, $p < .000001$). As the model theory predicts, conjunctions (59%) were more likely to yield condensations than disjunctions (43%; Wilcoxon test, $z = 6.565$, $p < .000001$). Modal logics including system T predict that all the responses in the experiment should be 'No'. Hence, the results corroborated the prediction that reasoners condense possibilities unless the contents elicit modulation that blocks the process.

Experiment 3

This experiment tested the model theory's third prediction of epistemic conclusions in Table 5. The principle predicts that reasoners should draw their own "possible" conclusions from premises asserting an epistemic possibility more often than they should draw "necessary" conclusions from premises asserting an epistemic necessity. The bias occurs because models represent possibilities rather than necessities. The experiment therefore tested inferences, such as:

93. It is possible that Alex is in Erie.

If Alex is in Erie than Eddy is in Fremont.

What, if anything, follows?

and their analogs with "necessary" in a premise, such as:

94. It is necessary that Alex is in Erie.

As Table 8 below shows, some of these inferences were valid in modal logic system T, whereas others were not.

Method. We tested 67 participants (27 men, 41 women; M = 44.22 years, SD = 12.47) from the same population as before. They acted as their own controls and drew their own spontaneous conclusions from 12 pairs of premises of the same sort as the example above. All the inferences had the underlying logic of modus ponens:

Kommentiert [PJ1]: The three numbers do not add up!

A

If A then B.

What, if anything, follows?

Four pairs of premises introduced an epistemic possibility: "it is possible that …", four pairs of premises introduced an epistemic necessity: "it is necessary that …", and four control inferences had a categorical premise, *C*, which was unrelated to the conditional, and so the participants should infer than nothing follows from the premises. The eight experimental inferences are shown in Table 8 below. The pairs of premises were presented to each participant in a random order.

Procedure

The procedure was the same as in the previous studies except that the participants had to type their own conclusions in the form of a short sentence to each pair of premises. The key instructions stated that they could take as much time as they needed, but they had to draw a conclusion that was true given that the premises were true. If no such conclusion existed for a pair of premises, they should respond accordingly.

Results and discussion. Table 8 presents the percentages of the three main sorts of conclusion – *it*

is possible that B, *B*, *and it is necessary that B* – that 45 participants drew for the eight experimental problems with possible and necessary premises. Of the original 67 participants, we dropped from the analysis 17 who failed to respond to all the pairs of premises and 5 who guessed conclusions for the control problems. In our evaluations of the conclusions, we counted as *possible* conclusions those that also used "may", "might", and "could", and as *necessary* conclusions those that also used "must". Miscellaneous conclusions included "nothing follows" and its synonyms. Performance on the four control problems for which nothing followed was 98% correct. As Table 8 shows, the principle of epistemic conclusions was reliably correlated: all 45 participants drew more *possible* conclusions from premises containing "possible" than they drew *necessary* conclusions from premises containing "necessary" (Binomial test, $p = .5^{45}$). Remarkably, hardly any participants framed necessary conclusions (only 1% of total conclusions): 36 out of the 45 participants drew only factual conclusions to these four problems (Binomial, $p < .0000001$). The rarity of such conclusions supports the principle that models represent possibilities, not necessities, and so necessity calls for a label on a model – a label that reasoners might easily neglect. In contrast, the participants drew 46% of *possible* conclusions from premises containing “possible”, and 47% of factual conclusions to these premises – again,

Table 8

The percentages of the three main sorts of conclusion that the participants drew in Experiment 3 depending on the different sorts of "possible" and "necessary" premises, and the validity or invalidity in system T. The balance of percentages in each row were other miscellaneous responses

a result that suggests that they build models of possibilities, but focus on explicit models

	Percentages of conclusions the participants drew			Status in system T
Premises with "possible"	Possible that B.	B.	Necessary that B.	
1. It is possible that A. If A then B.	40	51	0	Invalid
2. A. It is possible that if A then B.	44	51	0	Invalid
3. It is possible that A. It is necessary that if A then B.	42	51	0	Invalid
4. A. If A then it is possible that B.	58	33	0	*Possible that B* is valid
Premises with "necessary				
5. It is necessary that A. If A then B.	0	87	7	*Necessary that B* is invalid
6. A. It is necessary that if A then B.	0	93	4	*Necessary that B* is invalid
7. It is necessary that A. It is necessary that if A then B.	2	91	0	*Necessary that B* is valid
8. A. If A then it is necessary that B.	0	89	7	*Necessary that B* is valid

rather

than on the implicit models representing alternative but unspecified possibilities. The participants clustered into two groups in reasoning from premises containing “possible”: 17 participants drew at least three *possible* conclusions out of four, whereas 21 drew no more than one. As Table 8 shows, validity or invalidity in system T had no reliable effect on performance: only five participants drew logical conclusions more often than not (Binomial, $p < .99$). Hence, the overall results corroborated the model theory.

General Discussion

All three examples at the start of this article are inferences that the model theory predicts, but none of them is valid in any modal logic. In the first example, the premise, *Either Trump will be impeached or he won’t be*, elicits the conjunction of two possibilities in default of information to the contrary:

95. Trump will be impeached

    Trump will not impeached

The conjunction of these two possibilities is not a contradiction. And the first conjunct yields the required conclusion, *it’s possible that he will be impeached*, and so reasoners should accept it. From a logical standpoint, however, suppose that it is impossible for Trump to be impeached. The premise is then true, but the conclusion is false, and so the inference is invalid in modal logics. The model theory accommodates this case: it eliminates the relevant default possibility, and yields the remaining possibility as a conclusion: *Trump will not be impeached*. In the second of the opening inferences, the premise, *the probability of rain today is 90%*, also yields a conjunction of two models of default possibilities, but with tags of their numerical probabilities.

And so the treatment of the first example applies to them too. In the third of the opening inferences, the premises:

96. If it rains then the crops will survive.

It’s possible that it will rain.

introduce an epistemic possibility of rain in default of information to the contrary, and the conjunction of the mental models of the two premises yields the required modal conclusion, *It’s possible that the crops will survive*. Once again, the inference is invalid in modal logics, because nothing follows in the case that it doesn’t rain.

These three examples and the inferences described in the present article show that a modal logic, such as system T, cannot be the basis for a plausible psychological theory. There already existed a theory of mental models that accounted for deontic reasoning, i.e., reasoning about permissible and obligatory actions (see, e.g., Bucciarelli & Johnson-Laird, 2005; Johnson-Laird, 2006, Ch. 22). We therefore proposed a general theory of reasoning about possibilities based on models. The theory distinguishes three main interpretations of possibilities: deontic,

alethic, and epistemic (see Table 1). Deontic possibilities concern what is permissible. Alethic possibilities concern the relations between two sets of possibilities, e.g., those for premises and those for conclusions. Epistemic possibilities concern those to which propositions refer.

Underlying the three main interpretations of “possible”, the theory posits a single underlying meaning, which yields different interpretations depending on how it is modulated by background knowledge. The underlying meaning presupposes that individuals can categorize alternative outcomes to events into a small finite number of alternatives – each of which may be realized in infinitely many different ways. A “possibility” refers to a subset of these finite alternatives, which can each be represented in a model, and a “necessity” refers to all of them. This semantics is the same for everyday probabilities (see Johnson-Laird et al., 1999; Khemlani et al., 2012, 2015). The interpretation of modals depends on modulation from knowledge about such matters as deontic conventions, general principles of reasoning, and knowledge pertinent to possibilities. It ensures that deontic alternatives refer to what is permissible and that actions that violate an obligation do not refute it, that alethic alternatives concern relations between one set of assertions (premises or the input to analyses) and another set (conclusions), and that epistemic alternatives are akin to non-numerical probabilities. Hence, epistemic modals give rise to a scale, running from *impossible* to *certain* by way of intermediaries such as *hardly possible* and *almost certain*. These scalar interpretations correspond to degrees of probability, and they can be represented in iconic models of magnitudes that are not numerical (Khemlani et al., 2012, 2015).

The theory postulates that compound assertions, such as conditionals and disjunctions, have a conjunction of models of epistemic possibilities to which these assertions refer in default of information to the contrary. As a result, conditionals differ in meaning from the material conditional of logic (see its "possibility" table in Table 2). This meaning unifies the hitherto different results from studies of the judgments of truth or falsity of conditionals, in which participants judge evidence in which the conditional's *if*-clause is false as irrelevant to the conditional's truth value, and studies in which participants list include these possibilities in those to which true conditionals refer. It obviates the well-known paradoxes of the material conditional in which, for instance, a false *if*-clause validly entails any conditional with the same if-clause. And it resolves the well-known paradox of confirmation in which a claim such as:

97. If an animal is a unicorn then it is invisible

is corroborated by the observation of a kangaroo – a visible animal that is not a unicorn, and indeed is "vacuously" true in logic, because unicorns do not exist. The pattern of possibilities in

Table 2 also has consequences for probabilities: the probability of a conditional, *If A then B*, equals the conditional probability of *B* given *A* – an equation that is one of the keystones of probability logic (Adams, 1998), because the cases in which the *if*-clause is false are irrelevant to the conditional's truth or falsity, and therefore do not enter into its probability.

The model theory postulates that an assertion, such as:

98. It is possible that A

has the following mental models:

A

. . .

The ellipsis denotes epistemic possibilities other than *A*. And the theory specifies mental models and fully explicit models for assertions about the possibilities of all the main sorts of compound assertion (see Table 3).

Alethic reasoning relies on principles relating premise possibilities to those of a conclusion (see Table 4). Likewise, epistemic reasoning relies on the postulate that compound assertions refer to conjunctions of exhaustive possibilities that hold in default of information to the contrary. The theory accordingly makes three main predictions for alethic reasoning and three for epistemic reasoning (Table 5). The evidence that we reviewed showed that individuals are sensitive to the difference between necessary, possible, and impossible conclusions. It also showed that models represent alethic possibilities rather than necessities. As this hypothesis predicts, possible conclusions are easier to infer than necessary ones, whereas conclusions that are not necessary are easier to infer than those that are not possible. This interaction is a consequence of whether or not a single model of a possibility suffices for an inference. The principle of truth, which goes back to the original model theory, predicts the illusions that occur in alethic inferences. System 1 yields intuitive evaluations based on mental models, and so, for example, an inference based on two exclusive disjunctions:

99. Either A or else B, but not both.

Either not-A or else B, but not both.

Could both of these assertions be true at the same time?

tends to yield a positive answer (Johnson-Laird et al., 2012). The mental models of the two premises have *B* in common, which suggests that it is possible for both assertions to be true. Only the fully explicit models show that two premises have no possibility in common.

Studies also corroborated the theory's predictions about epistemic possibilities. They bore out the prediction that individuals infer conclusions about possibilities from disjunctions that made no reference to them, such as inferences of the sort (see Hinterecker et al., 2016):

100. A or B, or both.

Therefore, it is possible that A.

Our new experiments also corroborated the theory. Experiment 1 showed that reasoners tend to condense two possibilities into one:

101. It is possible that A and it is possible that B.

Therefore, it is possible that A and that B.

They also tended to reject inferences, which are valid in modal logic system T, in which a possibility to which the conclusion refers has no support from the premises (see Table 6). But, as Experiment 2 showed, when modulation reveals that the initial possibilities are inconsistent, as in:

102. It is possible that Ann is in Bath and it is possible that she is in Ayr.

Therefore, it is possible that Ann is in Bath and that she is in Ayr.

reasoners reject the conclusion that condenses them into one (see Table 7).

The assertion of an epistemic possibility elicits an additional implicit mental model. So, for instance:

103. It is possible that A

yields a mental model of *A* and an additional implicit model denoting other possibilities. As a consequence, reasoners make the following sort of inference even though it is invalid in modal logics (see Experiment 3):

104. It is possible that A.

If A then B.

Therefore, it is possible that B.

They draw analogous conclusions from premises containing "necessary", but as the principle of epistemic conclusions predicts, they tend to draw factual conclusions rather than ones that are explicitly necessary (see Table 8). These studies of epistemic reasoning showed that reasoners

reasoners tend to follow the model theory rather than modal logic, both in the conclusions that they accept and in the conclusions that they reject.

The corroboration of the theory supports the conclusion that possibilities underlie reasoning based on sentential connectives. We conjecture that it has an analogous role in modal

reasoning with quantifiers, as in:

105. It is possible that all the patients are hypochondriacal.

Viv is one of the patients.

So, it is possible that Viv is hypochondriacal.

Predicate modal logics aim to analyze these sorts of inference. They are more powerful than the sentential modal logics that we described earlier, but they are also more problematical (see, e.g., Girle, 2009). Some of the problems concern logic. For example, it makes sense to assert:

106. Necessarily, there is a side that wins the World cup.

But, it is unclear what exactly, if anything, the following sort of assertion means:

107. There is a side that necessarily wins the World cup

Some philosophers have followed Quine (1953) in arguing that such cases of "quantifying into" a modal context lack a coherent sense. Likewise, quantifiers raise the problem of who and want can be the values of variables in "possible worlds". Nevertheless, we know of no grounds for supposing that the finite semantics of the model theory cannot apply to inferences such as (105).

A major criticism of the model theory concerns its principle of conjunctive possibilities according to which compound assertions refer to conjunctions of default possibilities. Cruz, Over, and Oaksford (2017), echoing Baratgin et al. (2015), argue that this principle creates logical problems and counterintuitive consequences. Although these critics defend a probabilistic approach to reasoning, their arguments make a stand on behalf of orthodox logic (cf. Rips, 1994). They make two main points whose rebuttals allow us to clarify the model theory and to strengthen its support. Their first point is that if *A or B or both* refers to a conjunction of three possibilities, then almost every disjunction is true, because these possibilities hold for any pair of contingent assertions. In fact, the three possibilities are exhaustive in system 1, and so they imply that the fourth case is impossible: *not-A and not-B*. Two contingent assertions do not rule out this case, and so their disjunction would not be true. In a related point, Cruz et al. argue that the disjunction:

A or not-A

is false according to the model theory, because the conjunction *A and not-A* is impossible. But, consider an inclusive disjunction, *A or B or both*. It has a conjunction of three mental models of default possibilities:

A	
	B
A	B

If *not-A* replaces *B*, the result is two mental models:

A	
	not-A

The conjunction of *A* and *not-A* is indeed impossible, and so it yields the null model, which like the empty set, is a member of any set of possibilities.

The critics' second point concerns the inference known as *or*-introduction:

A.

Therefore, A or B or both.

It is valid in orthodox logic, and it follows in probabilistic logic too, because its conclusion is at least as probable as its premise. But, the model theory allows a more refined alethic evaluation of conclusions. The premise for *or*-introduction, *A*, does not support one of the possibilities in the conjunction to which the conclusion refers: *not A* & *B*. Hence, the conclusion is not necessary. But, it follows as an alethic possibility, because premise and conclusion have a possibility in common (see the principles in Table 3).

If *or*-introduction were invalid in the model theory, then Cruz et al. wonder how the following sort of inference could be made:

A.

If A or B or both then C.

Therefore, C.

The fallacy is to suppose that the inference can be made only using *or*-introduction. In fact, the conjunction of the two premises yields the following mental models of possibilities:

A		C
A	B	C

The conclusion, *C*, follows necessarily from these two models, because each of them imply it (see Table 3 for this inference). The crux is therefore whether people accept the following sort of inference, which makes no reference to probabilities:

Anna read a novel.

Does it follow that Anna read a novel or a newspaper, or both?

Most participants in an experiment said, "No" (Orenes & Johnson-Laird, 2012). But, as the model theory predicts, modulation can make the inference acceptable when the premise implies the possibilities to which the conclusion refers:

Eve read *Don Quixote*.

Therefore, she read *Don Quixote* or she read a novel.

Cruz et al. (2017) make a point that was once well-taken: it is not clear what "possible" means according to the model theory. We have, of course, proposed a theory of its meaning in the present article (see 3.2.).

As we remarked at the outset, because knowledge modulates interpretations, each inference has to be assessed on its own merits. Readers might therefore wonder whether the theory provides any normative basis by which to evaluate inferences. In fact, it has a long-standing theory at the computational level: reasoners who draw their conclusions for themselves should aim to maintain semantic information, to simplify, and to reach a new proposition that is not explicit in the premises (see, e.g., Johnson-Laird, 1983, p. 40, and Ch. 2 of Johnson-Laird & Byrne, 1991). The maintenance of semantic information guarantees that an inference is valid in that its conclusion is true in every case in which its premises are true (Jeffrey, 1981, Ch. 1). This definition of validity applies to any domain for which there is an account of when assertions are true. The alethic modality goes beyond the binary distinction between validity and invalidity.

It specifies the alethic status of sentential inferences provided that modulation is taken into account. As a result, it shows that system 1, which uses mental models, yields systematic fallacies – illusory inferences from certain premises. System 2, which uses fully explicit models, can in principle correct them: it embodies a normative account. However, sentential reasoning is not computationally tractable, and so many inferences are beyond its competence.

No orthodox modal logic appears to provide an appropriate system for everyday reasoning. These logics are monotonic, and so definitive evidence contrary to a valid conclusion does not call for the withdrawal of the conclusion; everyday reasoning is defeasible, and so individuals readily withdraw conclusions. Modal logics allow that any proposition whatsoever follows necessarily from a contradiction; everyday reasoning does not (see, e.g., Johnson-Laird et al, 2004). Modal logics are based on “possible worlds” semantics or other analogous infinities; everyday reasoning is based on a finite semantics that also underlies probabilities (Khemlani et al., 2012, 2015). Modal logics offer a unitary system for a given inference, whereas everyday assertions can combine different interpretations of “possibility”. No contradiction occurs in an assertion, such as:

It is possible for Hillary to seek the nomination again, but it is impossible that she will.

The most plausible modal logic for human reasoning is system T, which Osherson (1976) adapted as the basis for his pioneering theory. But, our evidence shows that individuals make inferences that are invalid in system T, and fail to make inferences that are valid in system T. For example, the condensation of possibilities yield the following sort of inference, which is invalid in system T: *It is possible that A and it is possible that B*; therefore, *it is possible that A and that B*. Because modulation can block this inference, defenders of modal logic could argue that system T is plausible as an account of reasoning that takes modulation into account. A crucial datum is thus those inferences valid in system T but that individuals reject. *Or*-introduction is one example. Another is an inference of this sort:

A or B but not both.

Therefore, A or B or both.

Reasoners reject it, but not because of the clash between “but not both” and “or both”, because they accept the converse inference, which has the same clash, more often. It is a case of “weak necessity” (see 4.1.). In sum, human reasoning cannot be based on a normal modal logic.

The model theory postulates an intimate connection between possibilities and probabilities. Epistemic modals even yield a scale of degrees of possibility, and a scale for degrees of necessity, akin to one that occurs for non-numerical probabilities. The model theory integrates the two domains of possibility and probability in a common finite semantics. Individuals make such probabilistic inferences as:

If there is a king or queen in the hand then there is an ace in the hand.

Therefore, an ace is more likely to be in the hand than the king.

They realize that an ace could be in the hand in the absence of a king (Johnson-Laird & Savary, 1996). The analysis of the alethic status of inferences (see Table 4) suggests that alethic modals can be extended to take into account the strength of an inference. An inclusive disjunction:

There is an ace or a king in the hand or both of them.

yields a stronger inference to the conclusion:

Possibly, an ace is in the hand,

than the corresponding exclusive disjunction. The difference is in the relative proportion of possibilities in which the putative conclusion holds. Hence, alethic possibilities yield a non-numerical conditional probability of the conclusion given the premises – a view that is compatible with Bayesian approaches that give up numerical probabilities in favor of sampling (Oaksford & Hall, 2016; Sanborn & Chater, 2016; Teglas et al., 2011). So, possibilities can underlie probabilities, but probabilities cannot underlie possibilities: probabilities have no way to distinguish between necessity and certainty.

6. Conclusions

The conclusions are plain. Evidence suggests that possibilities are the basis of human reasoning. They underlie sentential reasoning. The meanings of *possible*, *necessary*, and their cognates refer to small finite numbers of alternatives – a semantics common to everyday probability. Knowledge of a relevant body of knowledge can yield particular interpretations. Such knowledge concerns laws, rules, and conventions for deontic interpretations, principles of reasoning and analysis for alethic interpretations, and principles of specific knowledge of the world for epistemic interpretations. Inferences are always defeasible, and conditional and disjunctive premises yield conjunctions of possibilities, whereas conjunctive premises yield categorical claims that hold in all possibilities. Intuitive reasoning uses mental models of these possibilities in a system that considers them one at a time. They suffice for many sensible inferences, but they also yield illusory inferences about modal conclusions. Deliberations can flesh out mental models into fully explicit ones, and can consider alternative models, which may correct these fallacies. The patterns of modal inference according to modal logic differ from those underlying human reasoning. People make inf