CellPress

Logic, probability, and human reasoning

P.N. Johnson-Laird[1,2], Sangeet S. Khemlani[3], and Geoffrey P. Goodwin[4]

[1] Princeton University, Princeton, NJ, USA
[2] New York University, New York, NY, USA
[3] Naval Research Laboratory, Washington, DC, USA
[4] University of Pennsylvania, Philadelphia, PA, USA

This review addresses the long-standing puzzle of how logic and probability fit together in human reasoning. Many cognitive scientists argue that conventional logic cannot underlie deductions, because it never requires valid conclusions to be withdrawn – not even if they are false; it treats conditional assertions implausibly; and it yields many vapid, although valid, conclusions. A new paradigm of probability logic allows conclusions to be withdrawn and treats conditionals more plausibly, although it does not address the problem of vapidity. The theory of mental models solves all of these problems. It explains how people reason about probabilities and postulates that the machinery for reasoning is itself probabilistic. Recent investigations accordingly suggest a way to integrate probability and deduction.

The nature of deductive reasoning

To be rational is to be able to make deductions – to draw valid conclusions from premises. A valid conclusion is one that is true in any case in which the premises are true [1]. In daily life, deductions yield the consequences of rules, laws, and moral principles [2]. They are part of problem solving, reverse engineering, and computer programming [3–6] and they underlie mathematics, science, and technology [7–10]. Plato claimed that emotions upset reasoning. However, individuals in the grip of moderate emotions, even those from illnesses such as depression or phobia, reason better than controls, although only about matters pertaining to their emotion [11,12]. Deductive reasoning is an ability that varies vastly from one person to another, correlating with their intelligence and with the processing capacity of their working memory [13–15]. Our topic is the cognitive foundation of deductive reasoning, and we ask two fundamental questions:

(i) Does deduction depend on logic [16–20]?
(ii) How does deduction fit together with probabilities?

The first question is timely because of proposals that probability is the basis of human reasoning [21–23]. The second question has engaged theorists from the economist John Maynard Keynes [24] onward. Here we address both questions. We begin with logic (see Glossary) and present the arguments that logic alone cannot characterize deductive competence. These arguments motivated the turn to probability – a pivot that its proponents refer to as the 'new paradigm' [25–29]. Next, we outline the theory of mental models [30–34], which combines set theory with psychological principles. Finally, we present recent studies that suggest how to integrate deduction and probability.

Problems for logic as a theory of deductive reasoning

An ancient proposal is that deduction depends on logic (see also [16–20]). Sentential logic concerns inferences from premises such as conjunctions ('and') and disjunctions ('or'). Like most logics, it has two parts: proof theory and model theory [35]. Proof theory contains formal rules of inference for proofs. One major rule of inference in most formalizations is:

$A \rightarrow C$
A
therefore, C

where A and C can be any sentences whatsoever, such as: '2 is greater than 1' $\rightarrow$ '1 is less than 2'.

Proof theory specifies rules containing logical symbols such as $\rightarrow$, but not their meanings. Model theory defines their meanings. It specifies the truth of simple sentences such as '2 is greater than 1' with respect to a model, such as the natural numbers, and the truth of compound sentences containing connectives, such as $\rightarrow$, which is known as material implication. The meaning of $A \rightarrow C$ is defined as follows: it is true in any case except when A is true and C is false [1] and so it is analogous to 'if A then C'. This definition can be summarized in a truth table (Table 1). Model theory therefore determines the validity of inferences: a valid inference is one in which the conclusion is true in all cases in which the premises are true.

Logic is extraordinarily powerful and underlies the theory of computability [35–37]. Many cognitive scientists have accordingly supposed that human reasoning depends on unconscious formal rules of inference [16–20]. The hypothesis is plausible, but it runs into three difficulties.

First, conventional logic is monotonic; that is, if an inference is valid, its conclusion never needs to be withdrawn – not even when a new premise contradicts it. A contradiction validly implies any conclusion whatsoever [1]. However, human reasoners faced with a solid fact tend

Corresponding author: Johnson-Laird, P.N. (phil@princeton.edu).

1364-6613/
 http://dx.doi.org/10.1016/j.tics.2015.02.006

Glossary

Bayesian net: a directed graph in which each node represents a variable and arrows from one node to another represent conditional dependencies. It captures the complete joint probability distribution in a parsimonious way.
Consistency: a set of assertions is consistent if they can all be true at the same time.
Counterexample: in an inference, a possibility to which the premises refer but which is inconsistent with the conclusion.
Deductive reasoning: a process designed to draw a conclusion that follows validly from premises; that is, the conclusion is true in any case in which the premises are true.
Defeasible logics: also known as 'non-monotonic' logics. Unlike conventional logic, they allow conclusions to be weakened or withdrawn in the face of facts to the contrary.
Defective truth table: a truth table for a conditional, 'if *A* then *C*', that has no truth value when *A* is false (also known as the de Finetti truth table).
The Equation: the probability of a conditional, 'if *A* then *C*', equals the conditional probability of '*C* given *A*'.
Fully explicit model: unlike a mental model, it represents a possibility depicting each clause in the premises as either true or not. The fully explicit models of a disjunction, '*A* or *B* but not both', accordingly represent a conjunction of two possibilities: possibly(*A* & not-*B*) & possibly(not-*A* & *B*).
Kinematic model: a mental model that unfolds in time to represent a temporal succession of events.
Logic: the discipline that studies the validity of inferences. There are many logics, normally comprising two main components: proof theory, which stipulates rules for the formal derivation of proofs; and model theory, which is a corresponding account of the meanings of logical symbols and of the validity of inferences. In sentential logic, each proof corresponds one to one with a valid inference, but for other, more powerful logics not every valid inference can be proved.
Logical form: the structure of a proposition that dovetails with the formal rules of inference in a logic. No computer program exists to recover the logical form of propositions in daily life.
Material implication: a compound assertion in logic whose truth table is presented in Table 1 in main text. It is sometimes taken to correspond to a conditional, 'if *A* then *C*'. This view leads to logically valid but unacceptable 'paradoxes' such as that *C* implies 'if *A* then *C*'.
Mental model: an iconic representation of a possibility that depicts only those clauses in a compound assertion that are true. The mental models of a disjunction, '*A* or *B* but not both' accordingly represent two possibilities: possibly(*A*) and possibly(*B*).
Model theory: the component of a logic that accounts for the meaning of sentences in the logic and for valid inferences.
Modulation: the process in the construction of models in which content, context, or knowledge can prevent the construction of a model and can add information to a model.
Monotonicity: the property in conventional logic in which further premises to those of a valid inference yield further conclusions.
New paradigm: see probabilistic logic.
Probabilistic logic (p-logic): a paradigm for reasoning that focuses on four hypotheses: Ramsey's test, the defective truth table, the Equation, and p-validity.
Proof theory: the branch of a logic that provides formal rules of inference that can be used in formal proofs of conclusions from premises.
P-validity: an inference is p-valid if its conclusion is not more informative than its premises.
Ramsey's test: to determine your degree of belief in a conditional assertion, add its if-clause to your beliefs and assess the likelihood of its then-clause.
Recursive process: a loop of sequential operations performed either for a predetermined number of times or while a particular condition holds. If it has to be conducted an indefinite number of times, as in multiplication, it needs a working memory to hold intermediate results.
Syllogism: a form of inference that Aristotle formulated based on two premises and a conclusion, which each contain a single quantifier, such as 'all *A*', 'no *A*', or 'some *A*'.
Systems 1 and 2: the two systems of reasoning postulated in dual-process theories of judgment and reasoning, in which system 1 yields rapid intuitions and system 2 yields slower deliberations. Many versions of the theory exist.
Truth table: a systematic table showing the truth values of a compound assertion, such as a conjunction, as a function of the truth values of its clauses.
Validity: in logic, an inference is valid if its conclusion is true in every case in which its premises are true. In everyday reasoning, its premises should also be true in every case in which its conclusion is true.
Vapid deductions: valid inferences that yield useless conclusions, such as the conjunction of a premise with itself.

to withdraw any conclusion that it contradicts. Some theorists therefore defend so-called 'non-monotonic' or 'defeasible' logics developed in artificial intelligence, which allow conclusions to be withdrawn [38–42].

Second, conditional assertions (e.g., 'If she insulted him then he's angry') occur in all sorts of reasoning. However, they do not correspond to any connective in sentential logic. Theorists have treated them as material implications [16–19], but this interpretation yields deductions of the following sort:

He's angry.

Therefore, if she insulted him then he's angry.

As the truth table for $A \rightarrow C$ shows (Table 1), whenever *C* is true, the material implication is true. The preceding inference is therefore valid on this interpretation. It is also valid to infer a material implication from the falsity of *A*; for example:

She didn't insult him.

Therefore, if she insulted him then he's angry.

However, people usually reject both sorts of inference [43], which are called the 'paradoxes' of material implication. They are a major motivation for alternative foundations for human reasoning [21–23,25–29].

Third, logic yields infinitely many valid conclusions from any set of premises but many of them are vapid, such as a conjunction of the same premise with itself some arbitrary number of times; for example, '*A*, therefore, *A* and *A* and *A*'. Logic alone cannot characterize sensible inferences [8,30,31]. Psychological theories based on logic therefore resort to extralogical methods to prevent vapid inferences [18–20]. No one knows to what degree these methods work without preventing useful inferences.

A further practical difficulty is that formal rules of inference apply, not to sentences, but to logical forms that match those of the formal rules of inference. No computer program exists for extracting logical forms from sentences in natural language, let alone from the propositions that sentences express in context. No one knows in full how to identify these forms from their shadows cast in sentences [44].

Probability logic

As a consequence of the preceding arguments, some cognitive scientists propose that probability should replace logic. Their theories differ in detail but overlap enough to have a label in common – the new paradigm [25–29]. We refer to the paradigm as 'probability logic' or 'p-logic' for short. It presupposes that degrees of belief correspond to subjective probabilities [45–49], an idea that not all psychologists accept [50,51]. It focuses on conditionals, and one p-logician even allows that conventional logic could apply to other sorts of assertion [47]. P-logic's proponents engage with four main hypotheses.

First, individuals fix their degree of belief in a conditional, using Ramsey's test [45]. To assess, say, 'If she insulted him then he's angry', they add the content of the if-clause (she insulted him) to their beliefs and then assess the likelihood of the then-clause (he's angry).

Second, Ramsey's test or an analogous concept of a conditional event [46] defines the conditions in which a conditional is true or false. As Table 1 shows, they yield

Table 1. The truth table in logic for material implication *A* → *C*, the contrasting defective truth table (also known as the de Finetti truth table) for a conditional, 'if *A* then *C*', and the cases that are possible and impossible according to the mental model theory of conditionals

Contingencies in the world	Truth value of *A* → *C*	Truth-value of 'if *A* then *C*' in a defective truth table	Possible and impossible cases for 'if *A* then *C*' in the mental model theory
A & *C*	True	True	Possible
A & not-*C*	False	False	Impossible
Not-*A* & *C*	True	No truth value	Possible
Not-*A* & not-*C*	True	No truth value	Possible

Table 2. The views of various advocates of p-logic about the four key hypotheses of p-logic

	I	II	III	IV	V
1. Ramsey's test assesses the probability of conditionals	+?	+	+	+	+
2. Conditionals have the defective (de Finetti) truth table	+	+?	+?	+	
3. The Equation holds: p(if A then C) = $p(C\|A)$	+	+	+	+	+?
4. Reasoners rely on p-validity rather than logical validity	+	+	–	–	–

+, acceptance of a hypothesis; –, its rejection; ?, caveats on the decision.
The five sets of advocates are: (I) Oaksford and Chater [21,22]; (II) Evans, Over *et al.* [25–27]; (III) Oberauer *et al.* [76,77]; (IV) Pfeifer, Fugard, *et al.* [28,54]; and (V) Douven and Verbrugge [78,79], who do not propose any semantics for conditionals and so neither accept nor reject hypothesis 2.

a 'defective' truth table, also known as a de Finetti truth table, in which a conditional is void when its if-clause is false [21,26,46,47,52–54]. It has no truth value, like division by zero has no numerical value.

Third, as both of the previous hypotheses imply, the probability of a conditional, 'if *A* then *C*', is the ratio of the cases in which the conditional is true, '*A* & *C*', to the cases in which it has a truth value: (A & C) + (A & not C). Hence, the probability of a conditional, p(if A then C) equals the conditional probability $p(C|A)$. This relation is known as 'the Equation' and is a major consequence of p-logic [21–23,25–29].

Fourth, apart from exceptional individuals – mathematicians, perhaps – the concept of probabilistic validity should replace logical validity (e.g., [21,47]). A p-valid inference has a conclusion that is not less probable than its premises. All valid deductions are p-valid but not all p-valid inferences are valid deductions. The lower the probability of a proposition, *A*, the more informative it is. This idea derives from the Shannon measure of the statistical information, *I*, in a signal, *A*, which is inversely proportional to *A*'s frequency of occurrence [55]:

$$I(A) = \log_2(1/p(A)) = -\log_2 p(A).$$

Informativeness, however, treats $p(A)$ as referring, not to statistics in signals, but to the probabilities of events; that is, to the probability of the possibilities to which *A* refers [56]. Deductions in logic do not increase informativeness, whereas inductions do [57], and individuals distinguish between the two sorts of reasoning [58]. In general, an inference is p-valid provided that its conclusion's informativeness is not greater than the sum of the informativeness of each of its *n* premises [21,22,47]:

$$I(\text{conclusion}) \leq \sum_{i}^{n} I(\text{premise}_i)$$

Table 2 summarizes the various views of advocates of p-logic about the four hypotheses. Box 1 shows how p-logic applies to syllogisms. But, how does it cope with the problems for logic – its monotonicity, its treatment of conditionals, and its vapid conclusions?

P-logic is not monotonic. Assertions can vary in certainty and p-logic allows evidence to lower their probability, even to the value of zero [47].

P-logic does not yield the paradoxes of material implication. Consider this example:

The Taliban won't be defeated soon in Afghanistan.
Therefore, if the Taliban are defeated soon in Afghanistan then life exists on Jupiter.

The conclusion is much less probable than the premise and so the inference is not p-valid [47]. It is not even logically valid given the defective truth table, because a valid inference demands a true conclusion if its premise is true and in this case the falsity of the premise ensures that the conditional has no truth value.

P-logic does not address the problem of vapid conclusions – perhaps because it focuses on what is computed, not how it is computed. Its proponents have yet to describe the mental processes underlying, say, Ramsey's test or estimates of conditional probabilities. The theory may be able to avoid vapidity, but the conjunction of a premise with itself is just as probable as the premise alone and so the inference is p-valid. We consider the evidence for p-logic below, but first we consider a third theory.

The theory of mental models

During World War II, Kenneth Craik proposed that individuals simulate the world in mental models to make predictions, but that reasoning depends on verbal rules [59]. A more recent theory – the mental model theory – postulates that simulation underlies reasoning too [30–33]. It is a simple idea – people simulate possibilities – and most of its ramifications are integrated in a computer program, mReasoner [60], which is in the public domain at http://mentalmodels.princeton.edu/models/.

Mental models resemble models in logic: both treat deductions as invalid if they have a counterexample; that

Box 1. The application of p-logic to syllogistic reasoning

Aristotle formulated the logic of syllogisms. These dominated logic for two millennia and the psychology of reasoning for many decades. They have two premises and a conclusion, each in one of four moods: 'all *A* are *B*', 'no *A* is a *B*', 'some *A* are *B*', or 'some *A* are not *B*'. For example:

Some Greeks are men.
Some men are athletes.

An end term (e.g., 'Greeks') is in one premise, whereas a middle term (e.g., 'men') is in both premises. Middle terms have four possible arrangements, depending on whether they are first or last in each premise. Hence, there are 64 possible pairs of premise: 27 of them yield valid deductions interrelating the end terms and 31 yield p-valid inferences [114]. Syllogisms vary vastly in difficulty: 7-year-old children can cope with the easiest, whereas adults struggle with the hardest [88]. Given the premises above, many reasoners infer:

Therefore, some Greeks are athletes.

The inference is invalid, because the men who are Greeks need not be the men who are athletes. According to the 'probability heuristics model' [114], the inference is not a failure in logical reasoning but a success in p-logic, because p(Greek & athlete) > 0. Granted that premises can be ranked in order of decreasing informativeness ('all', 'most', 'few', 'some', 'none',' some are not'), reasoners can use three main heuristics to converge on p-valid conclusions.

(i) The preferred conclusion has the same quantifier as the least informative premise. In our example, the two premises have the same quantifier and so the preferred conclusion is one with the quantifier 'some', too.
(ii) The next-most preferred conclusion is a p-valid implication of the preceding conclusion; for example, 'Some _ are not _.'
(iii) If the least informative premise has an end term as its subject, it is the subject of the conclusion; otherwise, the end term in the other premise is the subject of the conclusion.

The virtue of p-logic is that it extends to quantifiers, such as 'most *A*', that are outside traditional syllogisms and that cannot be defined using the standard quantifiers of 'first order' logic in which variables range over individuals but not over properties (i.e., sets of individuals) [126]. The preceding heuristics and two other less important ones yield predictions about the conclusions that reasoners should draw for each of the 64 possible pairs of syllogistic premises. We assess these predictions below.

is, a model of the premises that is inconsistent with the conclusion. The mental model theory, however, is based on three psychological principles [31–33,60]. First, each mental model represents a distinct set of possibilities. For example, the disjunction:

Pat visited England or she visited Italy, or both

has three mental models, which we abbreviate using the names of Pat's destinations, although in reality models represent situations in the world:

1.	England	
2.		Italy
3.	England	Italy

Second, mental models represent only what is true in a possibility: what is false is left implicit. For example, the first model above does not represent that it is false that Pat visited Italy. This 'principle of truth' reduces the load on processing but yields systematic fallacies, which we illustrate below.

Third, with deliberation reasoners can use the meaning of assertions to flesh out mental models into fully explicit models. For the disjunction above, they are:

1.	England	not-Italy
2.	not-England	Italy
3.	England	Italy

The disjunction is true provided that each of these three cases is possible. Content, context, and knowledge can modulate fleshing out. For example, with a disjunction such as:

Pat visited Milan or she visited Italy

modulation blocks the model of Pat visiting Milan but not Italy because it is impossible, and so the assertion has only these two models of her visits:

1.	Milan	Italy
2.		Italy.

Like other 'dual-process' theories (e.g., [26,61–64]), the mental model theory depends on two systems. System 1 constructs mental models. It is rapid but often errs because it cannot use working memory to store intermediate results. System 2 has access to working memory and so can perform recursive processes such as the construction of fully explicit models, but is fallible when working memory becomes overloaded. Nevertheless, valid inferences deriving from mental models should be easier than those requiring fully explicit models [60]. One of the theory's major differences from logic and p-logic is that reasoning can depend on kinematic models that unfold in time [8,30–33]. Box 2 illustrates how they work.

Mental models solve the three problems for logic. The first problem is that everyday reasoning is not monotonic. Suppose, for instance, you believe that:

If someone pulled the trigger, then the gun fired.
Someone pulled the trigger.

However, you then discover that, in fact:

The gun did not fire.

What would you infer? Most people try to explain what might have happened [65] and conclude, for example, that:

Someone emptied the chamber and so there were no bullets in the gun.

Hence, contrary to many philosophical accounts (following [66]), individuals do not always accommodate an inconsistent fact with a minimal change to their beliefs [67]. Instead, they simulate what might have happened, generating a mental model (or models) of the situation that explain the inconsistency [65], and they rate such explanations as more probable than minimal changes [68]. Explanations of this sort depend on mental models and have the advantage of providing a guide to action. Neither non-monotonic logic nor p-logic creates explanations.

The mental model theory requires no special logic for the task (*pace* [38–42]). Simulation can be performed within a protected environment – an intellectual laboratory to try out hypotheses – to model other individuals' inferences, to envisage causal interventions or counterfactual possibilities, and to explain inconsistencies [65,68–70]. As a computer implementation shows, models are crucial at all stages [65]. First, an inconsistency is detected between the premises and the fact: they have no model in common. Facts weaken a belief that has a mental model inconsistent with them and so the fact that the gun did not fire weakens the conditional belief above rather than the categorical belief. The facts are accordingly updated to:

Someone pulled the trigger and the gun did not fire.

Finally, these facts trigger a search in knowledge for an explanation, where knowledge itself takes the form of fully

Box 2. Simulations in kinematic mental models

Anecdotes suggest that reasoning can depend on kinematic models of sequences of events; for example, Tesla imagined running a dynamo to infer the wear on its bearings. However, most people simulate not the simultaneous rotations of pulleys in a system, but the effect of one pulley's rotation on the next pulley, and so on [127]. A good test of kinematic deductions is from informal descriptions of computer programs, because they differ in complexity. Figure I depicts a railway track and the cars, *ABCDEF*, that have to be rearranged into a new order on the right track (e.g., *ACEBDF*). Individuals ignorant of programming can create informal programs for solving such rearrangements, even for trains of any number of cars [5]. A computer program, mAbducer, also creates such programs with recursive loops of moves for trains [5]. Programming depends on deducing the consequences of a program to check that it does what it is supposed to do. We invite readers to envisage the consequences of the following program for the train *ABCDEF* on the left track.

Move all but one of the cars to the siding.
While there is at least one car on the siding,
move one car from the left track to the right track,
move one car from the siding to the left track.
Otherwise, move one car from the left track to the right track.

Many people can deduce the rearrangement, which reverses the cars: *FEDCBA*. They report imagining the sequence of moves.

The difficulty of solving a rearrangement problem reflects the minimal numbers of moves in its solution. The difficulty of deducing the consequences of a program depends instead on the complexity of the program, which has a simple universal measure – Kolmogorov complexity [160]. This measure assesses the complexity of a string of symbols from the length of its shortest description in a standard language such as Common Lisp. The relative complexities of four sorts of program predicted the percentages of accurate deductions of their consequences. For instance, the reversal of the order of the cars calls for more moves than interleaving them, as in a 'Faro shuffle' of cards (e.g., *ADBECF*). However, the program for the Faro shuffle is more complex because it calls for the same car, *D*, to be moved to and from the siding twice and so, as predicted, it yielded fewer accurate deductions than the program for reversals [5].

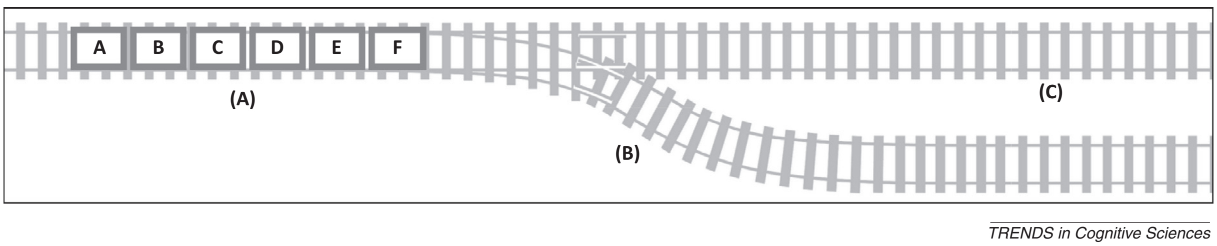

Figure I. A computer display of a simple railway environment [5]. It has a left track **(A)** on which there is an arrangement of cars and a siding **(B)** that can hold cars while others are moved from the left track to the right track **(C)**. Both the siding and the left track can act as stack-like memories and so the system has considerable computational power.

explicit models of causal relations, which can trigger one another to create a novel causal chain [65]. The search can fail, but it can also yield various putative explanations such as the one above.

The second problem for logic is its treatment of conditionals, which yields the 'paradoxes' of material implication [43]. Box 3 shows how mental models provide a solution. There are also paradoxes of disjunction, for example:

Pat visited England.
Therefore, Pat visited England or she visited Italy, or both.

Reasoners balk at this inference too [43]. However, it is logically valid, and it is p-valid because its conclusion cannot be less probable than its premise. By contrast, the inference is invalid for mental models. The disjunction is true provided that the three destinations for Pat are possible:

possibly(England & not-Italy) & possibly(not-England & Italy) & possibly(England & Italy).

The premise does not establish that the second and third cases are possible. Hence, the truth of the premise does not guarantee the truth of the disjunctive conclusion and so the inference is not valid. However, as the theory of mental models predicts, if modulation blocks the visit to Italy, which comes from nowhere, the inference is much more acceptable [43], for example:

Pat visited Milan.
Therefore, Pat visited Milan or she visited Italy.

A visit to Milan implies a visit to Italy and so the premise establishes both of the possibilities to which the conclusion refers. Reasoners accept the inference much more often than its unmodulated version and an analogous recipe works just as well for the paradoxes with conditionals [43].

The third problem for logic is that validity does not avoid vapidity. The mental model theory explains why individuals do not draw vapid conclusions such as the conjunction of a premise with itself. The solution hinges on the processes underlying reasoning, which construct models from meanings [60]. Modulation can prevent the construction of a model but cannot add models of new possibilities [71]. A description of the resulting models yields a conclusion. There is therefore no machinery to form vapid conclusions (e.g., to conclude 'It is not the case that the sun is not shining' from 'The sun is shining'). Regarding the practical difficulty for logic, models are constructed from the meanings of assertions and so they avoid the hard task of extracting logical form. We now turn to the evidence for p-logic and for mental models.

An evaluation of p-logic and mental models

P-logic has the great merit of allowing that reasoning can be tentative, uncertain, and probabilistic. It has inspired much ingenious research. However, what is the standing of its four key hypotheses?

Ramsey's test assesses the probability of conditionals. Perhaps reasoners use the test [45], but their estimates of

Box 3. Mental models of conditionals

A conditional such as:

If it rained then it was cold

has the mental model:

rain	cold

However, deliberation can lead to its fully explicit models [71,128], which are summarized in Table 1 in main text:

rain	cold
not-rain	not-cold
not-rain	cold

The same order occurs developmentally in children's interpretations of conditionals and the capacity of working memory predicts the number of possibilities that they envisage [129,130]. Modulation can block any model of a conditional [71]. For example, 'If it rains then it'll pour' has no model in which it pours but does not rain, because the meaning of 'pours' implies that it rains. If modulation blocks a conditional's mental model, however, it refutes the conditional. Experiments have corroborated modulations, including those that establish temporal and spatial relations between the events that a conditional describes [71,131,132].

Basic conditionals – those unaffected by modulation – refer to the three possibilities above. They are analogous to the truth table for material implication (see Table 1 in main text) and that analogy has misled many theorists into supposing that the model theory treats basic conditionals as material implications (e.g., [23,71,75,79]). On the contrary, it implies that a basic conditional, 'if *A* then *C*', is true only if all three situations in its fully explicit models are possible: possibly(*A* & *C*) & possibly(not-*A* & not-*C*) & possibly(not-*A* & *C*) and *A* & not-*C* is impossible. The conjunction of possibilities, which is the proper interpretation, shows that the paradoxes of material implication are invalid in the theory of mental models. Neither the falsity of *A* nor the truth of *C* implies that '*A* & *C*' is possible and so neither of them implies 'if *A* then *C*'. The conjunction also reveals the flaw in the proof in the main text for the existence of God: the truth of a conditional 'if *A* then *C*' implies that *A* is possible, not that it is true. Consider the conditional:

If God exists then atheism is correct (if *A* then *C*).

Atheism means that God does not exist and so modulation blocks the model of the possibility '*A* & *C*'. The conditional is therefore false. However, unlike material implication or the defective truth table, its falsity does not imply that *A* is true. '*A* & *C*' is impossible and so *A* may not be true.

conditional probabilities tend to violate the complete joint probability distribution (see below).

Conditionals have a defective truth table. Some experiments corroborate its predictions (e.g., [72]) but others do not [73]. Also, consider this conditional:

If God exists then atheism is wrong.

It is true. Its defective truth table (Table 1) therefore implies that its if-clause is true:

Therefore, God exists.

The inference is a valid deduction from a true premise and so is a sound proof for the existence of God. One way to avoid such bizarre proofs is to abandon truth for 'truthiness' and to treat the meaning of conditionals as conditional probabilities satisfying the Equation. This solution seems implausible, because conditionals can certainly be true (and false). Nevertheless, some theorists have endorsed it [21,47] or analogous proposals [74].

The probability of 'C *if* A' *equals the conditional probability of* 'C *given* A'. Some studies support the Equation, although to varying degrees [54,75–80], and others do not [81,82]. It may fail because of ignorance of the probability calculus: a recent study reports that the judgments of only a minority of well-educated individuals corroborated it and only for some sorts of conditional [83].

Reasoners rely on p-validity in which a conclusion is never more informative than its premises. The hypothesis explains the acceptance of certain invalid conclusions but, as we show presently, its account of syllogistic reasoning (Box 1) is of only middling accuracy.

The theory of mental models makes three crucial predictions that distinguish it from other accounts, as follows.

The principle of truth predicts the occurrence of systematic fallacies. They can be compelling cognitive illusions. Some of them concern conditionals [84] and are sometimes open to alternative explanations, but those based on disjunctions are hard to explain without the principle of truth [85–87]. Consider, for example, this problem:

1. Either the pie is on the table or else the cake is on the table.
2. 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? Most people answer 'yes': the mental models of both assertions represent the cake on the table [87]. However, the fully explicit models of the two disjunctions have no model in common:

1.		2.	
pie	not-cake	not-pie	not-cake
not-pie	cake	pie	cake

Hence, the correct answer is 'no'. Participants tend to be wrong about such illusory inferences but to be right about control inferences for which mental models yield the correct answer.

Reasoners should spontaneously use counterexamples to refute invalid deductions. They do so most often to refute conclusions that are consistent with the premises but that do not follow from them [88–90]. Figure 1 shows a particular region of the brain, the right frontal pole, that becomes active in reasoning only during a search for counterexamples [91].

Valid inferences should be easier from mental models (system 1) than from fully explicit models (system 2). They should be faster and more accurate. Experiments have corroborated this prediction for all of the main domains of reasoning, including reasoning about spatial, temporal, sentential, and quantified relations. Table 3 cites studies examining these three predictions.

A small demonstration may help you sharpen your intuitions. Would you accept or reject the following inference?

Viv visited Ireland or Scotland, but not both of them.

Therefore, Viv visited Ireland or Scotland, or both of them.

If you reject this inference, neither logic nor p-logic explains your reasoning because the inference is both valid and p-valid. It is valid because the conclusion holds in any case in which the premise holds [1]. It is p-valid because the conclusion holds in both cases in which the premise holds and in an additional case, and so the conclusion is less informative than the premise [47]. However, the inference is not valid in the model theory because the two cases to which the premise refers do not establish that it is possible that Pat visited both countries. So, mental models explain your rejection of the inference and other similar examples.

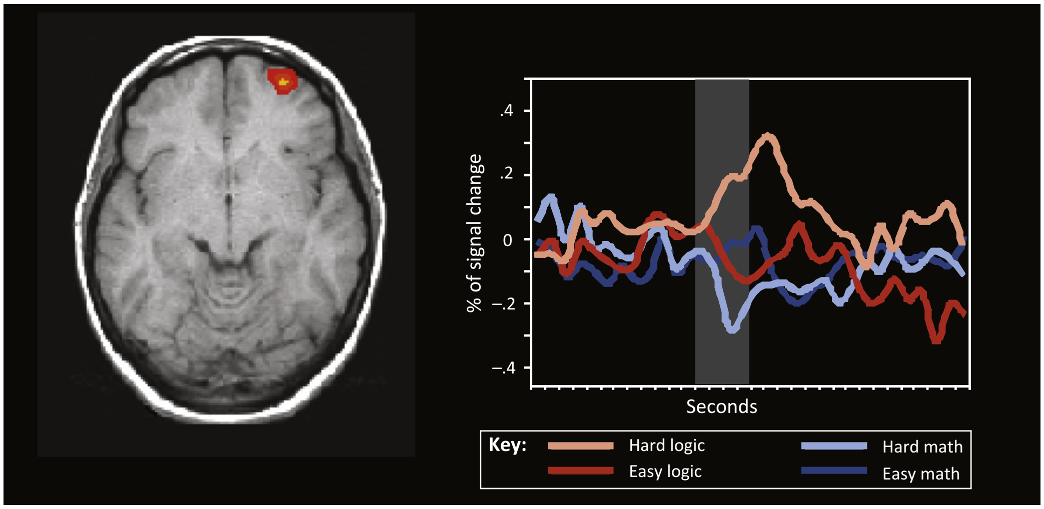

Figure 1. Activation in right frontal pole, Brodmann's area 10, while participants performed four sorts of task [91]. After the 8-s window of the presentation of the problems (shown in grey, allowing for the hemodynamic lag), hard logical inferences, which called for a search for counterexamples, activated this region more than the easy logical inferences did. There was no difference in activation between hard and easy mathematical problems and only the hard logical inferences showed activity above baseline. Reproduced, with permission, from [91].

Table 3. Studies corroborating the three principal predictions of the model theory that diverge from those of other theories[a]

Domain of reasoning	Use of counterexamples	Illusory inferences	Mental versus fully explicit models
Disjunctions	[89]	[85–87]	[133]
Conditionals	[89]	[84]	[31]
Spatial relations	–	[134]	[135–138]
Temporal relations	–	–	[139]
Causal relations	[105]	[106]	[106,107]
Syllogisms and other inferences with quantifiers	[88,91]	[140,141]	[88,111]

[a]Reasoners spontaneously use counterexamples to refute given invalid inferences, compelling illusory inferences occur when falsity matters, and valid inferences from mental models (system 1) are easier than those depending on alternative fully explicit models (system 2); these latter studies contrast the prediction with the length of formal derivations in theories based on logic [18,19]. The entries in the table are citations of the most recent papers only; – signifies that the authors know of no studies of the relevant prediction.

What we have not discussed so far are probabilistic deductions, and so the final sections of this review concern probabilities.

Do probabilities enter into pure deductions?

By 'pure' deductions, we mean those that make no reference to probabilities. P-logic, however, holds that probabilities are ubiquitous and that they unconsciously enter into pure deductions [21–23,25–29]. By contrast, the model theory implies that probabilities enter into the contents of reasoning only if invoked explicitly. We examine this proposal for conditional, syllogistic, and causal reasoning.

According to p-logic, a conditional such as:

> If the FDA approves a drug then it is safe for human consumption

asserts not a deterministic relation between approval and safety, but a probabilistic one [21,22]. Its inferential consequences should therefore be akin to those from the conditional:

> If the FDA approves a drug then probably it is safe for human consumption.

Several studies examining various inferential tasks show that reasoners distinguish between the two sorts of conditional [92]. Figure 2 presents striking differences between them that should not occur if pure conditionals are probabilistic [93].

For syllogisms, p-logic proposes that individuals assess p-validity (Box 1). However, a manipulation of the frequencies of the entities that the premises referred to did not corroborate p-logic [94]. By contrast, the mental model theory treats the meanings of syllogistic premises as expressing definite relations between sets (Table 4). System 1 in mReasoner uses the meaning of a premise to construct a mental model [60], which is iconic in that its structure corresponds to the structure of the sets it represents [95]. Hence, the premises:

> Some Greeks are men
> Some men are athletes

yield a mental model of the relevant individuals; for example, a model of four Greeks:

Greek	man	athlete
Greek	man	athlete
Greek	man	
Greek		

System 1 uses heuristics acquired from valid inferences to frame a conclusion:

> Some Greeks are athletes.

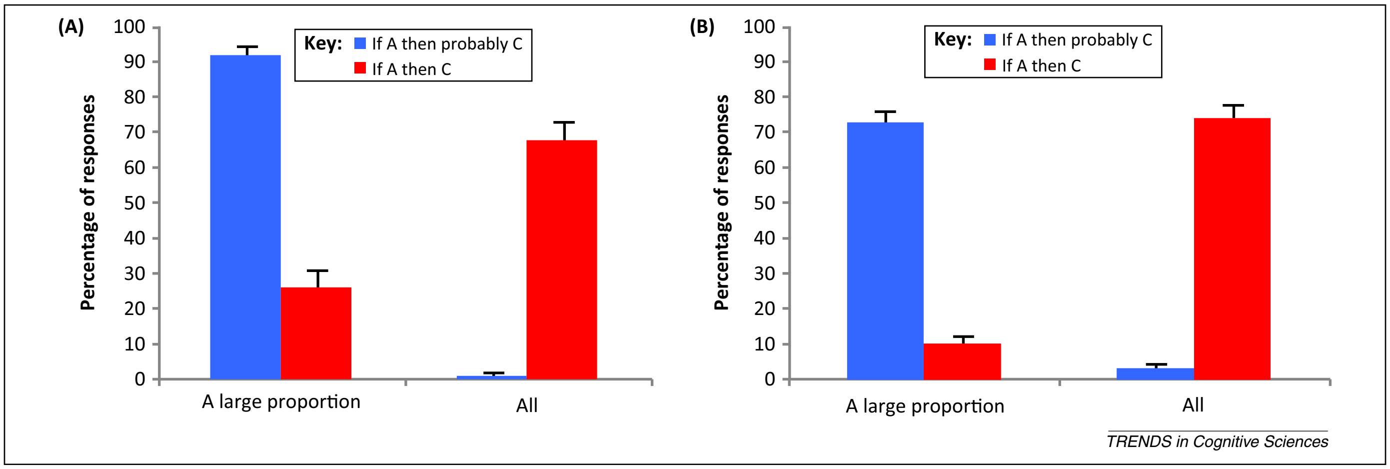

Figure 2. The results of two experiments contrasting judgments of pure conditionals and those referring to probability. **(A)** The percentages of participants who judged that 'all the *A*s are *C*s' and that 'a large proportion of *A*s are *C*s' for various everyday conditionals, 'if *A* then *C*', such as 'If the wine is Italian then it is red' and 'If the wine is Italian then it is probably red'. **(B)** The percentages of participants who judged that 'all the *A*s must be *C*s' and that 'a large proportion of *A*s must be *C*s' for the same sorts of conditional. From Experiments 5 and 6 in [93].

Reasoners will draw this conclusion (which p-logic also predicts) unless they use system 2 to search for a counterexample. System 2 finds a model in which the men who are Greeks are separate from the men who are athletes and so it withdraws the conclusion. In principle, there are nine responses to the pair of premises: eight sorts of conclusion (four quantifiers × two orders of the end terms 'Greeks' and 'athletes') and the response 'nothing follows'. A minimal goal for a theory is to predict which of these nine responses occur reliably, and which do not, for each of the 64 pairs of premises. A measure of the theory's accuracy is to express the sum of these two numbers as a proportion of the total number of possible responses (9 × 64). Figure 3 reports the results of such a meta-analysis [96] comparing various theories including logic, p-logic, and mental models. One account of syllogistic and other sorts of reasoning argues that 'many formal systems are required for psychology: classical logic, non-monotonic logics, probability logics, relevance logic, and others' [42]. However, this theory does not explain how the logics fit together or what its predictions are for each of the 64 syllogisms.

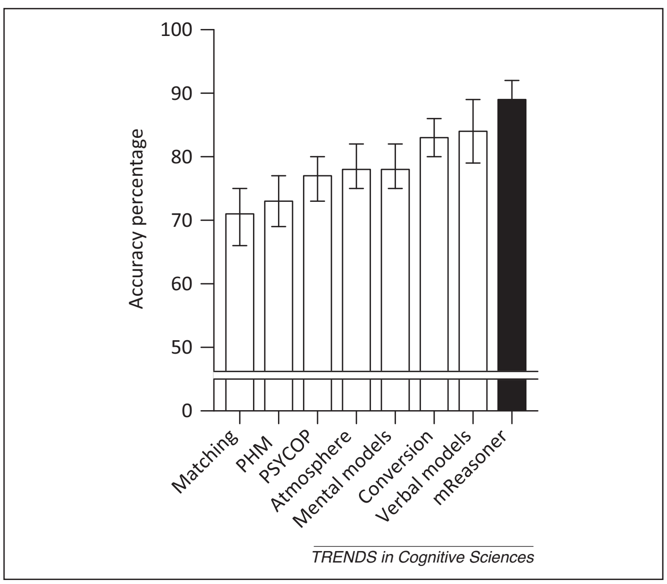

Figure 3. The percentages of accurate predictions of seven theories of syllogistic reasoning for six experiments [96]. The participants drew their own conclusions to the 64 possible pairs of premises. The accuracy measure described in the text had slight adjustments, as required in the meta-analysis standards of the American Psychological Association (APA), to reflect the numbers of participants in each experiment and the heterogeneity of the theory over different experiments. The seven theories are: the matching hypothesis that conclusions tend to match the quantifiers in the premises [142]; the probability heuristics model in which heuristics approximate p-valid conclusions (Box 1) [114]; logical rules of inference as embodied in the psychology of proof (PSYCOP) model [18]; the atmosphere hypothesis [143], which is a precursor to the matching hypothesis; the original mental model theory [88]; the theory that reasoners make illicit conversions of premises (e.g., from 'all *A* are *B*' to 'all *B* are *A*') [144]; verbal models theory in which verbal formulations are central and no search for alternative models occurs [145]; and the mReasoner implementation of the mental model theory [60]. Certain extant theories are missing from the figure because their predictions have not been published and their authors either did not regard them as full-fledged theories or were unable to make them available for analysis (for details, see [96]).

Since the advent of quantum mechanics, theorists have often advocated probabilistic accounts of causal relations, arguing that '*A* causes *C* to occur' means that the conditional probability of '*C* given *A*' is greater than some criterion such as the conditional probability of '*C* given not-*A*' [97–103]. Approaches of this sort fit nicely with Bayesian nets in artificial intelligence, which yield the probabilities of various events without the need to compute the complete joint probability distribution [104]. By contrast, the model theory [8,105–107] postulates a deterministic meaning for the causal assertion, which refers to three temporally ordered possibilities in which *A* does not occur after *C*:

A	*C*
not-*A*	not-*C*
not-*A*	*C*

The assertion '*A* enables *C* to occur' has a different meaning, referring to three temporally ordered possibilities in which *A* is necessary for *C* to occur:

A	*C*
A	not-*C*
not-*A*	not-*C*

The two relations accordingly refer to slightly different possibilities, but they both have the same mental model:

A C

Table 4. Treatment of quantified assertions as relations between sets *A* and *B*[a]

Sort of assertion	Its set-theoretic notation	An informal paraphrase of the notation
All *A* are *B*	$A \subseteq B$	*A* is included in *B*
Some *A* are *B*	$A \cap B \neq \emptyset$	The intersection of *A* and *B* is not empty
No *A* are *B*	$A \cap B = \emptyset$	The intersection of *A* and *B* is empty
Some *A* are not *B*	$A - B \neq \emptyset$	The set of *A* that are not B is not empty
At least three *A* are *B*	$\mid A \cap B \mid \geq 3$	The cardinality of the intersection of *A* and *B* ≥ 3
Most *A* are *B*	$\mid A \cap B \mid > \mid A - B \mid$	The cardinality of the intersection of *A* and *B* $>$ the cardinality of the *A* that are not *B*
More than half of the *A* are *B*	$\mid A \cap B \mid > \mid A \mid /2$	The cardinality of the intersection of *A* and *B* $>$ half of the cardinality of *A*

[a]Each sort of assertion has other equivalent assertions [60]. This account handles all of the quantifiers in natural language, including those such as 'more than half of the Greeks' that cannot be defined using the standard quantifiers ranging over individuals in logic [126].

This identity may explain why so many theorists have followed John Stuart Mill [108] in denying a difference in meaning between the two relations. However, when individuals list what is possible given the two sorts of assertion, they distinguish their meanings in the predicted way. They also succumb to illusory inferences about these relations and take a single counterexample, '*A* & not-*C*', to refute a causal claim rather than an enabling claim (Table 3). These phenomena are contrary to a probabilistic interpretation. Indeed, as a leading proponent of Bayesian nets emphasizes [49], probabilities cannot elucidate the falsity of the claim:

Mud causes rain

because a correlation does not distinguish between cause and effect. Of course, a correlation may provide evidence for causation, but evidence for a proposition should not be confused with the meaning of the proposition.

The probabilistic machinery of reasoning

At this point, readers may suspect that the model theory is deterministic through and through. In fact, it postulates that the machinery underlying reasoning – even for pure deductions – is probabilistic [30]. One illustration concerns evaluations of consistency – an important task because inconsistent beliefs can lead to disaster [109,110]. The only general way to use formal rules of inference [18] to establish the consistency of a set of assertions is to show that the negation of one assertion in the set cannot be proved from the remaining assertions. No psychologist has endorsed such an implausible procedure. In p-logic, two assertions are p-consistent if they can both be highly probable. Hence, to show that a set of assertions is consistent, one needs to establish that all of the assertions can be highly probable [47]. It is unclear how this task is to be done. In the mental model theory, a set of assertions is consistent if they all have a model in common [8]. We use inferences about consistency as an illustration of how probabilities can be part of the machinery of reasoning.

Consider this problem:

All Greeks are athletes.
Some athletes are Greeks.
Can both of these assertions be true at the same time? (Yes.)

A mental model of the first premise, which here represents three possible individuals:

```
Greek   athlete
Greek   athlete
Greek   athlete
```

also satisfies the second assertion and so it is easy to infer that both assertions can be true at the same time. As in deduction proper, problems that depend on a mental model only are easier than those that call for an alternative, fully explicit model [111]. This inference, for example, is harder because it calls for such a model:

All Greeks are athletes.
Some athletes are not Greek.
Can both of these assertions be true at the same time? (Yes.)

The model has to take this form:

```
Greek   athlete
Greek   athlete
Greek   athlete
        athlete
        athlete
```

Sets of the two sorts of inference yielded the predicted difference, but within each set the frequencies of endorsement varied [112,113]. Perhaps the probabilistic meanings of p-logic [114] could explain this variation. The model theory takes a different tack. As Figure 4 explains, probabilities determine the numbers and sorts of individuals in models and whether a search is made for fully explicit models [112]. The values of the parameters in Figure 4 provided a good fit with the data (Experiment 2 in [112]). The mReasoner program accordingly accounts for variations within system 1 and within system 2 inferences and for the difference between them. The machinery for pure deductions can be probabilistic even when the contents of inferences are not.

Reasoning about probabilities

The model theory can explain how people reason about probabilities of various sorts. Consider the following inference:

A sign of a particular viral infection – a peculiar rash – occurs only in patients who are infected, but some patients with the infection do not have the rash. Is the infection more likely than the rash? (Yes.)

This deduction follows from the mental model of the possible individuals. Here is a numerical example:

There is a box in which there is at least a red marble, or else there is a green marble and there is a blue marble, but not all three marbles.

Given the preceding assertion, what is the probability of the following situation?

In the box there is a green marble and a blue marble.

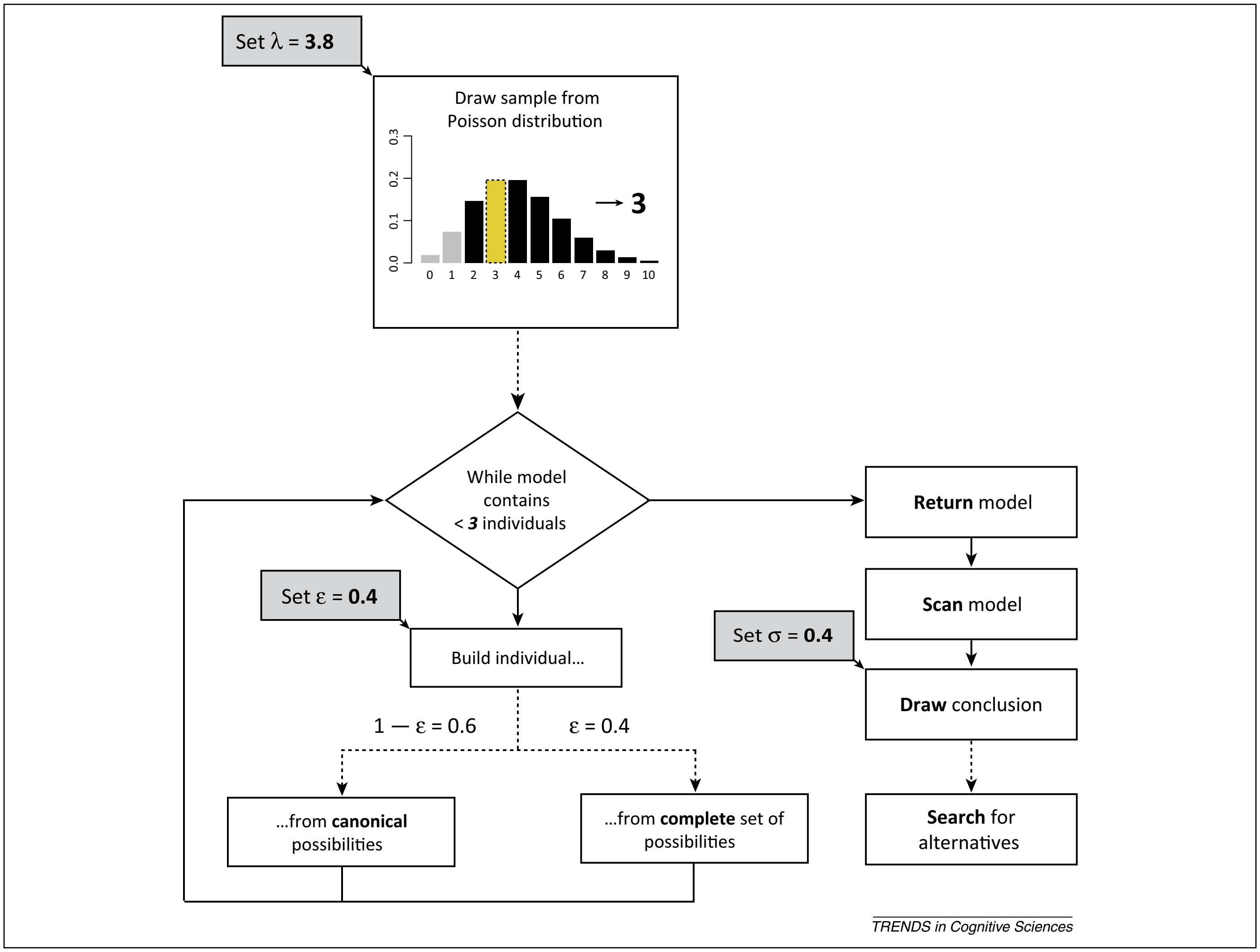

Figure 4. The program implemented in mReasoner for constructing an initial mental model of quantified assertions such as 'All the Greeks are athletes' [112]. As the first box in the diagram shows, the program begins by choosing the number of individuals to be represented in the model. This number is drawn at random from a small Poisson distribution whose mean and variance is set by the parameter λ, which in this instance equals 3.8. The number that is then drawn at random in this case equals 3. Hence, while the number of individuals in the model is less than 3, as shown in the decision in the diamond, the program constructs a new individual. This individual is either one from the complete set of possibilities compatible with the meaning of the assertion or one from the 'canonical' set corresponding to its mental model [88]. The probability of choosing from the complete set is fixed by the value of the parameter ε, which in this case equals 0.4. A mental model for 'All the Greeks are athletes' contains only individuals that are both Greeks and athletes, whereas the complete set includes in addition those that are neither Greeks nor athletes and those that are not Greeks and athletes. The final process scans the model to draw an initial conclusion and a third parameter, σ, which in this case is set to 0.4, determines the probability that the program then searches for and constructs an alternative model. Unbroken lines denote deterministic processes and broken lines denote probabilistic processes.

The mental models of the premise represent two possibilities:

red

green blue

Most participants in an experiment likewise inferred that the answer to the question was 50% [115]. When the question asked for the probability that the box contained a red marble and a blue marble, as the preceding models predict, most participants inferred a probability of zero. The inference is illusory. When the red marble is in the box, it is false that there is both a green and a blue marble. One way that this proposition can be false is when there is a blue marble without a green marble. So, red and blue marbles can co-occur.

In simple cases, infants, children, and non-numerate adults can infer probabilities from models of possibilities [116–118], and numerate adults can infer them from kinematic models. Given a randomly stacked tower of blocks, individuals can simulate what will happen, allowing for the asymmetry of the arrangement, and infer the probability that the blocks will tumble and the direction that they are most likely to fall [119]. Their simulations are unlikely to make accurate Newtonian predictions for every sort of problem [120]. Nevertheless, as in the rearrangements of trains (Box 2), kinematic models seem the most plausible basis for these inferences.

Neither frequencies nor physical asymmetries are applicable to some probabilities; for example:

What is the probability that advances in genetics will end the shortage of replacement organs in the next 15 years?

Some theorists regard such probabilities as improper [50,51], yet individuals happily infer them. In one study, the mean estimate for the preceding example was close to 38% [121]. The puzzle is where the numbers come from.

Box 4. How reasoners infer the probabilities of unique events

Individuals happily make numerical estimates of probabilities that cannot be assessed from any definitive set of frequencies; for example, the probability that Apple is profitable next year. The model theory [121,146] postulates that reasoners infer such probabilities from pertinent evidence, such as:

Most companies that are profitable one year are profitable the next year.

They construct a mental model from this evidence:

profitable this year	profitable next year
profitable this year	profitable next year
profitable this year	profitable next year
profitable this year	

Given Apple's profitability this year, system 1 translates the proportion in the model into a primitive analog representation of the sort postulated to represent magnitudes in non-numerate individuals [147–149]:

|–––––– |

The left-hand end represents impossibility and the right-hand end represents certainty. System 1 can translate this icon into a verbal description of the sort that non-numerate individuals use:

It's highly likely.

The icon can be pushed one way or another by other evidence and can represent the probability of a conjunction, a disjunction, or one event given another. The principle that human reasoners follow is that uncertainty translates into a lower probability. The process, which is deductive in nature, calls for minimal computational power. As system 1 in mReasoner shows, loops of a small fixed number of iterations suffice [146]. They combine evidence using a repertoire of primitive quasi-arithmetical operations of the sort that infants and adults in non-numerate cultures rely on [147–149]. One such operation resolves the uncertainty of divergent evidence by taking a primitive average of the analog magnitudes. System 2 maps these analog representations into numerical probabilities. Many estimates, in principle, fix the complete joint probability distribution, including each of the following three sets:

$p(A),\ p(B),\ p(A \text{ and } B)$

$p(A),\ p(B),\ p(A \text{ or } B, \text{ or both})$

$p(A),\ p(B),\ p(A|B)$

Figure I presents a summary of experimental results corroborating the primitive methods for resolving uncertainty and their predictions about participants' estimates of conjunctive, disjunctive, and conditional probabilities [121,146]. Conflicting estimates of $p(A)$ and $p(B)$ create uncertainty about both the probability of their conjunction and the probability of their disjunction. Reasoners therefore often assess the probabilities of both these compounds as a primitive average of the probabilities of their constituents, $p(A)$ and $p(B)$. They often estimate the conditional probability of $p(A|B)$ by nudging $p(A)$ a minimal amount depending on whether B increases or decreases its probability.

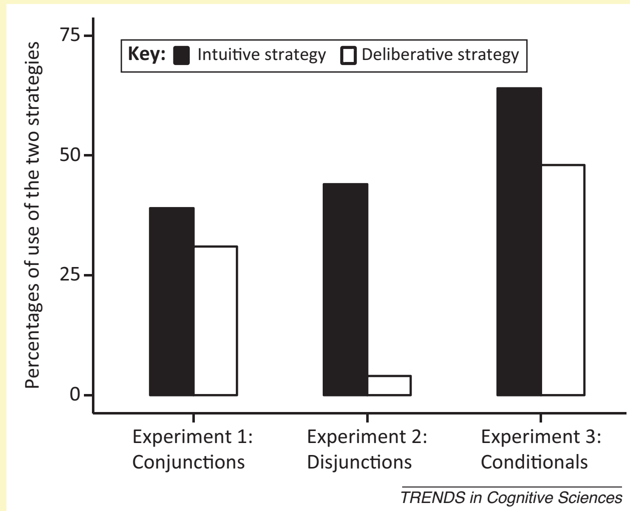

Figure I. The percentages of participants' estimates corroborating the model theory's intuitive strategy (system 1) and its deliberative strategy (system 2) for inferring the probabilities of unique events [146]. For Experiment 1, the intuitive strategy is that $p(A \& B)$ is a primitive average of $p(A)$ and $p(B)$ to resolve the conflict between them and the deliberative strategy is a multiplication of the two probabilities, which presupposes that they are independent, although in the experiment they were not independent. For Experiment 2, the intuitive strategy is that $p(A$ or B or both) is a primitive average of $p(A)$ and $p(B)$ to resolve the uncertainty between them and the deliberative strategy is the sum of the two probabilities except where it exceeds 100%, which presupposes that they are in an exclusive disjunction, although in the experiment they were not. For Experiment 3, the intuitive strategy is that a conditional probability, $p(B|A)$, is a small adjustment of $p(B)$ depending on whether A increases or decreases it and the deliberative strategy is to compute the ratio of $p(A \& B)$ to $p(A)$; that is, in accordance with Bayes' theorem.

P-logic, as yet, does not provide a solution. Box 4 shows how mental models solve the puzzle.

Concluding remarks

We began with two questions: does logic underlie human deductions and how do probabilities fit together with them? Despite the importance of logic to mathematics and the theory of computability, unconscious logical rules do not appear to be the basis of everyday reasoning. Arguments for this claim motivated the new paradigm in which reasoners rely instead on probability logic. It focuses on conditional assertions and postulates that individuals assess them by imagining that their if-clauses are true and then estimating the likelihood that their then-clauses are true. In consequence, when their if-clauses are false, conditionals are neither true nor false – they have a defective truth table (Table 1). As a result, the degree to which people believe a conditional equals the corresponding conditional probability. The evidence for p-logic is not decisive – especially outside the domain of simple conditional reasoning – and its defective truth table yields bizarre proofs (such as our earlier proof for the existence of God). It has no account of many important domains of reasoning such as spatial, temporal, and relational reasoning; and, according to a recent critique, it fails utterly with chains of inference [122]. The theory of mental models explains these various sorts of reasoning and appears to solve the difficulties that beset logical and p-logical accounts. It proposes that people can envisage possibilities, which they can also use to estimate probabilities. The mechanism for deduction, however, is probabilistic.

So far, no evidence has overturned the mental model theory's main predictions: the neglect of falsity in mental models leads to systematic fallacies, counterexamples occur spontaneously in the evaluation of inferences as invalid, and deductions depending only on mental models (system 1) are easier than those depending on a fully explicit model (system 2). These predictions have been corroborated for most domains of reasoning, although gaps in knowledge remain (Table 3). The theory is not a paragon

Box 5. Outstanding questions

- How robust is the Equation that *p*(if *A* then *C*) equals *p*(*C*|*A*) (see Table 2 in main text)? Does it depend on the use of Ramsey's test? If so, then, as some p-logicians have wondered [21], how is the test computed?
- Multiple quantifiers occur in assertions; for example, 'all the CEOs that have some assistants who know all their firm's plants are happy'. The logic required for these assertions is more complex than sentential or syllogistic logic [1] and there are a few psychological studies of the domain [20,31]. How do individuals make inferences from such assertions?
- According to the mental model theory, conditionals refer to a conjunction of possibilities. So, a single piece of evidence such as 'not-*A* & *C*' cannot be used to determine whether a conditional, 'if *A* then *C*', is true or false. Is the apparent evidence for the defective truth table a result of posing an impossible task to participants [150,151]?
- Do kinematic models underlie reasoning about causal relations and counterfactuals?
- Some dual-process theories of reasoning contrast inferences based on probabilities and those based on counterexamples [152], and evidence corroborates these theories [153,154]. However, other dual-process theories contrast emotion and reasoning [155] and bias from beliefs and no bias from beliefs [156]. Is it feasible to integrate all of these accounts?
- Massive differences occur in deductive ability and in susceptibility to bias from beliefs [156]. Some of the variance reflects the processing capacity of working memory (e.g., [157–159]). What accounts for the remaining variance?
- What are the origins of deductive competence? Past theorists have argued that it is innate or else constructed from experience. How could either of these processes work?
- Psychology has myriad theories of syllogistic reasoning, myriad theories of conditional reasoning, and so on. It is very difficult to get rid of theories. How can we solve this crisis?
- Leibniz dreamed of a calculus that settles any argument. Can cognitive scientists devise such a system?

(see the results in [123–125]). It leaves open several outstanding questions (Box 5). However, it does argue that counterexamples are fundamental to human rationality. So, if counterexamples to its principal predictions occur, the theory will at least explain its own refutation.

Acknowledgments

The authors thank Ruth Byrne, Rebecca Schwarzlose (*TiCS* editor), and two anonymous reviewers for their constructive criticisms of an earlier draft. They are also grateful to Igor Douven, Niki Pfeifer, Gernot Kleiter, and Klaus Oberauer for helping them to clarify their views about the four key hypotheses of the new paradigm. This research was supported by a Jerome and Isabella Karle Fellowship from the Naval Research Laboratory to S.S.K.

References

1 Jeffrey, R. (1981) *Formal Logic: Its Scope and Limits*. (2nd edn), McGraw-Hill
2 Bucciarelli, M. *et al.* (2008) The psychology of moral reasoning. *Judgm. Decis. Mak.* 3, 121–139
3 Lee, N.Y.L. *et al.* (2008) The psychological problem of Sudoku. *Think. Reason.* 14, 342–364
4 Lee, N.Y.L. and Johnson-Laird, P.N. (2013) Strategic changes in problem solving. *J. Cogn. Psychol.* 25, 165–173
5 Khemlani, S.S. *et al.* (2013) Kinematic mental simulations in abduction and deduction. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16766–16771
6 Lee, N.Y.L. and Johnson-Laird, P.N. (2013) A theory of reverse engineering and its application to Boolean systems. *J. Cogn. Psychol.* 25, 365–389
7 Beth, E.W. and Piaget, J. (1966) *Mathematical Epistemology and Psychology*, Reidel
8 Johnson-Laird, P.N. (2006) *How We Reason*, Oxford University Press
9 Baron, J. (2008) *Thinking and Deciding*. (4th edn), Cambridge University Press
10 Nickerson, R. (2011) *Mathematical Reasoning: Patterns, Problems, Conjectures, and Proofs*, Taylor & Francis
11 Blanchette, E. and Richards, A. (2010) The influence of affect on higher level cognition: a review of research on interpretation, judgement, decision-making and reasoning. *Cogn. Emot.* 15, 561–595
12 Gangemi, A. *et al.* (2013) Models and cognitive change in psychopathology. *J. Cogn. Psychol.* 25, 157–164
13 Kaufman, S.B. *et al.* (2011) General intelligence predicts reasoning ability even for evolutionarily familiar content. *Intelligence* 39, 311–322
14 Kyllonen, P.C. and Christal, R.E. (1990) Reasoning ability is (little more than) working memory capacity?! *Intelligence* 14, 389–433
15 Klauer, K.C. *et al.* (1997) Working memory involvement in propositional and spatial reasoning. *Think. Reason.* 3, 9–47
16 Osherson, D.N. (1974–1976) In *Logical Abilities in Children* (Vols1–4), Erlbaum
17 Macnamara, J. (1986) *A Border Dispute: The Place of Logic in Psychology*, MIT Press
18 Rips, L.J. (1994) *The Psychology of Proof*, MIT Press
19 Braine, M.D.S. and O'Brien, D.P., eds (1998) *Mental Logic*, Erlbaum
20 Rips, L.J. (2002) Reasoning. In *Stevens' Handbook of Experimental Psychology, Vol. 2: Memory and Cognitive Processes* (3rd edn) (Medin, D., ed.), pp. 317–362, Wiley
21 Oaksford, M. and Chater, N. (2007) *Bayesian Rationality*, Oxford University Press
22 Oaksford, M. and Chater, N. (2009) Précis of *Bayesian Rationality*: the probabilistic approach to human reasoning. *Behav. Brain Sci.* 32, 69–120
23 Pfeifer, N. and Kleiter, G.D. (2009) Framing human inference by coherence based probability logic. *J. Appl. Logic* 7, 206–217
24 Keynes, J.M. (1921) *A Treatise on Probability*, Macmillan
25 Over, D.E. (2009) New paradigm psychology of reasoning. *Think. Reason.* 15, 431–438
26 Evans, J.St.B.T. (2012) Questions and challenges for the new psychology of reasoning. *Think. Reason.* 18, 5–31
27 Elqayam, S. *et al.* (2013) Rationality in the new paradigm: strict versus soft Bayesian approaches. *Think. Reason.* 19, 453–470
28 Pfeifer, N. (2013) The new psychology of reasoning: a mental probability logical perspective. *Think. Reason.* 19, 329–345
29 Oaksford, M. and Chater, N. (2013) Dynamic inference and everyday conditional reasoning in the new paradigm. *Think. Reason.* 19, 346–379
30 Johnson-Laird, P.N. (1983) *Mental Models*, Harvard University Press
31 Johnson-Laird, P.N. and Byrne, R.M.J. (1991) *Deduction*, Erlbaum
32 Goodwin, G.P. and Johnson-Laird, P.N. (2011) Mental models of Boolean concepts. *Cogn. Psychol.* 63, 34–59
33 Johnson-Laird, P.N. and Khemlani, S.S. (2013) Toward a unified theory of reasoning. In *The Psychology of Learning and Motivation* (Ross, B., ed.), pp. 1–42, Elsevier
34 Koralus, P. and Mascarenhas, S. (2013) The erotetic theory of reasoning: bridges between formal semantics and the psychology of deductive inference. *Philos. Perspect.* 27, 312–365
35 Boolos, G. and Jeffrey, R. (1989) *Computability and Logic*. (3rd edn), Cambridge University Press
36 Turing, A.M. (1937) On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc. Ser.* 2, 230–265
37 Davis, M. (2000) *Engines of Logic: Mathematicians and the Origin of the Computer*, Norton
38 Antoniou, G. (1997) *Nonmonotonic Reasoning*, MIT Press
39 Pijnacker, J. *et al.* (2011) Reasoning with exceptions: an event-related brain potentials study. *J. Cogn. Neurosci.* 23, 471–480
40 Baggio, G. *et al.* (2014) Logic as Marr's computational level: four case studies. *Top. Cogn. Sci.* Published online November 24, 2014. http://dx.doi.org/10.1111/tops.12125
41 Oaksford, M. and Chater, N. (2014) Probabilistic single function dual process theory and logic programming as approaches to non-monotonicity in human vs artificial reasoning. *Think. Reason.* 20, 269–295
42 Achourioti, T. *et al.* (2014) The empirical study of norms is just what we are missing. *Front. Psychol.* 5, 1–15

43 Orenes, I. and Johnson-Laird, P.N. (2012) Logic, models, and paradoxical inferences. *Mind Lang.* 27, 357–377
44 Johnson-Laird, P.N. (2010) Against logical form. *Psychol. Belg.* 50, 193–221
45 Ramsey, F.P. (1990) Truth and probability. In *Philosophical Papers* (Mellor, D.H., ed.), pp. 52–109, Cambridge University Press
46 de Finetti, B. (1995) The logic of probability (transl. Angell, R.B. of 1936 original). *Philos. Stud.* 77, 181–190
47 Adams, E.W. (1998) *A Primer of Probability Logic*, Center for the Study of Language and Information
48 Tversky, A. and Kahneman, D. (1983) Extension versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90, 292–315
49 Pearl, J. (2009) *Causality: Models, Reasoning, and Inference*. (2nd edn), Cambridge University Press
50 Cosmides, L. and Tooby, J. (1996) Are humans good intuitive statisticians after all? Rethinking some conclusions of the literature on judgment under uncertainty. *Cognition* 58, 1–73
51 Gigerenzer, G. and Marewski, J.N. (2014) Surrogate science: the idol of a universal method for scientific inference. *J. Manag.* Published online September 2, 2014. http://dx.doi.org/10.1177/0149206314547522
52 Baratgin, J. *et al.* (2013) Uncertainty and de Finetti tables. *Think. Reason.* 19, 308–328
53 Baratgin, J. *et al.* (2014) New psychological paradigm for conditionals and general de Finetti tables. *Mind Lang.* 29, 73–84
54 Fugard, A.J. *et al.* (2011) How people interpret conditionals: shifts toward the conditional event. *J. Exp. Psychol. Learn.* 37, 635–648
55 Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Communication*, University of Illinois
56 Bar-Hillel, Y. (1964) *Language and Information Processing*, Addison-Wesley
57 Johnson-Laird, P.N. (1988) A taxonomy of thinking. In *The Psychology of Human Thought* (Sternberg, R.J. and Smith, E.E., eds), pp. 429–457, Cambridge University Press
58 Heit, E. and Rotello, C.M. (2010) Relations between inductive reasoning and deductive reasoning. *J. Exp. Psychol. Learn.* 36, 805–812
59 Craik, K. (1943) *The Nature of Explanation*, Cambridge University Press
60 Khemlani, S. and Johnson-Laird, P.N. (2012) The processes of inference. *Argument Comput.* 4, 4–20
61 Verschueren, N. *et al.* (2005) A dual-process specification of causal conditional reasoning. *Think. Reason.* 11, 239–278
62 Kahneman, D. (2011) *Thinking Fast and Slow*, Farrar, Strauss, Giroux
63 Evans, J.St.B.T. *et al.* (2013) Dual-process theories of higher cognition advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241
64 Evans, J.St.B.T. (2014) Reasoning, biases and dual processes: the lasting impact of Wason (1960). *Q. J. Exp. Psychol. (Hove)* Published online August 27, 2014. http://dx.doi.org/10.1080/17470218.2014.914547
65 Johnson-Laird, P.N. *et al.* (2004) Reasoning from inconsistency to consistency. *Psychol. Rev.* 111, 640–661
66 James, W. (1907) *Pragmatism*, Longmans, Green
67 Walsh, C.R. and Johnson-Laird, P.N. (2009) Changing your mind. *Mem. Cognit.* 37, 624–631
68 Khemlani, S. and Johnson-Laird, P.N. (2011) The need to explain. *Q. J. Exp. Psychol. (Hove)* 64, 276–288
69 Byrne, R.M.J. (2005) *The Rational Imagination: How People Create Alternatives to Reality*, MIT Press
70 Dixon, J.E. and Byrne, R.M. (2011) "If only" counterfactual thoughts about exceptional actions. *Mem. Cognit.* 39, 1317–1331
71 Johnson-Laird, P.N. and Byrne, R.M.J. (2002) Conditionals: a theory of meaning, pragmatics, and inference. *Psychol. Rev.* 109, 646–678
72 Johnson-Laird, P.N. and Tagart, J. (1969) How implication is understood. *Am. J. Psychol.* 82, 367–373
73 Schroyens, W. (2010) Mistaking the instance for the rule: a critical analysis of the truth-table evaluation paradigm. *Q. J. Exp. Psychol. (Hove)* 63, 246–259
74 Jeffrey, R.C. (1991) Matter of fact conditionals. *Aristotelian Soc. Suppl.* 65, 161–183
75 Evans, J. *et al.* (2013) Reasoning to and from belief: deduction and induction are still distinct. *Think. Reason.* 19, 267–283
76 Oberauer, K. and Wilhelm, O. (2003) The meaning(s) of conditionals: conditional probabilities, mental models and personal utilities. *J. Exp. Psychol. Learn.* 29, 688–693
77 Geiger, S.M. and Oberauer, K. (2010) Toward a reconciliation of mental model theory and probabilistic theories of conditionals. In *Cognition and Conditionals. Probability and Logic in Human Thinking* (Oaksford, M. and Chater, N., eds), pp. 289–307, Oxford University Press
78 Douven, I. and Verbrugge, S. (2010) The Adams family. *Cognition* 117, 302–318
79 Douven, I. and Verbrugge, S. (2013) The probabilities of conditionals revisited. *Cogn. Sci.* 37, 711–730
80 Cruz, N. and Oberauer, K. (2014) Comparing the meanings of "if" and "all". *Mem. Cognit.* 42, 1345–1356
81 Schroyens, W. *et al.* (2008) "The" interpretation(s) of conditionals. *Exp. Psychol.* 55, 173–181
82 Girotto, V. and Johnson-Laird, P.N. (2004) The probability of conditionals. *Psychologia* 47, 207–225
83 Barrouillet, P. and Gauffroy, C. (2015) Probability in reasoning: a developmental test on conditionals. *Cognition* 137, 22–39
84 Johnson-Laird, P.N. and Savary, F. (1999) Illusory inferences: a novel class of erroneous deductions. *Cognition* 71, 191–229
85 Khemlani, S. and Johnson-Laird, P.N. (2009) Disjunctive illusory inferences and how to eliminate them. *Mem. Cognit.* 37, 615–623
86 Goodwin, G. and Johnson-Laird, P.N. (2010) Conceptual illusions. *Cognition* 114, 253–265
87 Johnson-Laird, P.N. *et al.* (2012) The consistency of disjunctive assertions. *Mem. Cognit.* 40, 769–778
88 Bucciarelli, M. and Johnson-Laird, P.N. (1999) Strategies in syllogistic reasoning. *Cogn. Sci.* 23, 247–303
89 Johnson-Laird, P.N. and Hasson, U. (2003) Counterexamples in sentential reasoning. *Mem. Cognit.* 31, 1105–1113
90 Schroyens, W. and Schaeken, W. (2008) Deductive rationality in validating and testing conditional inferences. *Can. J. Exp. Psychol.* 62, 163–173
91 Kroger, J.K. *et al.* (2008) Distinct neural substrates for deductive and mathematical processing. *Brain Res.* 1243, 86–103
92 Markovits, H. and Thompson, V. (2008) Different developmental patterns of simple deductive and probabilistic inferential reasoning. *Mem. Cognit.* 36, 1066–1078
93 Goodwin, G.P. (2014) Is the basic conditional probabilistic? *J. Exp. Psychol. Gen.* 143, 1214–1241
94 Oberauer, K. *et al.* (2004) Logical reasoning and probabilities: a comprehensive test of Oaksford and Chater (2001). *Psychol. Bull. Rev.* 11, 521–527
95 Peirce, C.S. (1931–1958) . In *Collected Papers of Charles Sanders Peirce* (Hartshorne, C. *et al.*, eds), Harvard University Press
96 Khemlani, S. and Johnson-Laird, P.N. (2012) Theories of the syllogism: a meta-analysis. *Psychol. Bull.* 138, 427–457
97 Suppes, P. (1970) *A Probabilistic Theory of Causality*, North-Holland
98 Cheng, P.W. (1997) From covariation to causation: a causal power theory. *Psychol. Rev.* 104, 367–405
99 Sloman, S.A. (2005) *Causal Models: How We Think About the World and its Alternatives*, Oxford University Press
100 Sloman, S. *et al.* (2009) A causal model theory of the meaning of cause, enable, and prevent. *Cogn. Sci.* 33, 21–50
101 Fernbach, P.M. and Erb, C.D. (2013) A quantitative causal model theory of conditional reasoning. *J. Exp. Psychol. Learn.* 39, 1327–1343
102 Rehder, B. (2014) Independence and dependence in human causal reasoning. *Cogn. Psychol.* 72, 54–107
103 Rehder, B. (2014) The role of functional form in causal-based categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* Published online August 11. http://dx.doi.org/10.1037/xlm0000048
104 Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufman
105 Frosch, C.A. and Johnson-Laird, P.N. (2011) Is everyday causation deterministic or probabilistic? *Acta Psychol. (Amst.)* 137, 280–291
106 Johnson-Laird, P.N. and Goldvarg-Steingold, E. (2007) Models of cause and effect. In *The Mental Models Theory of Reasoning: Refinement and Extensions* (Schaeken, W. *et al.*, eds), pp. 167–189, Erlbaum
107 Khemlani, S. *et al.* (2014) Causal reasoning with mental models. *Front. Hum. Neurosci.* 8, 1–15

108 Mill, J.S. (1874) *A System of Logic, Ratiocinative and Inductive*. (8th edn), Harper
109 Johnson-Laird, P.N. (2012) Mental models and consistency. In *Cognitive Consistency* (Gawronski, B. and Strack, F., eds), pp. 225–244
110 Khemlani, S. and Johnson-Laird, P.N. (2012) Hidden conflicts: explanations make inconsistencies harder to detect. *Acta Psychol. (Amst.)* 139, 486–491
111 Ragni, M. *et al.* (2014) The evaluation of the consistency of quantified assertions. *Mem. Cognit.* 42, 53–66
112 Khemlani, S. *et al.* (2015) Immediate inferences from quantified assertions. *Q. J. Exp. Psychol. (Hove)* (in press)
113 Forgues, H.L. and Markovits, H. (2012) In search of the elusive initial model. *Exp. Psychol.* 59, 322–331
114 Chater, N. and Oaksford, M. (1999) The probability heuristics model of syllogistic reasoning. *Cogn. Psychol.* 38, 191–258
115 Johnson-Laird, P.N. *et al.* (1999) Naive probability: a mental model theory of extensional reasoning. *Psychol. Rev.* 106, 62–88
116 Téglás, E. *et al.* (2008) Pure reasoning in 12-month-old infants as probabilistic inference. *Science* 332, 1054–1059
117 Girotto, V. and Gonzalez, M. (2008) Children's understanding of posterior probability. *Cognition* 106, 325–344
118 Fontanari, L. *et al.* (2014) Probabilistic cognition in two indigenous Mayan groups. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17075–17080
119 Battaglia, P.W. *et al.* (2013) Simulation as an engine of physical scene understanding. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18327–18332
120 Marcus, G.F. and Davis, E. (2013) How robust are probabilistic models of higher-level cognition? *Psychol. Sci.* 24, 2351–2360
121 Khemlani, S. *et al.* (2012) The probability of unique events. *PLoS ONE* 7, 1–9
122 Hlobil, U. (2015) Chains of inference and the new paradigm in the psychology of reasoning. *Rev. Philos. Psychol.* Published online January 23, 2015. http://dx.doi.org/10.1007/s13164-015-0230-y
123 Monti, M. *et al.* (2009) The boundaries of language and thought in deductive inference. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12554–12559
124 Bonnefon, J.F. and Vautier, S. (2008) Defective truth tables and falsifying cards: two measurement models yield no evidence of an underlying fleshing-out propensity. *Think. Reason.* 14, 231–243
125 Ali, N. *et al.* (2011) The mental representation of causal conditional reasoning: mental models or causal models. *Cognition* 119, 403–418
126 Barwise, J. and Cooper, R. (1981) Generalized quantifiers and natural language. *Linguist. Philos.* 4, 159–219
127 Hegarty, M. (2004) Mechanical reasoning by mental simulation. *Trends Cogn. Sci.* 8, 280–285
128 Vergauwe, E. *et al.* (2013) Chronometric evidence for the dual-process mental model theory of conditional. *J. Cogn. Psychol.* 25, 174–182
129 Gauffroy, C. and Barrouillet, P. (2011) The primacy of thinking about possibilities in the development of reasoning. *Dev. Psychol.* 47, 1000–1011
130 Barrouillet, P. and Gauffroy, C. (2013) Dual processes and mental models in the development of conditional reasoning. In *The Developmental Psychology of Reasoning and Decision-Making* (Barrouillet, P. and Gauffroy, C., eds), pp. 95–121, Psychology Press
131 Quelhas, A.C. *et al.* (2010) The modulation of conditional assertions and its effects on reasoning. *Q. J. Exp. Psychol. (Hove)* 63, 1716–1739
132 Juhos, C. *et al.* (2012) Temporal and spatial relations in sentential reasoning. *Cognition* 122, 393–404
133 García-Madruga, J.A. *et al.* (2001) Are conjunctive inferences easier than disjunctive inferences? A comparison of rules and models. *Q. J. Exp. Psychol. A* 54, 613–632
134 Mackiewicz, R. and Johnson-Laird, P.N. (2012) Reasoning from connectives and relations between entities. *Mem. Cognit.* 40, 266–279
135 Brüssow, S. *et al.* (2013) Premise annotation in mental model construction: an ACT-R approach to processing indeterminacy in spatial relational reasoning. *Cogn. Syst. Res.* 24, 52–61
136 Knauff, M. and Ragni, M. (2011) Cross-cultural preferences in spatial reasoning. *J. Cogn. Culture* 11, 1–21
137 Knauff, M. *et al.* (2013) Spatial belief revision. *J. Cogn. Psychol.* 25, 147–156
138 Ragni, M. and Knauff, M. (2013) A theory and a computational model of spatial reasoning with preferred models. *Psychol. Rev.* 120, 561–588
139 Schaeken, W. and Johnson-Laird, P.N. (2000) Strategies and temporal reasoning. *Think. Reason.* 6, 193–219
140 Yang, Y. and Johnson-Laird, P.N. (2000) How to eliminate illusions in quantified reasoning. *Mem. Cognit.* 28, 1050–1059
141 Kunze, N. *et al.* (2010) Illusions of consistency in quantified assertions. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (Ohlsson, S. and Catrambone, R., eds), pp. 2028–2032 Cognitive Science Society
142 Wetherick, N.E. and Gilhooly, K.J. (1995) "Atmosphere", matching, and logic in syllogistic reasoning. *Curr. Psychol.* 14, 169–178
143 Begg, I. and Denny, J.P. (1969) Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning. *J. Exp. Psychol.* 81, 351–354
144 Revlis, R. (1975) Two models of syllogistic reasoning: feature selection and conversion. *J. Verb. Learn. Verb. Behav.* 14, 180–195
145 Polk, T.A. and Newell, A. (1995) Deduction as verbal reasoning. *Psychol. Rev.* 102, 533–566
146 Khemlani, S.S. *et al.* (2014) Naive probability: model-based estimates of unique events. *Cogn. Sci.* Published online November 3, 2014. http://dx.doi.org/10.1111/cogs.12193
147 Barth, H. *et al.* (2006) Nonsymbolic arithmetic in adults and young children. *Cognition* 98, 199–222
148 Carey, S. (2009) *The Origin of Concepts*, Oxford University Press
149 Gordon, P. (2004) Numerical cognition without words: evidence from Amazonia. *Science* 306, 496–499
150 Sevenants, A. *et al.* (2012) Is the truth table task mistaken? *Think. Reason.* 18, 119–132
151 Sevenants, A. *et al.* (2013) Truth table task: working memory load, latencies, and perceived relevance. *J. Cogn. Psychol.* 25, 339–364
152 Verschueren, N. *et al.* (2005) Everyday conditional reasoning: a working memory-dependent tradeoff between counterexample and likelihood use. *Mem. Cognit.* 33, 107–119
153 Markovits, H. *et al.* (2012) More evidence for a dual-process model of conditional reasoning. *Mem. Cognit.* 40, 736–747
154 Markovits, H. *et al.* (2013) Direct evidence for a dual process model of deductive inference. *J. Exp. Psychol. Learn.* 39, 1213–1222
155 Gubbins, E. and Byrne, R.M.J. (2014) Dual processes of emotion and reason in judgments about moral dilemmas. *Think. Reason.* 20, 245–268
156 Ball, L.J. (2013) Microgenetic evidence for the beneficial effects of feedback and practice on belief bias. *J. Cogn. Psychol.* 25, 183–191
157 De Neys, W. (2014) Conflict detection, dual processes, and logical intuitions: some clarifications. *Think. Reason.* 20, 169–187
158 von Bastian, C.C. and Oberauer, K. (2013) Distinct transfer effects of training different facets of working memory capacity. *J. Mem. Lang.* 69, 36–58
159 Santamaría, C. *et al.* (2013) Deductive reasoning and metalogical knowledge in preadolescence: a mental model appraisal. *J. Cogn. Psychol.* 25, 192–200
160 Li, M. and Vitányi, P. (1997) *An Introduction to Kolmogorov Complexity and its Applications*. (2nd edn), Springer-Verlag