Minds and Machines (2024) 34:19
https://doi.org/10.1007/s11023-024-09662-4

Models of Possibilities Instead of Logic as the Basis of Human Reasoning

P. N. Johnson-Laird[1,2] · **Ruth M. J. Byrne**[3] · **Sangeet S. Khemlani**[4]

Received: 25 August 2023 / Accepted: 7 January 2024

Abstract
The theory of mental models and its computer implementations have led to crucial experiments showing that no standard logic—the sentential calculus and all logics that include it—can underlie human reasoning. The theory replaces the logical concept of validity (the conclusion is true in all cases in which the premises are true) with necessity (conclusions describe no more than possibilities to which the premises refer). Many inferences are both necessary and valid. But experiments show that individuals make necessary inferences that are invalid, e.g., *Few people ate steak or sole*; therefore, *few people ate steak*. Other crucial experiments show that individuals reject inferences that are not necessary but valid, e.g., *He had the anesthetic or felt pain, but not both*; therefore, *he had the anesthetic or felt pain, or both*. Nothing in logic can justify the rejection of a valid inference: a denial of its conclusion is inconsistent with its premises, and inconsistencies yield valid inferences of any conclusions whatsoever including the one denied. So inconsistencies are catastrophic in logic. In contrast, the model theory treats all inferences as defeasible (nonmonotonic), and inconsistencies have the null model, which yields only the null model in conjunction with any other premises. So inconsistences are local. Which allows truth values in natural languages to be much richer than those that occur in the semantics of standard logics; and individuals verify assertions on the basis of both facts and possibilities that did not occur.

Keywords Defeasible inferences · Logic · Mental models · Possibilities · Reasoning · Verification

1 Introduction

Reasoning is a process that starts with information—a scene or a description—and ends with a conclusion that follows from it. The value of the conclusion depends on the veracity of its starting point, and the quality of the inferential

Extended author information available on the last page of the article

Published online: 04 June 2024

process. Almost all humans can make correct inferences. Take away this ability, and most of science, society, and culture would collapse, and you would not be reading this article. Large language models have led to programs, such as GPT-4, that mimic simple inferences but that cannot be a guide to correct reasoning. Logic provides such a criterion. Logical inferences preserve truth, and so if premises are true, conclusions are true. For over a century, most students of cognition have therefore supposed that some sort of logic underlies human reasoning. Nowadays, they take the idea for granted, defend it only if they need to, and rely on it in circuitous ways. Reasoning, they argue, depends on abstract structures that map onto logic (Piantadosi et al., 2016), adults' intuitions match logical evaluations (Bago & De Neys, 2017) as do infants' (Cesana-Arlotti et al., 2020), patterns of human reasoning are logical (Bringsjord & Sundar Govindarajulu, 2020), the language of thought has a logical structure (Quilty-Dunn et al., 2022) and the conventions of discourse explain otherwise logically erroneous inferences (Aloni, 2022).

Over thirty years ago experiments showed that the logical methods of formal proof are not the way in which humans reason (Evans et al., 1993). Yet the truth values of sentential connectives such as 'or' seemed feasible, and the original theory of mental models adopted these meanings (e.g., Johnson-Laird, 1983). Developments in the theory, however, have led to enough anomalous results to overturn logic as the basis of human reasoning. The rest of this article recounts what happened. It begins with an outline of standard logics (2), and then deals with the proposal that the probability calculus is a better basis for human reasoning (3)—a bold hypothesis but one with limited applications. It next describes the revised theory of mental models (4), distinguishing between intuitive and deliberative models (4.1), and between factual and counterfactual conditionals (4.2). It outlines the principal findings from recent crucial experiments (5). They report contrasts between necessary but invalid inferences (5.1), contrasts between valid inferences that are not necessary (5.2), and the use of truth values outside logic in the verifications of descriptions. These results refute standard logics and corroborate mental models. The article ends in a summary of its main arguments and of the benefits of mental models to human reasoning (6).

2 The Nature of Standard Logics

Standard logics can be defined as the classical sentential calculus and any logic that includes it. They are therefore:

- The sentential calculus, which handles inferences that depend on idealizations of negation, 'not', and of sentential connectives, such as 'if', 'or', and 'and' (see Jeffrey, 1981).
- The predicate calculi, which add to the sentential calculus a logic for properties, relations, and the quantifiers, *some* and *all*. In the simplest calculus—the 'first

order' one, quantifiers range over individual entities; in the calculus one level up in complexity, they also range over sets of individuals (i.e., properties), and so on.

- Modal logics that add to the sentential or predicate calculi a logic for *possible* and *necessary*. There are many modal logics—a countable infinity of them.

The three sorts of logic are our principal concern, but standard logics also include those for the elementary arithmetic of natural numbers, (0, 1, 2,...), i.e., their addition and multiplication (e.g., Gödel, 1931/1965). Extensions of standard logics also yield probability calculi (Demey et al., 2019). All standard logics have a grammar specifying the logical forms of well-formed sentences, a system for proofs that uses formal rules of inferences to derive conclusions, and a separate semantic system that specifies the truth values of compound assertions, such as disjunctions, from the truth values of their constituent clauses.

The concept of validity applies as the criterion for correct reasoning in all standard logics. This concept can be defined as follows (Jeffrey, 1981, p 1):

> A *valid* inference has a conclusion that is true in every case in which all its premises are true.

Hence, there can be no counterexample to a valid inference, i.e., no case in which the premises are true and the conclusion is false. This criterion for good reasoning seems sensible, but it is too exorbitant to be plausible for everyday reasoning—a point to be considered later.

Computer programs implement many standard logics. Yet, no program exists that uses logic to assess the inferences of everyday life—'a scandal of human existence', as a distinguished logician remarked more than fifty years ago (Bar-Hillel, 1969, p 256). To understand why such a program has yet to be implemented, you need to understand how standard logics work. So, suppose you know two things:

> Bernie is a communist or he is a socialist.
> He is not a communist.

You can infer:

> So, he is a socialist.

Your inference has the same form as this one:

> The bug in the program is syntactic or it is conceptual.
> It is not syntactic.
> So, it is conceptual.

Both inferences are valid.

The sentential calculus has two different ways to capture the similarity of the two preceding inferences. The first way uses a set of formal rules (and also, in some formulations, axioms assumed to be true) that depend on the logical *forms* of inferences, as the grammar for the logic defines them. Here is a relevant formal rule in which variables have informative labels instead of being single letters:

Sentence-1 *or* sentence-2.
Not sentence-1.
Therefore, sentence-2.

The italicized terms have constant interpretations in the logic: 'or' includes the case of 'or both,' and 'not' has a semantics that switches a true sentence to being false and vice versa. If sentence-1 is *Bernie is a communist*, and sentence-2 is *Bernie is a socialist*, the rule delivers the following conclusion: *Bernie is a socialist*. It also yields the analogous conclusion for the inference about the bug: *it is conceptual*. So, a single formal rule proves an unlimited number of inferences with different contents. Of course there are other rules (or axioms), and their use in sequences of multiple steps yields more complex proofs. Computer programs implement these systems, including Rips's (1994) pioneering psychological venture. But none of them, not even his, determines the logical forms of everyday inferences. They finesse the problem, and have inputs that are logical forms themselves as in the rule above. To determine logical forms in natural languages is so difficult that, as yet, no algorithm can do the job—a point that Sect. 4.3 illustrates.

The second way to capture similar inferences in standard logics also relies on logical form, but its principles are semantic and concern truth values. Standard logics have only two: *true* and *false* (Jeffrey, 1981). The premise with 'or' in the inference about Bernie is true when at least one of its two clauses is true, and false only when both of its clauses are false. The second premise denies one of its clauses, e.g., *it is not the case that Bernie is a communist*. If the disjunctive premise is true, its other clause is therefore true, and so *Bernie is a socialist*. Standard logics use this sort of treatment for all the other logical connectives, such as 'if' and 'and'. If the premises in the pair of inferences are true, then their conclusions are too. So, both inferences are valid.

A set of assertions can be inconsistent, that is, they cannot all be true at the same time, e.g.:

Bernie is a communist or he is a socialist.
He is not a communist.
He is not a socialist.

The exorbitant cost of inconsistencies in standard logics is that any conclusions whatsoever follow validly from them, e.g.:

Therefore, Bernie admires Trump.

There can be no case in which the inconsistent set of assertions are all true. Hence, there can be no case in which they are all true and a conclusion is false. So, the preceding inference is valid: anything follows from inconsistencies in a standard logic. Their consequences are catastrophic in logic.

A further cost is that nothing can justify the withdrawal of the conclusion to a valid inference. Suppose, for instance, you establish that such a conclusion is false. So, you deny it. Alas, you have now created an inconsistent set of assertions, and they validly imply any conclusion, including the one that you know is false. You must withdraw

from logic in order to withdraw a valid inference. And this argument holds for all standard logics.

Modal logics concern inferences about possibility and necessity, and so they add principles dealing with them to sentential or to predicate logic. The semantic systems for most modal logics are based on *possible worlds* (Kripke, 1963). A sentence about the real world, such as:

Possibly, it is hot in Milan

is true if there is at least one possible world relevant to the real world—or, as logicians say, 'accessible' to it—in which the following categorical sentence is true:

It is hot in Milan.

In each possible world accessible to the real world, the corresponding categorical sentence is either true or false. So, possible worlds are far too vast to fit inside anyone's head (Partee, 1979). What differs from one modal logic to another are their assumptions about which sorts of possible world are accessible to a given world, and so their systems for formal proofs differ in parallel too. In consequence, inferences valid in one modal logic may be invalid in another. Three interpretations of 'possibility' are frequent in daily life, and they differ in assumptions about accessibility (see, e.g., Girle, 2009):

- A *deontic* interpretation concerns actions (or inactions) that are permissible or obligatory.
- An *epistemic* interpretation concerns the occurrence or non-occurrence of events according to knowledge.
- An *alethic* interpretation concerns conceptual or inferential relations that are possible or necessary.

No standard logics can predict which conclusion reasoners ought to infer, because any premises yield infinitely many valid conclusions. Most of them are ludicrous, such as the conjunction of a premise with itself five times, but standard logics do not legislate against ludicrosity. So, psychological theories based on logic focus on the evaluation of given conclusions (Rips, 1994). Henceforth, when the present article uses the naked terms 'logic' and 'model', they refer to a *standard* logic and to a *mental* model, respectively.

3 Probabilistic Theories as an Alternative to Logic

Inferences from uncertain premises can lead to modifications in beliefs and to the withdrawal of conclusions. Some theorists therefore made the commendable step of abandoning logic as the basis of human reasoning. They replaced it with the probability calculus (e.g., Adams, 1998; Oaksford & Chater, 2020; Over, 2020). Probabilities can be consistent with one another (a.k.a. *coherent*), and an inference is correct if it is probabilistically valid: in essence, conclusions should not be less

probable than their premises (Demey et al., 2019). The approach had some success in its applications to conditional inferences and simple quantified inferences (Oaksford & Chater, 2020), but it describes no algorithm for the mental processes of reasoning, and it has no explanations for many results (Knauff & Gazzo Castañeda, 2021). Its fundamental problem, however, is that individuals who have not mastered the probability calculus do not know how to estimate the probabilities of conjunctions or disjunctions. And so they make many inconsistent judgments of probabilities. The best known are instances of the *conjunction fallacy* in which they judge a conjunction to be more probable than one of its constituent clauses (Tversky & Kahneman, 1983). But their estimates of other sorts of probability, such as those for disjunctions, can be inconsistent too (Khemlani et al., 2015).

Probabilists sometime claim that what matters is a mathematical characterization of aggregate human reasoning, that sampling errors are bound to occur (Zhu et al., 2022), and that no need exists to identify correctness (Elqayam, 2017). Yet the approach fails to predict the preceding aggregates. Sampling captures variation, but not the systematic inconsistencies in judgments of probabilities (e.g., Hinterecker et al., 2016; Khemlani et al., 2015). And naive individuals do not know how to assess the consistency of sets of probabilities. It is not native to them, but depends on a calculus created from a numerical measure for possibilities (Khemlani et al., 2015). The probability calculus is normative for judgments of probability, but its computations are beyond the competence of naive individuals.

4 The Theory of Mental Models

The idea of mental models is due to Craik (1943) though it has intimations in Nineteenth century physics. He treated them as simulating only the same input–output pairs—so his ideal model is Kelvin's tidal predictor, which predicts the tides but has a structure quite remote from that of the earth and its orbiting moon. He also restricted models to their role in decision making, and took human reasoning to depend on 'verbal rules'. The re-invention of mental models differs from Craik's account (e.g., Byrne, 2005; Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Khemlani & Johnson-Laird, 2022). Mental models underlie reasoning and their structures correspond as far as possible to those of the situations under description (see also Koralus, 2023). The present section concentrates on the three principle elements that distinguish the model theory from standard logics.

4.1 Models of Possibilities as the Basis of Reasoning

Comprehension leads to the construction of models built from the meanings of assertions and knowledge. The early theory was consistent with standard logics though it relied, not on formal rules of inference, but on models themselves. Inferences from scenes follow from the structure of their spatial models (e.g., Knauff,

2013; Ragni & Knauff, 2013; Stoll & Hegarty, 2016; Tversky, 2019), those about sequences of events follow from kinematic models with structures that unfold in time (e.g., Bucciarelli et al., 2021), and those about causes follow from dynamic models, i.e., kinematic models of possibilities (e.g., Gerstenberg et al., 2021).

Human reasoners aim to draw conclusions that follow of necessity. Failing that, they draw conclusions that are at least possible—a category that includes those that have a probability, numerical or non-numerical. The theory's definition of alethic necessity, which replaces logical validity, is as follows:

> An inference is *necessary* if its conclusion describes no more than the possibilities to which its premises refer.

If a conclusion refers to only some of the possibilities to which the premises refer it should not exclude the remaining ones. Consider this example:

> There's a triangle or a square, or both.
> So, there may be a square.

The conclusion describes one of the premise's possibilities, and so the inference is necessary. In contrast, conclusions that introduce a new possibility, such as:

> There's a triangle.
> So, there's a triangle or a square, or both.

are not necessary inferences, because nothing in the premise implies the additional possibility of a square (Johnson-Laird & Ragni, 2019).

Epistemic possibilities are akin to subjective probabilities, which can even be assigned numerical values in numerate cultures (Lassiter, 2017). If a speaker tells you:

> It may rain

then the model theory postulates that it presupposes that:

> It may not rain.

The idea is presaged in Aristotle's *De Interpretatione* (21b34-6) in which he wrote:

> … 'possible to be' and 'possible not to be' may be thought actually to follow from one another.

But, the model theory instead treats the two assertions as presupposing one another (Johnson-Laird & Ragni, 2019), and a presupposition holds for both an assertion and its negation. Presuppositions are not part of logics, and the inference that the possibility of an event implies the possibility its non-occurrence is invalid in standard modal logics. In all but the simplest of these logics, *it is raining* validly implies that *it is possible that it is raining*. But, if the latter implies that *it is possible that it is not raining*, the resulting three assertions:

It is raining.
It is possible that it is raining.
It is possible that it is not raining.

would be inconsistent with one another, and cannot all be true at the same time. Modal logics take the first pair to embody a valid inference; the model theory takes the second pair to embody a mutual presupposition. To accommodate the latter inference in logic, a typical move is of the sort that the philosopher Grice (1989) envisaged. He argued that a co-operative speaker wouldn't hold back information, and so 'It may rain' conveys as much information as is available to the speaker, who therefore does not know that it will rain. You are therefore entitled to make a tentative inference that it may not rain. From a logical standpoint, if you make this inference, you are committed to both the inferences that yield the three inconsistent assertions above. So, a Gricean appeal to the conventions of discourse may be misguided in this case. Grice was adamant—at least in conversation—that his concern was, not the processes of reasoning, but the conventions of discourse needed to defend logic against apparent violations. Others, however, have transformed his ideas and their own into psychological theories (Holler & Levinson, 2019; Noveck & Spotorno, 2022; Sperber & Wilson, 1995). But they do not analyze the status of the modal inferences above.

A general principle of the model theory is that individuals condense different possibilities into one provided that they are consistent with one another (Johnson-Laird & Ragni, 2019)—a step that is in violation of modal logics. Yet, it accounts for 'free choice permissions' in deontic inferences, such as:

You may leave now or later.
So, you may leave now.

These inferences have perplexed logicians (Kamp, 1973), because they are contrary to the logical meaning of 'or'. Hence, theorists have made many efforts to explain them both in Gricean terms (e.g., Aloni, 2022; Kratzer & Shimoyama, 2017) and in post-Gricean ones (e.g., Bar-Lev & Fox, 2020). Other accounts are akin to the model theory (Geurts, 2005; Zimmermann, 2000). But, unlike these precursors, it explains these inferences as consequences of the condensation of possibilities into one from which they follow necessarily (Johnson-Laird et al., 2021).

4.2 Meanings and Intuitive and Deliberative Models

The meanings of compound assertions formed from sentential connectives such as 'or' refer to possibilities, though a description yielding only one possibility refers to a fact (e.g., Byrne & Johnson-Laird, 2020; Espino et al., 2020; Khemlani et al., 2018). Here's a simple expository example of multiple possibilities:

There is a triangle or a star, or both.

The structure of its mental models corresponds to that of the three possibilities to which it refers. Readers can therefore construct its models from these different possibilities: one possibility is that there's a triangle, another possibility is that there's a star, and a third possibility is that there's both a triangle and a star. Intuitive models of these three possibilities are depicted here on separate rows:

(1) Δ
(2) ☆
(3) Δ ☆

The models are schematic in that the semantic representation from which they are constructed puts no constraints on the particular sizes, shapes, colors, interrelations etc. of the triangle and star. So, each model captures what is common to many different realizations of a possibility—each realization corresponds, if you prefer, to a different possible world. The three possibilities have the force of a conjunction:

> Possibly there's a triangle, and possibly there's star, and possibly there's a triangle and a star.

Each possibility holds only in default of knowledge to the contrary, so they can be withdrawn without contradiction. But, for the assertion to be true, at least one of them must be true.

Intuitive models, such as the ones above, represent what is true, but not what is false, and their construction does not rely on access to a working memory for the results of intermediate computations. When individuals deliberate, their models can also represent what is false. These *deliberative* models of the disjunction, 'There is a triangle or a star, or both' are:

(1) Δ ¬ ☆
(2) ¬ Δ ☆
(3) Δ ☆

The symbol '¬' denotes negation, and so if an assertion is true its negation is false, and vice versa. The negation of a compound assertion yields its complement of models. So the negation of the preceding disjunction has these models:

¬ Δ ¬ ☆

The theory is corroborated in the occurrence of compelling illusory inferences—a consequence from the lack of information about what is false in intuitive models (Johnson-Laird et al., 2000; Khemlani & Johnson-Laird, 2017; Sablé-Meyer & Mascarenhas, 2022). For example, most people answer 'yes' to the following problem (Goldvarg & Johnson-Laird, 2000):

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

There is a king in the hand or there is an ace, or both.
There is a queen in the hand or there is an ace, or both.
There is a jack in the hand or there is a 10, or both.
Is it possible that there is an ace in the hand?

If there were an ace in the hand then the first two assertions would be true, contrary to the stated constraint that only one of the three assertions is true. So, the correct answer is 'No'. Deliberation has access to working memory, and so it can rectify mistaken inferences from intuitive models. This idea of a 'dual system' for reasoning is due to Wason (e.g., 1968), and it is implemented in computer simulations of the model theory (Khemlani & Johnson-Laird, 2022; Ragni et al., 2018). It is also an informal part of many other cognitive theories (notably, Evans, 2008; and Kahneman, 2011).

A fact that eliminates a model yields an inference, e.g.:

There is a triangle or a star, or both.
In fact, there isn't a triangle.

The models of the triangle then change their status from possibilities to counterfactual possibilities, i.e., they were once possible but did not occur. The elimination of these models from the intuitive models above yields only a single model:

☆

So reasoner can draw this conclusion for themselves:

Therefore, there is a star.

Like many inferences, it is both valid in logic and necessary in the model theory.

All sentential inferences depend on a conjunction of possibilities, and the procedure is straightforward. Consider these two assertions:

There is a circle and a triangle.
There is a triangle and a square.

Their respective models are:

◯ □ and △ □

The two models are consistent with one another, and their conjunction is:

◯ □ △

with no unnecessarily duplication of entities. In comparison, consider the conjunction of these two models:

◯ △ and ¬ ◯ △

They are inconsistent because one model represents a circle and the other model represents its negation. Their conjunction yields the null model:

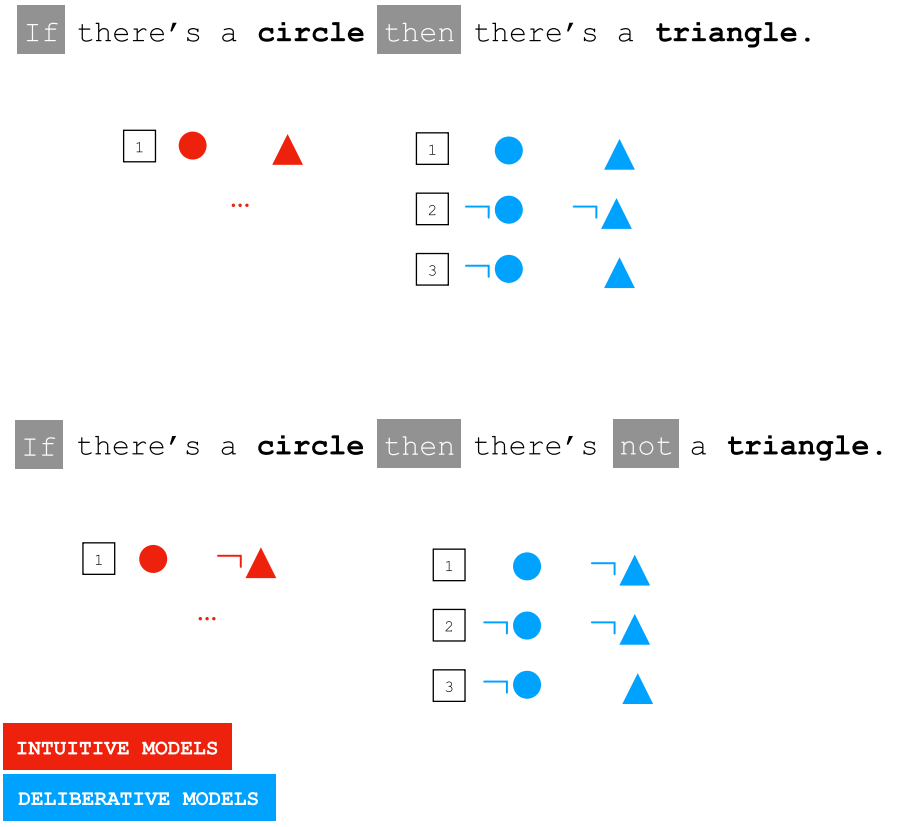

Fig. 1 Illustrations of the intuitive models (in red) and the deliberative models (in blue) of a conditional assertion: *If there's a circle then there's a triange,* and its negation: *If there's a circle then there's not a triangle*

nil

It denotes an inconsistency, which is necessarily false. The conjunction of the null model with any other model yields only the null model again. It follows of necessity only that the premises are inconsistent. Hence, inconsistencies are not like those in logic: their effects are local.

4.3 Factual and Counterfactuals Conditionals

For conditionals, such as:

If there's a circle then there's a triangle

the *if-clause* asserts the possibility of a circle in default of knowledge to the contrary, and the *then*-clause asserts that in this possibility there's a triangle. The possibility that there's a circle presupposes the possibility that there isn't a circle (as described

in 4.1), and in this case there are no constraints on whether or not there is a triangle. Presuppositions hold for the negation of assertions, and Fig. 1 presents the conditional's intuitive and deliberative models, and those for its negation in which the presupposed possibilities also hold (Johnson-Laird & Ragni, 2019).

The possibilities in Fig. 1 predict the main phenomena of conditional reasoning. For example, individuals infer the three possibilities the figure shows for conditionals (Barrouillet et al., 2000; Gauffroy & Barrouillet, 2009). Likewise, it is easy to infer that *there's a triangle* from the conditional, *if there's a circle then there's a triangle*, and the additional premise, *there's a circle*: intuitive models suffice. But it is harder to infer that *there isn't a circle* from the conditional and the assertion that *there isn't a triangle* (Evans et al., 1993); only deliberative models suffice. A conditional's presuppositions are irrelevant to its truth value and to its probability, because they hold for both its affirmation and its negation. So, the conditional above is true given at least one observation of a circle with a triangle and no observation of a circle without a triangle Likewise, its probability equals the conditional probability of a triangle given a circle. This equation, which probabilists often treat as axiomatic (e.g., Adams, 1998), follows from the models of conditionals (Lopéz-Astorga et al., 2022).

Consider the factual conditional (in the indicative mood):

If Oswald didn't shoot Kennedy then someone else did,

and a similar counterfactual (in the subjunctive mood):

If Oswald hadn't shot Kennedy then someone else would have.

The factual conditional is true, whereas the counterfactual is open to doubt (Adams, 1970). So, in one theory (Lewis, 1973), factual conditionals are handled in the sentential calculus, whereas counterfactual conditionals have a *possible worlds* semantics. However, you know that someone shot Kennedy, which verifies the first conditional, but not the second one (Byrne & Johnson-Laird, 2020). An unbiased match with the counterfactual is therefore a factual conditional that refers to an open possibility:

If Oswald hasn't shot Kennedy then someone else will.

Inferences from genuine matches between the two sorts of conditional run in parallel (Byrne & Johnson-Laird, 2020; Espino & Byrne, 2020; Goodwin & Johnson-Laird, 2018; Orenes et al., 2022).

One reason that logical form is so difficult to determine is that the meaning of an assertion and general knowledge can modulate the interpretation of connectives such as, 'or,' 'and,' and 'if' (Quelhas & Johnson-Laird, 2017; Quelhas et al., 2017). A modulation of the assertion:

It rained or it poured

elicits the knowledge that *poured* means *rained heavily*, and so the assertion cannot have a model representing that it didn't rain but poured. The assertion is therefore not a disjunction, but refers to two possibilities, one in which it rained and did not

pour, and one in which it rained and poured. So, it follows of necessity that it rained, and reasoners draw this conclusion (Quelhas et al., 2019). Likewise:

Ben drinks gin and he drinks tonic, but he doesn't drink gin and tonic

is a self-contradiction in standard logics. But the conjunction 'gin and tonic' can mean the two *together*, and so the assertion is consistent in daily life. Modulation can also add relations between elements of models, as it does for temporal relations between events, which people infer often without awareness of doing so (Juhos et al., 2012).

Of course there is more to the theory than the preceding principles, and the next section of the article illustrates some of its refinements in outlining the crucial differences between logic and models, and in presenting experiments that examined these differences.

5 Three Sorts of Crucial Experiment

Many inferences are both valid in logic and necessary in the model theory. Crucial experiments therefore contrast the two accounts, and so this section reports studies of necessary inferences that are invalid, valid inferences that are not necessary, and verifications of assertions with truth values outside logic but not the model theory.

5.1 Studies of Necessary but Invalid Inferences

The model theory predicts that certain inferences are necessary, though invalid in logic. Here in outline are the results from three recent studies.

• People infer that the possibility of an event implies the possibility of its non-occurrence, and vice versa (see 4.1 above) Given necessary, though invalid inferences, such as:

It is possible that it rains.
So, it is possible that it does not rain.

participants tended to accept them (Ragni & Johnson-Laird, 2020). As Sect. 4.1 argued, the hypothesis that they depend on conventions of discourse is implausible.

• The model theory predicts that individuals should accept each of the following necessary inferences in default of knowledge to the contrary (see 4.2):

There is a triangle or a star, or both.
So, it is possible that there is a triangle.
So, it is possible that there is a star.
So, it is possible that there is a triangle and a star.

Reasoners make such inferences from everyday disjunctions (Hinterecker et al., 2016). A reviewer argued that they were 'obviously valid'—they certainly seem so. In fact, they are invalid in standard modal logics. That is because if, say, it is

impossible that there's a triangle but there is a star, the premise above is true, but both conclusions about the triangle are false. Individuals make analogous inferences from conditional assertions to the possibilities shown in Fig. 1 in 4.3 (e.g., Barrouillet et al., 2000).

Third, model theory predicts that individuals condense separate possibilities into one provided that they are consistent (4.1). In fact, individuals make inferences of this sort that have nothing to do with permissions, but that violate the logic of disjunctions (Rasga et al., 2022):

Phil likes red or white wine.
So, Phil likes red wine.

They also make free choice permissions that are both outside Gricean (e.g., Aloni, 2022; Kratzer & Shimoyama, 2017) and post-Gricean theories (e.g., Bar-Lev & Fox, 2020). For instance, they accept the following sort of inferences (Rasga et al., 2022; Sklarek et al., 2023):

Imagine that your professor told you that you are *permitted* to do only one of the following actions:

You can do your homework.
You can do the presentation slides.

So, you are *permitted* to do your homework.

The inference depends on three premises, whereas Gricean discourse conventions concern single assertions (Cohen, 1971).

5.2 Studies of Valid Inferences that are not Necessary

In logic, nothing can justify the rejection of a valid inference—not even conventions of discourse (see Sect. 2). Yet, in three studies, participants rejected valid inferences that are not necessary according to the model theory.

• Inferences of this sort are valid in logic:

It may rain.
So, it may rain or freeze, or both.

Participants tended to reject such inferences (Hinterecker et al., 2016; Orenes & Johnson-Laird, 2012), which are not necessary, because nothing in the premise establishes that it may freeze. And so the inference is only a possible one.

• Inferences of the following sort are valid, because the truth of the premise guarantees the truth of the conclusion:

It's possible that it is raining or else that it is freezing, but not both.
So, it's possible that it is raining or that it is freezing, or both.

Participants rejected such inferences (Ragni & Johnson-Laird, 2020). They are not necessary, because nothing in the premise establishes the possibility that both

events occur. The mismatch between 'or both' and 'but not both' is not an adequate explanation on the result, because participants were more likely to infer the premise from the conclusion than vice versa.

• In logic, inconsistent premises imply any conclusions whatsoever (see Sect. 2), e.g.:

Boris always lies.
Boris does not always lie.
So, a hippopotamus is in his bath.

Inferences of this sort are valid but so bizarre that it would be silly to test whether participants reject them. Experiments have instead examined premises, such as:

If someone pulled the trigger, then the gun fired.
Someone pulled the trigger, but the gun did not fire.

Individuals notice the inconsistency, but rather than infer any conclusions, they try to formulate an explanation that resolves it (Johnson-Laird et al., 2004; Khemlani & Johnson-Laird, 2012).

5.3 Studies of Verifications Outside Logic

In standard logics, every sentence is either true or false (see Sect. 2). In daily life, assertions can have many other sorts of truth value, e.g., *neither true nor false; could be true and could be false*, *true and couldn't be false*, and so on. In logic, your arrival at one destination in this assertion:

You arrived in Venice or in Treviso

establishes that the assertion is true. Individuals have the same intuition in daily life. But, consider a counterfactual assertion, such as:

You would have arrived in Venice or in Treviso (if your flight hadn't been cancelled).

You didn't arrive at either destination, and so in logic the disjunction is false. Yet it could be true (Espino & Byrne, 2021; Orenes et al., 2019; Quelhas et al., 2018). So, counterfactuals are often analyzed in a 'possible worlds' semantics (see 4.3). The model theory postulates an alternative to possible worlds. Individuals start with a model of the actual situation—you did not arrive at either Venice or Treviso—and then modify it to represent the counterfactual event in which your flight was not cancelled (Byrne, 2005). You were flying to Venice, but in the event of fog there you would have landed in Treviso. So the counterfactual disjunction is true. The verification of factual and counterfactual assertions runs in parallel.

In a study of verification, participants spontaneously evaluated, not just the facts, but also pertinent counterfactual possibilities (Johnson-Laird et al., 2023). They saw a picture of a journey in which their car arrived, say, at Bath but could have arrived at Slough, which was the only other possible outcome. They then verified the assertion:

Table 1 A synopsis of how standard logics and mental models differ

Topic	Standard logics	Mental models
(A) Truth, representations, and verification		
Truth values:	Any sentence is *true* or *false*	Various sorts, e.g., *absolutely true, true, could be true, probably true*
Interpretation of terms such as 'if,' 'or', 'and'	They are constant	Their context can modulate them, as in: *I drink gin* and *I drink tonic, but not gin* and *tonic*
Representations of sentences, e.g., *There is a star or a diamond, or both*	Logical form, e.g., *sentence-1* or *sentence-2*, where the variables have values such as 'There is a star'	Intuitive models: ☆ ◊ — Deliberative models: ☆ not ◊ / not ☆ ◊
Verification of sentences, e.g., *She's in Ayr or Ulm:*	*True*, if she is in one of the two places, but otherwise false	*True but could have been false* if she is in one of the two places, but could not have been in the other (Johnson-Laird et al., 2023)
(B) Reasoning		
How people reason:	They use rules of inference that match the logical forms of sentences	They conjoin pairs of mental models that are consistent with one another
Criterion of correctness:	*Validity*: the conclusion is true in all cases in which the premises are true. Nothing justifies the withdrawal of valid inferences	*Necessary inferences*: the conclusion refers only to the premises' possibilities. *Possible inferences*: it also refers to additional possibilities. Any inference is defeasible
Consequences of inconsistent premises:	Any inference whatsoever from them is valid	Something is wrong with the premises (e.g., Khemlani & Johnson-Laird, 2012)
Crucial inferences, e.g.: *Phil likes red or white wine.* So, *Phil likes red wine*	Invalid	Necessary, and reasoners tend to accept it (Rasga et al., 2022)
Crucial inferences, e.g.: *Possibly, Di is in Ulm or else Ed is in Ayr, but not both.* So, *Possibly Di is in Ulm or Ed is in Ayr, or both*	Valid	Not necessary, and reasoners tend to reject it (Ragni & Johnson-Laird, 2020)

You arrived at Bath or at Slough,

and tended to choose the option:

True and it couldn't have been false.

But, if instead the road to Slough was blocked, they evaluated the assertion as:

True but it could have been false.

That is, the assertion would have been false if they had chosen the road to Slough, because they would have been unable to reach it. They used similar variants of falsity. These counterfactual truth values cannot be expressed in logic.

6 Conclusions

Table 1 summarizes the argument of this paper: experiments show that reasoners draw conclusions that are necessary in the model theory, and reject conclusions that are not necessary, regardless of whether they are valid in standard logics. No robust results are known to corroborate logic and to refute the model theory, though future research could lead to such a refutation, e.g., it could show that validity rather than necessity is decisive for the acceptance of certain sorts of inference. The model theory specifies what the brain computes in reasoning, which is the alethic status of inferences. This goal is computable: various programs simulate the process. And brain-imaging studies have corroborated neural events corresponding to the manipulation of mental models (Alfred et al., 2020; Cortes et al., 2022; Treur, 2021; van Ments, & Treur, 2022). The theory is consistent with the hypotheses that thinking depends on discrete symbols (Dehaene et al., 2022), structured representations (Andonovski, 2022; Lagnado, 2021; Radvansky, 2015; Tversky, 2019), and neural realizations of programs (Sablé-Meyer et al., 2022). Reasoning is algorithmic, but probabilistic principles can help to simulate human variation in its processes (Khemlani & Johnson-Laird, 2022). Standard logics exclude *true* and *false* from the language of proofs: they are confined to the semantic system (Tarski, 1944). No such segregation occurs in a natural language, and it allows a rich variety of truth values, which individuals are happy to use.

Many non-standard logics exist, and the model theory has been identified with such a logic (Bringsjord et al., 2020). If proponents of this view are happy to treat logic as giving an account of the truth or falsity of factual assertions about the world, as making no distinction between the language of reasoning and the language of semantics, as having no logical constants and no formal rules of inference, as replacing validity with alethic necessity, as treating all inferences with empirical content as defeasible from knowledge to the contrary, as allowing complex truth values that include counterfactual components, and as yielding, by design, erroneous inferences from intuitive models, then so be it. Perhaps a working definition of logic should be Quine's (1986): the systematic study of logical truths. In any case, the contrast between the model theory and standard logics remains secure.

The end of the story is that reasoning based on mental models has several benefits for humanity:

- You can draw your own conclusions from premises.
- You can modify or withdraw any conclusion—even one that otherwise follows of necessity—in case, say, you discover that its conclusion is false.
- Inconsistencies are local. Everyone has a vast set of beliefs, and the task of checking whether they are consistent is computationally intractable (Cook, 1971/2023), and so inconsistencies occur—one famous case concerned Frege's formulation of a logic (see Russell, 1902/1967). Combine a standard logic with the likelihood of inconsistent beliefs, and the result could be a logical catastrophe. It perplexed Wittgenstein and Turing (for transcriptions of their interchanges on the topic, see Wittgenstein & Bosanquet, 1989). But, in the model theory, inconsistencies are errors isolated in their effects—which explains why they can benefit thinking rather than wreck it (Wilczek, 2002).
- Reasoning depends on meanings. So, there is no need to recover the logical forms of assertions, which cannot be done without considering meanings (Johnson-Laird et al., 2023). That is why no program exists to apply logic to everyday inferences. But, if you start with meaning, you don't need logical form to build models, and so programs implementing the model theory cope with inferences from daily life.

Logic is a supreme achievement of human thought, not the converse. No case against logic is made in the present article, which shows only that it cannot underlie your everyday reasoning. The evidence suggests that you rely instead on models of possibilities. If you depend on logic as a theoretical touchstone, then you don't have to abandon it, but only to remember that it differs from human reasoning. If you have to reason from complicated premises, you should try to enumerate their possibilities, and to think about what might happen. A frequent cause of disasters in daily life is to overlook a possibility. One such possibility is a consequence of the future development of machines for correct reasoning—the obsolescence of its human counterpart.

Funding None.

Declarations

Competing Interests All authors declares that they have no conflict of interest to disclose.

References

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

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

Alfred, K. L., Connolly, A. C., Cetron, J. S., & Kraemer, D. J. (2020). Mental models use common neural spatial structure for spatial and abstract content. *Communications Biology, 3*(1), 17. https://doi.org/10.1038/s42003-019-0740-8

Aloni, M. (2022). Logic and conversation: the case of free choice. *Semantics and Pragmatics*. https://doi.org/10.3765/sp.15.5

Andonovski, N. (2022). Episodic representation: A mental models account. *Frontiers in Psychology, 13*, 899371. https://doi.org/10.3389/fpsyg.2022.899371

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition, 158*, 90–109. https://doi.org/10.1016/j.cognition.2016.10.014

Bar-Hillel, Y. (1969). Colloquium on the role of formal languages. *Foundations of Language, 5*, 256–284.

Bar-Lev, M. E., & Fox, D. (2020). Free choice, simplification, and innocent inclusion. *Natural Language Semantics, 28*(3), 175–223. https://doi.org/10.1007/s11050-020-09162-y

Barrouillet, P., Grosset, N., & Lecas, J.-F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition, 75*, 237–266. https://doi.org/10.1016/S0010-0277(00)00066-4

Bringsjord, S., & Sundar Govindarajulu, N. (2020). Rectifying the mischaracterization of logic by mental model theorists. *Cognitive Science, 44*, e12898. https://doi.org/10.1111/cogs.12898

Bucciarelli, M., Mackiewicz, R., Khemlani, S. S., & Johnson-Laird, P. N. (2021). The causes of difficulty in children's creation of informal programs. *International Journal of Child-Computer Interaction, 31*, 100443. https://doi.org/10.1016/j.ijcci.2021.100443

Byrne, R. M. J. (2005). *The rational imagination*. MIT Press.

Byrne, R. M., & Johnson-Laird, P. N. (2020). *If* and *or*: Real and counterfactual possibilities in their truth and probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(4), 760–780. https://doi.org/10.1037/xlm0000756

Cesana-Arlotti, N., Kovács, Á. M., & Téglás, E. (2020). Infants recruit logic to learn about the social world. *Nature Communications, 11*(1), 5999. https://doi.org/10.1038/s41467-020-19734-5

Cohen, L. J. (1971). The logical particles of natural language. In Y. Bar-Hillel (Ed.), *Pragmatics of Natural Language* (pp. 50–68). Reidel. https://doi.org/10.1007/978-94-010-1713-8

Cook, S. A. (1971). The complexity of theorem-proving procedures, *STOC'71: Proceedings of the Third Annual ACM Symposium on Theory of Computing*, *151–158.* Rep. in Cook, S. A. (2023). The complexity of theorem-proving procedures. In *Logic, Automata, and Computational Complexity: The Works of Stephen A. Cook* (pp. 143–152). https://doi.org/10.1145/3588287.3588297

Cortes, R. A., Peterson, E. G., Kraemer, D. J., Kolvoord, R. A., Uttal, D. H., Dinh, N., Weinberger, A. B., Daker, R. J., Lyons, I. M., Goldman, D., & Green, A. E. (2022). Transfer from spatial education to verbal reasoning and prediction of transfer from learning-related neural change. *Science Advances*. https://doi.org/10.1126/sciadv.abo3555

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

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: A hypothesis about human singularity. *Trends in Cognitive Sciences, 26*, 751–766. https://doi.org/10.1016/j.tics.2022.06.010

Demey, L., Kooi, B., & Sack, J. (2019). Logic and Probability. In Zalta, E. N. (Ed.) *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). https://plato.stanford.edu/archives/sum2019/entries/logic-probability/

Elqayam, S. (2017). The new paradigm in psychology of reasoning. In L. Ball & V. A. Thompson (Eds.), *International Handbook of Thinking and Reasoning* (pp. 130–150). Routledge.

Espino, O., & Byrne, R. M. (2020). The suppression of inferences from counterfactual conditionals. *Cognitive Science, 44*(4), e12827. https://doi.org/10.1111/cogs.12827

Espino, O., & Byrne, R. M. (2021). How people keep track of what is real and what is imagined: The epistemic status of counterfactual alternatives to reality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(4), 547–570. https://doi.org/10.1037/xlm0000965

Espino, O., Byrne, R. M., & Johnson-Laird, P. N. (2020). Possibilities and the parallel meanings of factual and counterfactual conditionals. *Memory & Cognition, 48*, 1263–1280. https://doi.org/10.3758/s13421-020-01040-6

Evans, J. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human Reasoning: The Psychology of Deduction*. Erlbaum.

Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. *Developmental Review, 29*(4), 249–282. https://doi.org/10.1016/j.dr.2009.09.002

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review, 128*(5), 936–975. https://doi.org/10.1037/rev0000281

Geurts, B. (2005). Entertaining alternatives: Disjunctions as modals. *Natural Language Semantics, 13*(4), 383–410. https://doi.org/10.1007/s11050-005-2052-4

Girle, R. (2009). *Modal logics and philosophy* (2nd ed.). Routledge.

Gödel, K. (1931/1965). On formally undecidable propositions of Principia Mathematica and related systems I. In Davis, M. (Ed.) ed., *The Undecidable* (Pp. 5–38). (Trans., E. Mendelson, of original publication in 1931.) Hewlett, NY: Raven Press, 1965.

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

Goodwin, G. P., & Johnson-Laird, P. N. (2018). The truth of conditional assertions. *Cognitive Science, 42*(8), 2502–2533. https://doi.org/10.1111/cogs.12666

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

Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2016). Modality, probability, and mental models. *Journal of Experimental Psychology Learning, Memory, and Cognition, 42*(10), 1606–1620. https://doi.org/10.1037/xlm0000255

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences, 23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

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

Johnson-Laird, P. N. (1983). *Mental models*. Harvard University Press.

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

Johnson-Laird, P. N., Byrne, R. M. J., & Khemlani, S. S. (2023). Truth, verification, and reasoning. *Psychological and Cognitive Sciences*. https://doi.org/10.31234/osf.io/spb83

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*(3), 640–661. https://doi.org/10.1037/0033-295X.111.3.640

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science, 288*(5465), 531–532. https://doi.org/10.1126/science.288.5465.531

Johnson-Laird, P. N., Quelhas, A. C., & Rasga, C. (2021). The mental model theory of free choice permissions and paradoxical disjunctive inferences. *Journal of Cognitive Psychology, 33*(8), 951–973. https://doi.org/10.1080/20445911.2021.1967963

Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition, 193*, 130950. https://doi.org/10.1016/j.cognition.2019.04.019

Juhos, C., Quelhas, A. C., & Johnson-Laird, P. N. (2012). Temporal and spatial relations in sentential reasoning. *Cognition, 122*(3), 393–404. https://doi.org/10.1016/j.cognition.2011.11.007

Kahneman, D. (2011). *Thinking fast and slow*. Farrar.

Kamp, H. (1973). Free choice permission. *Proceedings of the Aristotelian Society, 74*, 57–74.

Khemlani, S. S., Byrne, R. M., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive Science, 42*(6), 1887–1924. https://doi.org/10.1111/cogs.12634

Khemlani, S. S., & Johnson-Laird, P. N. (2012). Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta Psychologica, 139*(3), 486–491. https://doi.org/10.1016/j.actpsy.2012.01.010

Khemlani, S. S., & Johnson-Laird, P. N. (2017). Illusions in reasoning. *Minds and Machines, 27*, 11–35. https://doi.org/10.1007/s11023-017-9421-x

Khemlani, S., & Johnson-Laird, P. N. (2022). Reasoning about properties: A computational theory. *Psychological Review, 129*(2), 289–312. https://doi.org/10.1037/rev0000240

Khemlani, S. S., Lotstein, S., & Johnson-Laird, P. N. (2015). Naive probability: Model-based estimates of unique events. *Cognitive Science, 39*, 1216–1258. https://doi.org/10.1111/cogs.12193

Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. MIT Press.

Knauff, M., & Gazzo Castañeda, L. E. (2021). When nomenclature matters: Is the "new paradigm" really a new paradigm for the psychology of reasoning? *Thinking & Reasoning*. https://doi.org/10.1080/13546783.2021.1990126

Koralus, P. (2023). *Reason and Inquiry: The Erotetic Theory*. Oxford University Press.

Kratzer, A., & Shimoyama, J. (2017). Indeterminate pronouns: The view from Japanese. In C. Lee, F. Kiefer, & M. Krifka (Eds.), *Contrastiveness in Information Structure, Alternatives and Scalar Implicatures* (pp. 123–143). Springer. https://doi.org/10.1007/978-3-319-10106-4_7

Kripke, S. A. (1963). Semantical analysis of modal logic i normal modal propositional calculi. *Mathematical Logic Quarterly, 9*(5–6), 67–96. https://doi.org/10.1002/malq.19630090502

Lagnado, D. A. (2021). *Explaining the Evidence: How the Mind Investigates the World*. Cambridge University Press. https://doi.org/10.1017/9780511794520
Lassiter, D. (2017). *Graded modality: Qualitative and quantitative perspectives*. Oxford University Press.
Lewis, D. (1973). *Counterfactuals*. Wiley.
Lopéz-Astorga, M., Ragni, M., & Johnson-Laird, P. N. (2022). The probability of conditionals: A review. *Psychonomic Bulletin & Review, 29*(1), 1–20. https://doi.org/10.3758/s13423-021-01938-5
Noveck, I. A., & Spotorno, N. (2022). Experimental pragmatics. In J. Verschueren & J. O. Östman (Eds.), *Handbook of Pragmatics* (pp. 1555–1577). John Benjamins.
Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology, 71*, 305–330. https://doi.org/10.1146/annurev-psych-010419-051132
Orenes, I., Espino, O., & Byrne, R. M. (2022). Similarities and differences in understanding negative and affirmative counterfactuals and causal assertions: Evidence from eye-tracking. *Quarterly Journal of Experimental Psychology, 75*(4), 633–651. https://doi.org/10.1177/17470218211044085
Orenes, I., García-Madruga, J. A., Gómez-Veiga, I., Espino, O., & Byrne, R. M. (2019). The comprehension of counterfactual conditionals: Evidence from eye-tracking in the visual world paradigm. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2019.01172
Orenes, I., & Johnson-Laird, P. N. (2012). Logic, models, and paradoxical inferences. *Mind & Language, 27*(4), 357–377. https://doi.org/10.1111/j.1468-0017.2012.01448.x
Over, D. E. (2020). The development of the new paradigm in the psychology of reasoning. In S. Elqayam, I. Douven, J. B. T. Evans, & N. Cruz (Eds.), *Logic and Uncertainty in the Human Mind* (pp. 243–263). Routledge. https://doi.org/10.4324/9781315111902-15
Partee, B. H. (1979). Semantics—Mathematics or psychology? In R. Bäuerle, U. Egli, & A. von Stechow (Eds.), *Semantics From Different Points of View* (pp. 311–360). Springer-Verlag.
Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review, 123*(4), 392–424. https://doi.org/10.1037/a0039980
Quelhas, A. C., & Johnson-Laird, P. N. (2017). The modulation of disjunctive assertions. *Quarterly Journal of Experimental Psychology, 70*(4), 703–717. https://doi.org/10.1080/17470218.2016.1154079
Quelhas, A. C., Rasga, C., & Johnson-Laird, P. N. (2017). A priori true and false conditionals. *Cognitive Science, 41*, 1003–1030. https://doi.org/10.1111/cogs.12479
Quelhas, A. C., Rasga, C., & Johnson-Laird, P. N. (2018). The relation between factual and counterfactual conditionals. *Cognitive Science, 42*(7), 2205–2228. https://doi.org/10.1111/cogs.12663
Quelhas, A. C., Rasga, C., & Johnson-Laird, P. N. (2019). The analytic truth and falsity of disjunctions. *Cognitive Science, 43*(9), e12739. https://doi.org/10.1111/cogs.12739
Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2022). The best game in town: The re-emergence of the language of thought hypothesis across the cognitive sciences. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X22002849
Quine, W. V. O. (1986). *Philosophy of logic*. Harvard University Press.
Radvansky, G. A. (2015). *Human memory* (2nd ed.). Psychology Press.
Ragni, M., & Johnson-Laird, P. N. (2020). Reasoning about epistemic possibilities. *Acta Psychologica, 208*, 103081. https://doi.org/10.1016/j.actpsy.2020.103081
Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review, 120*(3), 561–588. https://doi.org/10.1037/a0032460
Ragni, M., Kola, I., & Johnson-Laird, P. N. (2018). On selecting evidence to test hypotheses: A theory of selection tasks. *Psychological Bulletin, 144*(8), 779–796. https://doi.org/10.1037/bul0000146
Rasga, C., Quelhas, A. C., & Johnson-Laird, P. N. (2022). An explanation of or-deletions and other paradoxical disjunctive inferences. *Journal of Cognitive Psychology, 34*(8), 1032–1051. https://doi.org/10.1080/20445911.2022.2091576
Rips, L. J. (1994). *The psychology of Proof*. MIT Press.
Russell, B. A. W. (1967). Letter to Frege, 1902. In J. van Heijenoort (Ed.), *From Frege to Gödel* (pp. 124–125). Harvard University Press.
Sablé-Meyer, M., Ellis, K., Tenenbaum, J., & Dehaene, S. (2022). A language of thought for the mental representation of geometric shapes. *Cognitive Psychology, 139*, 101527. https://doi.org/10.1016/j.cogpsych.2022.101527
Sablé-Meyer, M., & Mascarenhas, S. (2022). Indirect illusory inferences from disjunction: A new bridge between deductive inference and representativeness. *Review of Philosophy and Psychology, 13*(3), 567–592. https://doi.org/10.1007/s13164-021-00543-8

Sklarek, B., Knauff, M., & Johnson-Laird, P. N. (2023). Assertions, metaassertions, and mental models. *PsyArXiv*. https://doi.org/10.31234/osf.io/dxz5h

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell.

Stull, A. T., & Hegarty, M. (2016). Model manipulation and learning: Fostering representational competence with virtual and concrete models. *Journal of Educational Psychology, 108*(4), 509–527. https://doi.org/10.1037/edu0000077

Tarski, A. (1944). The semantic conception of truth: And the foundations of semantics. *Philosophy and Phenomenological Research, 4*(3), 341–376. https://doi.org/10.2307/2102968

Treur, J. (2021). Mental models in the brain: On context-dependent neural correlates of mental models. *Cognitive Systems Research, 69*, 83–90. https://doi.org/10.1016/j.cogsys.2021.06.001

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293

Tversky, B. (2019). *Mind in motion: How action shapes thought*. Hachette.

van Ments, L., & Treur, J. (2022). Dynamics, adaptation and control for mental models: a cognitive architecture. In *Mental Models and Their Dynamics, Adaptation, and Control A Self-Modeling Network Modeling Approach.* (pp. 3–26). Cham, Germany: Springer.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*(3), 273–281. https://doi.org/10.1080/14640746808400161

Wilczek F (2002) A piece of magic The Dirac equation. In: Farmelo G (Ed) It Must Be Beautiful Great Equations of Modern Science. London: Granta. Pp. 132–160

Wittgenstein, L., & Bosanquet, R. G. (1989). *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge, 1939*. University of Chicago Press.

Zhu, J. Q., Newall, P. W., Sundh, J., Chater, N., & Sanborn, A. N. (2022). Clarifying the relationship between coherence and accuracy in probability judgments. *Cognition, 223*, 105022. https://doi.org/10.1016/j.cognition.2022.105022

Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics, 8*(4), 255–290. https://doi.org/10.1023/A:1011255819284

Authors and Affiliations

P. N. Johnson-Laird[1,2] · Ruth M. J. Byrne[3] · Sangeet S. Khemlani[4]

✉ P. N. Johnson-Laird
phil@princeton.edu

[1] Department of Psychology, Princeton University, Princeton, NJ 08544, USA

[2] Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA

[3] School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Dublin 2, Ireland

[4] Navy Center for Applied Research in Artificial Intelligence, The US Naval Research Laboratory, 4555 Overlook Dr. SW, S.W., Washington, DC 20375, USA