The probability of conditionals: A review

Miguel López-Astorga[1] · Marco Ragni[2,3] · P. N. Johnson-Laird[4,5]

Abstract

A major hypothesis about conditionals is the Equation in which the probability of a conditional equals the corresponding conditional probability: $p(if\ A\ then\ C) = p(C|A)$. Probabilistic theories often treat it as axiomatic, whereas it follows from the meanings of conditionals in the theory of mental models. In this theory, intuitive models (system 1) do not represent what is false, and so produce errors in estimates of $p(if\ A\ then\ C)$, yielding instead $p(A\ \&\ C)$. Deliberative models (system 2) are normative, and yield the proportion of cases of A in which C holds, i.e., the Equation. Intuitive estimates of the probability of a conditional about unique events: *If covid-19 disappears in the USA, then Biden will run for a second term*, together with those of each of its clauses, are liable to yield joint probability distributions that sum to over 100%. The error, which is inconsistent with the probability calculus, is massive when participants estimate the joint probabilities of conditionals with each of the different possibilities to which they refer. This result and others under review corroborate the model theory.

Keywords Conditionals · Conditional probabilities · Mental models · Probabilities · The Equation · Subadditivity

Introduction

> Someone with only the most modest knowledge of probability mathematics could have won himself the whole of Gaul in a week.
> — Hacking (1975, p. 3).

✉ Miguel López-Astorga
  milopez@utalca.cl

  Marco Ragni
  ragni@sdu.dk

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

[1] Instituto de Estudios Humanísticos 'Juan Ignacio Molina', Universidad de Talca, Av. Lircay s/n, 3460000 Talca, Chile

[2] Danish Institute for Advanced Studies and Institute for Communication and Design, South Denmark University, Kolding, Denmark

[3] Cognitive Computation Laboratory, University of Freiburg, Freiburg, Germany

[4] Department of Psychology, Princeton University, Princeton, NJ 08540, USA

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

What are the probabilities of the following two conditionals?
If the next throw of this fair dice yields an even number then it will be a six.
If the US vaccination program fails then the Democrats will lose control of the Senate.

We invite readers to estimate both probabilities. In a distinction due to Tversky and Kahneman (1983), the first example calls for an *extensional* estimate, and the second example calls for an *intensional* estimate. Extensional estimates are based on knowledge of the frequencies with which mutually exclusive alternatives occur, or on the ability to infer them. Naive but numerate individuals can infer that given a fair toss yielding an even number, the probability of a six is 1/3, where "naïve" means only that they have not studied the probability calculus. Unique events have no prior frequencies of occurrence, and so they call for intensional estimates, which are based on evidence and sometimes heuristics, such as how representative an entity is of a category (Tversky & Kahneman, 1983). They can yield a "conjunction fallacy" in which the probability of a conjunction, $p(A\ \&\ C)$, is judged to be higher than the probability of a conjunct, $p(A)$ – a difference that is inconsistent with the probability calculus (Kolmogorov, 1950).

The contrast between extensional and intensional inferences mirrors the long controversy over whether probabilities refer to frequencies (e.g., Venn, 1888) or to degrees of belief (e.g., Ramsey, 1990a/1926). Some psychologists reject the probabilities of unique events as meaningless or outside the scope of the probability calculus (e.g., Cosmides & Tooby, 1996; Gigerenzer, 1994). Most others assume that individuals estimate their degree of belief in a proposition as a subjective probability. Quite how individuals make such estimates, however, is also a matter of debate. Controversy promotes research. And there are many papers on these issues (see, e.g., Baron, 2008; Nickerson, 2015).

The present article concerns the probabilities of conditionals, which it defines as any sentence with the grammatical structure in English of *if A then C*, where *A* and *C* are declarative clauses, or any sentence equivalent in meaning, such as, *C if A*. Its goal is to determine how people make such estimates, and whether they equal the corresponding conditional probabilities. This relation is the well-known Equation, so named in Edgington (1995), but a psychological conjecture first mooted in Rips and Marcus (1977):

The Equation: p(if A then C) = p(C|A).

The content or context of conditionals can elicit various estimates of their probabilities (Dieussaert et al., 2002). Because the probability of any assertion depends on its meaning, the present article begins with the main theories of their meanings: standard logic, probability logic, and mental models. Logic includes the sentential calculus that deals with conditionals, and we use the term "logic" throughout this article to refer to this standard account (e.g., Jeffrey, 1981). It treats conditionals as material implications, and so *if A then C* equals *not-A or C*, where "*not-A*" denotes the negation of *A*, and the disjunction is inclusive, i.e., both its clauses can hold. Probability logic differs (Adams, 1998). It is based on the assumption that the meaning of a conditional is the conditional probability in the *Equation*. The model theory has developed over many years, and so we distinguish its basic principles, applying to any sort of reasoning, from its account of the meanings of conditionals, which in the theory's original formulation were consistent with logic, but which later diverged from it in a radical way. The theory predicts that intuitions yield conjunctive estimates of the probabilities of conditionals, but deliberations yield normative estimates that fit the Equation. The paper aims to assess all the experiments in the literature that have investigated the Equation. Finally, it discusses the implications of their results for the different theories of conditionals.

Conditionals in logic and probability logic

In logic, the *partition* of a compound assertion, such as *if A then C*, is the set of all possible conjunctions of affirmations and negations of its two clauses:

$$A \ \& \quad C$$
$$A \ \& \ \neg \ C$$
$$\neg \ A \ \& \quad C$$
$$\neg \ A \ \& \ \neg \ C$$

where the symbol "&" here and later denotes logical conjunction in which *A & C* is true in case *A* is true and *C* is true, and is otherwise false; and the symbol "¬" denotes negation in which ¬ *A* is true in case *A* is false, and vice versa. The probabilities of the four conjunctions in the partition for a compound assertion make up its joint probability distribution (JPD). In the probability calculus, their sum equals 100%, because they are mutually exclusive and exhaustive (Kolmogorov, 1950). A *subadditive* function in arithmetic is one whose value as a whole is less than the sum of the values of its parts, for example, square-roots are subadditive: $\sqrt{(4 + 4)} < \sqrt{4} + \sqrt{4}$. In contrast, probabilities are additive, not subadditive:

$$p(A) = p(A \ \& \ C) + p(A \ \& \ \neg \ C).$$

Hence, p(*A & C*) can never be greater than p(A), and the conjunction fallacy is subadditive: it yields a JPD that can sum to 100% only if one or two of its cases have negative probabilities. Estimates that contain no conjunctions, however, can also produce a subadditive JPD, for example, p(*A*), p(*C*), and p(*C|A*). "Support" theory, which is due to Tversky and Koehler (1994), predicts that when descriptions are unpacked into separate disjunctive components, their probability estimates are subadditive, for example, probability estimates of deaths from heart disease, from cancer, or so on, are subadditive in comparison with those of deaths from natural causes. What determines the probability of a conditional, as we mentioned, is its meaning, and so we now consider conditionals in logic.

Conditionals in logic In (standard) logic, conditional assertions are truth functional, that is, their truth or falsity depends solely on the truth or falsity of their clauses. Hence, *material* conditionals, as they are known, are true for any case in their partition except the one in which their *if*-clause is true and their *then*-clause is false (see, e.g., Jeffrey, 1981, Ch. 4). Table 1 presents the corresponding truth table for the material conditional, and it also presents other tables that we come to, by and by. Some theorists accept the material conditional (e.g., Grice, 1989; but cf. Byrne & Johnson-Laird, 2019). It implies that the probability of a conditional, *If A then C*, should equal the probability of an inclusive disjunction, *not-A or C*, which has

Table 1 Three truth tables and one probability table for a conditional, *If A then C*: The material conditional of logic, the partial truth table with no truth-values when a conditional's *if*-clause is false (de Finetti, 1995/ 1935), the model theory's truth table, and Jeffrey's (1991) table of probabilities. Each row shows a conjunction in the partition, and, given its truth, the value of the conditional according to the four tables

The partition of cases		The material conditional	The partial truth-table	The model truth-table	The Jeffrey table of probabilities
A	C	True	True	True	1
A	¬ C	False	False	False	0
¬ A	C	True	Void	Possibly true & possibly false	p(C|A)
¬ A	¬ C	True	Void	Possibly true & possibly false	p(C|A)

the same truth table. But, in fact, estimates of the probability of a conditional seldom equal those of a material conditional. It also yields striking "paradoxes," for example, the falsity of its *if*-clause ensures that it is true (see Table 1).

Philosophical theories have treated conditionals as referring to "possible worlds" (Lewis, 1973; Stalnaker, 1968). Roughly speaking, a conditional is true if in a world in which its *if*-clause is true, and which otherwise doesn't differ from the actual world, its *then*-clause is also true. We go no further into details, because there are too many possible worlds relevant to the real world, and so – to paraphrase Partee's (1979) assessment – they cannot fit in anyone's head.

The probabilistic meaning of conditionals An influential view about conditionals is in a footnote describing what is now known as Ramsey's test (1990b/1929, p. 155). We quote it in its entirety:

> If two people are arguing "If *p*, then *q*?" and are both in doubt as to *p*, they are adding *p* hypothetically to their stock of knowledge and arguing on that basis about *q*; so that in a sense "If *p*, *q*" and "If *p*, *not-q*" are contradictories. We can say that they are fixing their degree of belief in *q* given *p*. If *p* turns out false, these degrees of belief are rendered *void*. If either party believes *not p* for certain, the question ceases to mean anything to him except as a question about what follows from certain laws or hypotheses.

The test concerns estimates of the subjective probabilities of conditionals, not their meanings, which Ramsey took to be truth-functional (but cf. Ramsey, 1990c). If so, why should people ignore the cases in which the *if*-clause is false, when they argue about conditionals? Ramsey's answer, we suspect, would be that the claim is obvious. His test is central to recent probabilistic theories (p-theories) of conditionals (e.g., Evans & Over, 2004; Fugard et al., 2011; Jeffrey, 1991; Oaksford & Chater, 2007; Over, 2009). These proponents refer to their theories as instances of a "new paradigm" in which subjective probabilities replace logic in theories of human reasoning. Not every probabilist, however, accepts the Equation implicit in Ramsey's test (e.g., Howson & Urbach, 1993, p. 82; Lewis, 1976). So, it is an assumption, not a fact.

Some theorists have transformed Ramsey's test into a semantics in which a conditional, *If A then C*, has only a partial truth table. It is true in case *A & C* holds, false in case *A & ¬ C* holds, but void in any other case, i.e., in any case in which *A* is false. One theorist has sometimes treated "void" as though it were a third truth-value (de Finetti, 1992/1936; see also Baratgin et al., 2013; Politzer et al., 2010), but at other times as an epistemological attitude (de Finetti, 1995/1935). Others have treated "void" as signifying that the truth function for conditionals returns no truth-value. Table 1 presents this *partial* truth table.

The partial truth table yields "paradoxes" too. As Table 1 shows, a conditional that is true implies that its *if*-clause is true too; likewise, a conditional that is false implies that its *if*-clause is true. People judge certain conditionals to be true, and certain others to be false (Quelhas et al., 2017). A conditional that they judge to be false is, for example:

If Anna has flu, then she is healthy.

And so it follows from the partial truth table that Anna has flu. Such inferences are absurd.

The partial truth table treats cases of ¬ *A* as void, and so they are irrelevant to the probability of a conditional. It therefore depends on the following ratio

$$p(\text{if } A \text{ then } C) = \frac{p(A\&C)}{p(A\&C) + p(A\&\neg C)}$$

And the ratio equals p(*C*|*A*). An ingenious solution that yields the Equation without treating the cases of ¬ *A* as void is Jeffrey's (1991) probability table. He argued that confidence in a conditional is defined in a table of probabilities (see Table 1), not truth-values (see also Sanfilippo et al., 2020). It blocks the "paradoxes" of the partial truth table, but does not state the cases in which conditionals are true or are false.

Adams ([1998](#)) goes one step further. He argues that conditionals do not have truth values. They have only imitation ones, which he refers to as "ersatz." He assumes that the meaning of a conditional is the conditional probability in the Equation. A corollary of this account is that conditionals cannot be embedded within other compounds or negated, because the negation of p(C|A) make no sense (ibid., p. 115). Likewise, inferences concerning conditionals cannot be valid in logic: they can only be probabilistically valid (p-valid). P-validity requires the uncertainty of a conclusion not to exceed the sum of the uncertainties of its premises, where the uncertainty of a proposition, A, equals 1 – p(A). Various inferences are valid in logic but not p-valid, such as the contrapositive inference: *If A then C*; therefore, *If not-C then not-A* (see also Bennett, [2003](#)).

Although p-logic avoids the paradoxes of material conditionals and of partial truth tables, it suffers from implausibility. Conjunctions and disjunctions have truth values, whereas conditionals do not (ibid., p. 187). And, as we have already illustrated, people do judge conditionals to be true or false. The following disjunction in logic:

> Wordsworth wrote *The Ancient Mariner,* or Coleridge did, or both of them did

refers to the same cases as the conditional:

> If Wordsworth didn't write *The Ancient Mariner* then Coleridge did.

They both rule out only one case in the partition – the one in which neither of the two poets wrote the poem. Individuals infer conditionals from disjunctions and vice versa (e.g., Richardson & Ormerod, [1997](#)). So, they infer an assertion with a truth-value from one without a truth value.

The cure seems no better than the earlier paradoxes, and it postulates the Equation without providing any grounds for it.

A characteristic of everyday reasoning, which was first addressed in artificial intelligence
(AI), is its non-monotonicity (or defeasibility):

> The term "non-monotonic logic" covers a family of formal frameworks devised to capture and represent *defeasible inference*. Reasoners draw conclusions defeasibly when they reserve the right to retract them in the light of further information (Strasser & Antonelli, [2019](#)).

Despite aiming for a defeasible logic, Adams failed to produce one. Hence, some p-theorists have imported the non-monotonic system P from AI into p-theories in order to capture defeasibility (e.g., Gilio, [2002](#); Pfeifer & Kleiter, [2003](#).)

In logic, given that *A* implies *C*, it follows that *if A then C*. An "inferentialist" account of the meaning of conditionals proposes the converse relation: a conditional is true given that *A* implies *C*, though the relation can be inductive rather than deductive (Douven, [2015](#); Douven et al., [2018](#); Ryle, [1949](#)). This hypothesis may explain the finding that the Equation seems to apply only when the probability of *A* is reasonably high (Skovgaard-Olsen et al., [2019](#)). When it is zero, the conditional probability p(C|A) is undefined in the probability calculus. The inferentialist hypothesis is a component of a probabilistic account of conditionals, but, unlike p-logic, it allows that conditionals can have truth values. Inferentialism, however, yields its own "paradox." It implies that conditionals can be used only to assert inferential relations. Yet, *If the company is in debt then it's not in their books*, makes no such assertion. Debts don't imply omission from the books.

Some proponents of p-theories take for granted the Equation as embodied in p-logic (see Johnson-Laird et al., [2015a](#), for a review). Some derive it from the partial truth table (e.g., Fugard et al., [2011](#); Oaksford & Chater, [2007](#)), or from de Finetti's notion of a conditional event (Fugard et al., [2011](#); Pfeifer, [2013](#); Pfeifer & Kleiter, [2007](#)), or from the Jeffrey table (Coletti & Scozzafava, [2002](#)). And it is to the credit of the approach that it has led to a burgeoning of studies of the probabilities of conditionals, and some of its proponents liken it to a paradigm shift in science (but cf. Knauff & Castañeda, [2021](#)). To succeed on Kuhn's ([1962](#)) account, a new paradigm needs to solve puzzles that its predecessors could not. But, as we show, the new paradigm has yet to do so, and it has even led to some results for which it offers no adequate explanation.

The theory of mental models

The great pioneer was Craik ([1943](#)), who argued that people base decisions on their mental models of the world. A more recent theory differs from Craik's conjecture. This theory proposes that perception and comprehension of discourse yield mental models, and so they also underlie reasoning (Johnson-Laird, [1970](#), [1975](#)), whereas Craik took reasoning to depend on verbal rules (ibid., p. 79). The newer theory treats models as iconic insofar as possible, i.e., their structure corresponds to the structure of what they represent, whereas Craik took models to have only the same input-output relations as what they represent, for example, Kelvin's tidal predictor, which has a structure remote from the moon, the earth, and the tides (ibid., p. 51). And, the newer theory allows for more than one model of premises, whereas Craik made no provision for multiple models.

In this section, we begin with the *model* theory's original basics, then introduce its progressive steps from logic to defeasible reasoning, from material conditionals to conjunctions of possibilities, and from modal logics to a semantics of *possibly A* in which it presupposes *possibly not-A*. These three steps, which each have empirical corroborations, lead to its account of the probability of conditionals and the Equation.

The basics The original model theory postulated several fundamental principles that still hold (see, e.g., Johnson-Laird, 1983). Reasoning is a semantic process. Its goal is a conclusion that maintains the meaning of the premises, and expresses it in a parsimonious conclusion that is new, not the regurgitation of a premise (ibid., p. 40). An inference is valid if its conclusion holds in all the models of the premises, i.e., it has no counterexample, which is a model of the premises in which the conclusion is false. Reasoning has two principal levels of expertise (ibid., Ch. 6) in what is now known as a "dual-process" theory. This idea is due to Wason (1966), and it was first implemented in an algorithm for his well-known "selection" task, which concerns the selection of evidence to test conditional hypotheses (Johnson-Laird & Wason, 1970). Its algorithm predicts numerous subsequent results (Ragni et al., 2018), and is more accurate than probabilistic accounts of the task (Evans, 1977; Kirby, 1994; Oaksford & Chater, 1996, 2007). Psychologists later developed other dual-process theories, for example, Evans (2008), Kahneman (2011), and Sloman (1996); and the approach even has its critics (e.g., Keren & Schul, 2009). Perhaps because Wason's account did not name two systems, it has been overlooked (see Manktelow, 2021, p. 3). So, the model theory refers to its *intuitive* level as system 1, and its *deliberative* level as system 2. System 1 lacks a working memory for intermediate results, which is available to system 2. As a consequence, system 2 makes correct inferences. It is normative, and decides for any putative inference whether its conclusion is necessary, possible, or impossible, bearing in mind the computational intractability of sentential reasoning (Cook, 1971), i.e., as the number of clauses in an inference increases, there comes a point where it is beyond the human brain to make the decision.

For a conditional, such as:

If it is raining, then it is hot

system 1 constructs two intuitive models (Johnson-Laird & Byrne, 1991, p. 47):

raining   hot
　　　　. . .

The first model represents the case in which it is raining and hot – we use words in these diagrams for simplicity. The second model, the ellipsis, has no explicit content, but allows for cases in which it is not raining. The further premise:

It is raining

allows reasoners to drop the second model, and to draw the conclusion of a modus ponens inference from the first model:

Therefore, it is hot.

Instead, the further premise for an inference of modus tollens:

It is not hot

calls for reasoners to drop the first model, but the ellipsis yields no conclusion, and so they respond:

Nothing follows.

This response is outside logic, which allows a countable infinity of valid conclusions from any premises, for example, a disjunction of the premises. System 2 can flesh out the intuitive models above into *explicit* models of the conditional (ibid., 48):

raining	hot
¬ raining	¬ hot
¬ raining	hot

where "¬ raining" denotes: *it is not raining*. The premise for modus tollens, *It is not hot*, eliminates the first and last models, and so reasoners can infer from the second model:

Therefore, it is not raining.

So, modus ponens is easier than modus tollens, and mental logic makes the same prediction, which experiments corroborate (e.g., Johnson-Laird & Byrne, 1991; Rips, 1994). The difference
between the two sorts of inference is a robust result.
The model theory's first basic prediction is:
1. *More models mean more work: inferences take longer, and are more prone to error.*
   A biconditional such as:

If, and only if, it is raining then it is hot

has only two explicit models:

raining	hot
¬ raining	¬ hot

Hence, it follows that modus tollens should be easier to infer from a biconditional than from a conditional. In contrast, mental logic yields the same formal proof from both sorts of premise (Rips, 1994, p. 178). Experiments corroborated the model theory (Johnson-Laird et al., 1992).
The theory's second basic prediction is:
2. *Certain premises yield systematic fallacies (or illusory inferences).*
The prediction follows from the theory's principle of truth, which postulates that system 1's intuitive models represent

what is true and ignore what is false. The fallacies can be so compelling that their system 2 corrections in a program implementing the theory seemed at first like a bug (Johnson-Laird & Savary, 1999). An early experiment tested extensional probabilities, such as:

If one of these assertions is true about a specific hand of cards then so is the other:

There is a jack if and only if there is a queen.
There is a jack.

The participants tended to estimate the probabilities of a jack and of a queen being in the hand as the same (Johnson-Laird & Savary, 1996). It's a fallacy, which follows from the intuitive model of both assertions being true:

jack	queen

However, both assertions could be false. In which case, the queen is bound to be in the hand, but the jack may not be. Such "illusory" inferences are a telltale sign of the use of intuitive models (Johnson-Laird et al., 2000).

The theory's third basic prediction is:

3. *The neglect of models of possibilities should yield erroneous conclusions that are consistent with the premises, but that do not follow from them.*

When there are many possibilities, it is easy to overlook one. The results in the laboratory are conclusions that take only some possibilities into account (e.g., Bauer & Johnson-Laird, 1993; Johnson-Laird et al., 1992). The results in daily life can be catastrophic, such as the capsizing of *The Herald of Free Enterprise*, a cross-channel car ferry. Its master overlooked the possibility that the bow-doors had not been closed, and the sea rushed into the vessel (Johnson-Laird, 2012).

As the theory developed, it has maintained these basic predictions and corroborated them in many studies. Its starting point was consistent with logic, and it treated the meaning of *if* as a material conditional (Johnson-Laird & Byrne, 1991). What changed are its semantics, its range of applications, and its account of reasoning.

Defeasible reasoning The theory's first step away from standard logic was to implement the defeasibility of inferences. From indeterminate premises about spatial relations, reasoners created a preferred model, which they might adjust to maintain consistency with a later premise (Johnson-Laird & Byrne, 1991, Ch. 9; see also Ragni & Knauff, 2013). Robust facts lead individuals to abandon even a valid conclusion, to amend a premise to restore consistency, and to try to construct an explanation to resolve the inconsistency (Johnson-Laird et al., 2004). The mSentential program implementing the theory carries out the same steps, which include assembling a causal explanation in a probabilistic way from pre-existing links in its knowledge-base (ibid., p. 661) Such explanations tend to be rated as more probable than mere amendments of premises (ibid., p. 654). Standard logic does not allow defeasible inferences, and AI systems do not generate explanations. The theory accordingly predicts:

> If a salient fact rebuts a conclusion, reasoners withdraw it; amend a premise, if need be to restore consistency; and try to explain the origins of the inconsistency.

Conditionals refer to conjunctions of default possibilities The next step away from logic was to treat the models of conditionals as referring to a conjunction of possibilities (see its presage in Johnson-Laird et al., 1994, p. 735). The four cases in a partition are mutually exclusive and so their conjunction is a self-contradiction. But, the conjunction of their possibilities is not contradictory, just as: *It may rain and it may not rain* is consistent. Barrouillet and his colleagues corroborated this account. They showed that when participants list what is possible given a conditional, *If A then C*, albeit with sensible contents, they tended to list all the possibilities in the conditional's explicit models:

A	C
¬ A	¬ C
¬ A	C

and they listed the remaining case as impossible:

A	¬ C

So, the conditional refers to a *conjunction* of possibilities. And the listing above is in the order in which children acquire them. Very young children list only the first possibility, and then acquire the others in the same order that adults list them as shown above (e.g., Barrouillet et al., 2000). Analogous listings of possibilities also occur for other sorts of assertions, such as disjunctions (Hinterecker & Johnson-Laird, 2016).

One influential critique claimed that even though conditionals in the model theory refer to possibilities (Johnson-Laird & Byrne, 2002), "the extensional semantics underlying this theory is equivalent to that of the material, truth-functional conditional" (Evans et al., 2005, p. 1040). This claim is provably false. Barrouillet's results showed that people make inferences, such as:

If A then C.
Therefore, it is possible that A and C.

But, for a material conditional, the inference is invalid, as a counterexample demonstrates. Suppose that *A* is impossible.

It follows both that the material conditional is true (because its *if*-clause is false, see Table 1), and that the conjunctive conclusion is false (because *A* is impossible). The inference leads from a true premise to a false conclusion, and so it cannot be valid in truth-functional logic. As Johnson-Laird and Byrne (2002, p. 673) wrote: "Conditionals are not truth functional. Nor, in our view, are any other sentential connectives in natural language." Yet, p-theorists are incorrigible in identifying the model theory as truth-functional (Evans & Over, 2004, p. 152; Fugard et al., 2011; Pfeifer, 2013; and more).

The possibilities to which a conditional refers each hold in default of knowledge to the contrary. Such knowledge can modulate the interpretation of a conditional both by preventing the construction of a model and by adding information to a model (Johnson-Laird & Byrne, 2002). To refute a conditional, it is therefore necessary to rule out the default possibilities to which it refers, and the critical one – for reasons we explain below – is the first one, *A* & *C*. Studies have shown that people make the different predicted interpretations of conditionals that modulation can yield (e.g., Quelhas et al., 2010) and that they judge certain conditionals to be true, and others to be false, from the general knowledge they use in modulation (Quelhas et al., 2017). The general principle is therefore:

Conditionals refer to a conjunction of possibilities that each hold in default of knowledge to the contrary.

The theory extends to counterfactual conditionals, because their meanings parallel those of factual conditionals (Adams, 1975; Byrne, 2005; Byrne & Johnson-Laird, 2019). A factual conditional, such as:

If it rains today then it is hot

refers to the three possibilities and the one impossibility illustrated earlier. A *counterfactual* conditional is one in the subjunctive mood in English and in which its *if*-clause is known to be false. The following conditional is therefore counterfactual, and it parallels the preceding factual conditional:

If it had rained today (and it didn't) then it would have been hot.

It refers to a fact and to two counterfactual possibilities, which are cases that were once possible but that did not occur (Byrne, 2005; Johnson-Laird & Byrne, 2002), whereas what was impossible given the factual conditional remains impossible:

rain	hot	– a counterfactual possibility
¬ rain	¬ hot	– a fact
¬ rain	hot	– a counterfactual possibility
rain	¬ hot	– an impossibility

The definitive truth values of factual conditionals with false *if*-clauses depend on parallel counterfactuals. You claim, for example:

If it rains today then it will be hot

but we claim on the contrary:

If it rains today then it won't be hot.

In fact, it doesn't rain. In logical truth tables, the two conditionals are both true; and in partial truth tables, they are both void (see Table 1). But, in the model theory, they are each possibly true, and their definitive truth value depends on their parallel counterfactuals. Your parallel counterfactual is: *if it had rained today then it would have been hot;* our parallel counterfactual is: *if it had rained today then it wouldn't have been hot*. The definite truth of the conditional depends on which of the two counterfactuals is true (Byrne & Johnson-Laird, 2019). It can be difficult to verify counterfactuals, but it is possible in some cases, and they do not always depend on experiments (pace Shpitser & Pearl, 2007). Those who believe otherwise have never watched a game of cricket, in which umpires make judgments of whether the ball would have hit the wicket had the batsman's body not interposed itself. On appeal, the Hawkeye TV system usually confirms these judgments.

Possibilities and their presuppositions The model theory has long advocated that reasoning concerns possibilities (e.g., Johnson-Laird & Byrne, 2002). Its most recent assumption provides a primordial basis for them. What underlies possibilities and probabilities is the human ability, in almost any situation, to model a small number of exhaustive and mutually exclusive alternatives. They can each be realized in an indefinite number of different ways (Johnson-Laird & Ragni, 2019). Toss a dice, and it has six possible outcomes, but each of them as in a "possible worlds" semantics can occur in infinitely many ways, depending on the number of times the dice spins, its speed of its rotation, its mass, and so on and on. This finite foundation is common to all assertions of possibilities, but modulation from content and context can yield different interpretations (Johnson-Laird, 1978; Kratzer, 1977). Hence, the modal auxiliary verb "may" is not ambiguous, but it can be interpreted as *deontic* in giving permission, *epistemic* in expressing a possibility or a probability based on knowledge (Lassiter, 2017), or *alethic* in asserting a relation, such as one between premises and conclusion. When a fire-chief tells the inhabitants of a building after a fire: *You may go back to your apartments now*, she can be both giving them permission and asserting that it is possible to do so (Johnson-Laird & Ragni, 2019).

Presuppositions entered the model theory in its account of negation (Khemlani et al., 2014), and their key feature is that they hold for both an assertion and its denial.

The assertion:

It's possible that it's raining

presupposes:

It is possible that it is not raining.

If the latter weren't possible, then it would be certain that it is raining. Likewise, the latter assertion presupposes the former one. Indeed, people accept the inferences from one to the other (Ragni & Johnson-Laird, 2021). Modal logics deal with possibilities, and there is a countable infinity of them (Hughes & Cresswell, 1996). Yet, all normal modal logics reject the preceding inferences as invalid.

The presuppositions of possibilities have consequences for conditionals. A conditional, *if A then C*, refers to the possibility of *A*, and so it in turn presupposes the possibility of *not-A*. The conditional therefore refers to the default conjunction of:

A	C	– a possibility
¬ A	¬ C	– a presupposed possibility
¬ A	C	– a presupposed possibility
A	¬ C	– an impossibility

The presuppositions also hold for the conditional's denial:

It is not the case that if A then C.

So, it refers to the default conjunction of:

A	¬ C	– a possibility
¬ A	¬ C	– a presupposed possibility
¬ A	C	– a presupposed possibility
A	C	– an impossibility

These cases correspond to those for: *if A then not-C*, which is the most frequent denial of a conditional (Handley et al., 2006; Khemlani et al., 2014). It is now clear why the possibility of *A & C* is crucial for *if A then C*. It is possible only if the conditional is true; whereas *A & ¬ C* is possible only if it is false; and the other two possibilities can hold in either case. Table 1 above states this truth table.

A conditional can be particular, in which case only one case in its partition can hold; and a conditional can be general, in which case several cases in if partition can hold, for example:

If it is a black hole, then it has massive gravity.

Given that "it" can refer to anything, the conditional is true if at least one black hole has massive gravity (*A & C*), and there are no black holes without massive gravity (*A & ¬ C*). The other two cases in its partition can occur, and they are consistent with the truth or falsity of this conditional. These same conditions apply to the induction of scientific hypotheses (Nicod, 1950, p. 219). This account also resolves the well-known "paradox" of confirmation (Hempel, 1945). In logic, the preceding hypothesis implies its contrapositive:

If it hasn't a massive gravity then it is not a black hole.
A teddy bear doesn't have a massive gravity and it's not a black hole, and so it corroborates this latter claim, and thus in logic it also corroborates the hypothesis about black holes. But, of course teddy bears imply nothing whatsoever about the gravity of black holes. The very idea is absurd. In the model theory, however, a teddy bear is consistent with both the black-hole hypothesis, *if A then C*, and its denial, *if A then not-C*. But, as an instance of ¬ A & ¬ C, it is irrelevant to the confirmation of either the hypothesis or its negation.

Presuppositions also solve a puzzling anomaly between two sets of experimental results. When people have to make a list of the possibilities to which a conditional refers, they include those in which its *if*-clause is false (e.g., Barrouillet et al., 2000), but when they chose evidence relevant to the truth or falsity of a conditional, they do not select these cases, which they judge to be "irrelevant" (e.g., Johnson-Laird & Tagart, 1969). So, cases that are possible given a conditional are judged to be irrelevant to its truth-value. Psychologists have tried to reconcile this discrepancy (e.g., Barrouillet, Gauffroy, Lecas, 2008). But, the model theory resolves it in a general principle:

Cases in which the if-*clause of a conditional is false are possible, but they are presuppositions, which hold whether the conditional is true or false.*

Earlier, we asked why people ignore the cases in which the *if*-clause is false in Ramsey's test – a question for which he offers no answer. But, the reason is now clear: they ignore these cases because they hold whether or not the conditional is true. We now turn to the probabilities of conditionals.

The model theory of the Equation

Cases in which the *if*-clause of a conditional is false, namely, ¬ A & C and ¬ A & ¬ C, have no bearing on whether the conditional is true or false, and they have no bearing on its probability, either. Hence, its probability is embodied in this relation:

$$p(\text{if } A \text{ then } C) = \frac{p(A\&C)}{p(A\&C) + p(A\&\neg C)}$$

The right-hand side of this relation equals p(C|A), and so the model theory implies the Equation. In general, there are three main ways to infer a conditional probability. One is to use Bayes's theorem to compute the corresponding conditional probability, but individuals are unlikely to know its formula. Another is to use the ratio on the right-hand side of the equation above as advocated in the ratio procedure (Evans et al., 2003). And the third depends on the *subset* procedure (Johnson-Laird et al., 1999, p. 80). The latter is easy to overlook, but it has the singular consequence that the conditional probability can be inferred in ignorance of the probabilities of the two conjunctions above in the ratio procedure. Their occurrence on the right-hand side of the equation above simplifies to:

$$\frac{p(A\&C)}{p(A\ )}$$

Elementary arithmetic shows that this expression is equivalent to the proportion of cases of *A* that are *C*. Consider the following problem, as an example. Among a number of shapes in a box, there are three triangles, of which one is red. Someone selects a shape at random from the box. It's a triangle, so what's the probability that it's red? The subset procedure yields a conditional probability of 1/3. Yet, the statement of the problem does not even fix the probability of selecting a triangle, let alone the probabilities of the two conjunctions in the ratio: *triangle & red*, and *triangle & not red*.

The subset procedure has a corollary. It fixes only part of the JPD, and for the example above:

triangle	red	1
triangle	not-red	2

Subadditivity can occur only in one of the two remaining cases in the JPD for shapes that are not triangles. Hence, random assignments of probabilities to p(C|A), p(A), and p(C), yield one subadditive entry in the JPD on a mean of 1/2 the occasions, discounting those in which p(A) is assigned zero. In contrast, p(A & C) fixes only one case in the JPD, and so subadditivity can occur in two of the three remaining cases in the JPD. Hence, random assignments of probabilities to p(A & C), p(A) and p(C) yield subadditive JPDs on 1/2 + 1/3 = 5/6 of occasions (Khemlani et al., 2015, Appendix B). Despite their smaller likelihood of subadditivity, conditional probabilities are on the borderline of individuals' competence, whereas those of conjunctions are beyond them. A true but telling example is that two pilots disagreed about the probability that both engines on a twin-engined plane fail. The ex-fighter pilot said: double the probability of one engine failing; the glider pilot said: halve it. For conditional probabilities, one difficulty is to realize that they are required for certain problems, such as

Monty Hall's famous dilemma (e.g., Bar-Hillel & Falk, 1982; Falk, 1992; Johnson-Laird et al., 1999; Nickerson, 1996). How people estimate them depends on whether they are making an extensional or intensional estimate.

Extensional probabilities The model theory of extensional probabilities was formulated in Johnson-Laird et al. (1999), and it predicts that intuitive models should yield an erroneous estimate of the probabilities of conditionals, whereas explicit models predict normative estimates that fit the Equation. The extensional probability of an event depends on the proportion of equiprobable models from the partition in which the event occurs, or, if relevant, on the sum of their frequencies or chances (ibid., p. 68). People tend to use intuitive models and so they are susceptible to illusory inferences about extensional probabilities (ibid., p. 75 et seq.). Given a JPD, one can in principle compute any conditional probability in its domain. So, from a relevant JPD, the probability of this conditional:

If Pat has symptom *A*, then she has disease *C*

can be inferred in the three ways outlined earlier, which include the subset procedure. The intuitive models of the conditional are:

system A	disease C
. . .	

Suppose that the JPD is stated in chances out of 10, and the chance of the first of the two models above is 3. But, out of how many? There are various possibilities, and the most salient is that the chances are out of JPD as a whole, and so the probability of the conditional is 3 chances out of 10. Such estimates are wrong: they are for the probability of the conjunction that Pat has symptom A and disease C. In contrast, System 2 can construct explicit models of the two relevant possibilities, and import their chances from the JPD:

symptom A	disease C	3
symptom A	¬ disease C	2

The subset of cases of symptom A in which disease C occurs has 3 chances out of 5 (ibid., p. 79). The procedure is less ambiguous than using intuitive models, but more complicated. It calls for holding two models and their numerical chances in mind, and computing the subset. So, it should take longer than system 1's estimate, which depends on only one model. In sum, system 2's estimates should yield values of p(*if A then C*) that fit the Equation. They should take longer to make, be less variable, and have a higher probability than those from system 1.

Intensional probabilities The model theory explains how people make estimates of intensional probabilities (Khemlani et al., 2011, 2012, 2015). Consider the probability of this event:

The Democrats will lose control of the Senate in the mid-term elections.

At the time of writing, people make estimates such as: "It's quite likely," and even assign a numerical probability, such as: 60%. A major mystery is where the numbers come from in intensional estimates of the probabilities of unique events. Khemlani et al. (2015) proposed this solution: Individuals use the same machinery as for extensional estimates, but use instead models of evidence, and the numbers come from subsets in these models. This solution is implemented for systems 1 and 2 in the mReasoner program at: https://www.modeltheory.org/models/mreasoner/.

This is how the process works (ibid., p. 1227 et seq.) System 1 adduces one or two pieces of relevant evidence, represents them in intuitive models, and accumulates from these models a single coarse, granular, and non-numerical proportion. A relevant piece of evidence for the example above is:

Most mid-terms result in a loss of Senators from the administration's party.

It yields a single intuitive model of a small but arbitrary number of mid-terms in which an appropriate proportion of losses occurs (see Khemlani & Johnson-Laird, 2021, for models of quantifiers such as "most mid-terms"). As the administration is Democrat, and has a majority of only two senators, the chances that the party loses control of the Senate have a similar probability. System 1 represents the probability in an iconic line:

$$|\text{———}|$$

The left vertical represents impossibility, the right vertical represents certainty, and the line between them represents a probability as a direct analogy: the longer the line, the greater the probability (ibid., p. 1229). Similar iconic representations appear to be used by animals, infants, and adults in cultures without numbers (see, e.g., Carey, 2009). Further evidence can push the line one way or the other: it exists to accumulate the effects of separate pieces of evidence on a probability. So, the evidence that few administrations that enact popular measures lose senate seats, shrinks the line, so it is just above the mid-point between the verticals. System 1 maps the line onto an intuitive scale from *impossible*, *nearly impossible*, and so on up to *certainty*, to yield an informal description, such as: "It's likely." System 2, however, can measure the line and transform it into a numerical probability, such as: 60%.

Analogous methods apply to estimates of the probabilities of conditionals, such as:

If Biden's vaccination program fails to eliminate Covid-19, then in the mid-term elections the Democrats will lose control of the Senate.

System 1 estimates the probability of the *then*-clause, p(*in the mid-term elections the Democrats will lose control of the house*), from models of relevant evidence as we described above, and then treats the *if*-clause, *Biden's vaccination program fails to eliminate Covid-19*, as an additional piece of evidence, adjusting the line to take into account its effects. The estimate treats the *if*-clause as evidence, but neglects its probability. In general, estimates of p(A), p(C), together with an estimate of either p(if A *then* C) or p(C|A), are therefore liable to yield subadditive JPDs. Because of the way in which individuals estimate the probability of conjunctions, this tendency is greater for p(A), p(C), and p(A & C), as corroborated experimentally (Khemlani et al., 2015), and the difference between conditionals and conjunctions is therefore independent from unpacking a category (Tversky & Koehler, 1994). In contrast, system 2 can use the subset procedure, which yields a correct estimate from the numerical proportion of A's that support C in models of evidence. These estimates yield additive JPDs.

The model theory's predictions Extensional estimates of the probability of a conditional derive from at least part of its JPD or an equivalent. Intuitive models yield a conjunctive error; deliberative models are amenable to the subset procedure, and, whether extensional or intensional, should fit the Equation. They should be slower, less variable, and more accurate than intuitive estimates. And, intuitions should yield considerable subadditivity when individuals estimate the joint intensional probability of a conditional with each of the four cases in its partition. We now assess whether experimental studies support these predictions.

Experimental studies of the probabilities of conditionals

A skeptical, if not cynical, reviewer suggested that we selected only some results from the literature, and adapted the model theory retroactively to fit their results. In fact, we searched the literature using both our knowledge and Google scholar, and we have reviewed all the experiments on the Equation that we could find. Their results are fairly uniform. The model theory of extensional probabilities antedates all these studies (Johnson-Laird et al., 1999), and led to Girotto and Johnson-Laird's (2004) experiments testing the Equation. Likewise, the model theory of intensional probabilities (Khemlani et al., 2015) was developed to explain estimates of conditional

probabilities, not the probabilities of conditionals, and we did not adjust it to hark back to earlier experimental results. Instead, it inspired intensional studies of the Equation, including two that yield results contrary to p-theories (Byrne & Johnson-Laird, 2019; Goodwin & Johnson-Laird, 2018).

Probabilists who defend the Equation sometimes take for granted that individuals know how to compute conditional probabilities. In fact, their grasp is tentative at best, for example, they often infer a conditional probability solely from its converse (Johnson-Laird, 2006, p. 203). Nor do they know the Equation. So, ideal tests of its predictions call for experiments in which each participant makes numerical estimates of the probabilities of conditionals and of the analogous conditional probabilities. Our analyses include such studies, but also those that neglect one side of the Equation.

Experiments on extensional estimates One of two pioneering studies of the probabilities of conditionals is due to Evans et al. (2003). They gave participants JPDs of the frequencies of four sorts of shape in a pack of cards, and the overall frequencies were either large or small. For example, the JPD in one condition was:

1 yellow circle
4 yellow diamonds
16 red circles
16 red diamonds

The study assumed that participants knew how to compute conditional probabilities, because it called for no estimates of them. The investigators instead computed them from the JPDs they presented in the experiment. The participants' estimates of the truth of a conditional, *If A then C*, such as:

If the card is yellow then it has a circle printed on it,

and the falsity of the conditional, summed to around 100%. Likewise, as the frequency of *A & C* increased, so the participants tended to estimate a higher probability for a conditional, and as the frequency of $A \& \neg C$ increased they tended to estimate a lower probability. The participants differed one from another in whether their estimates of the probability of conditionals correlated more with the computed conditional probability or more with the probability of the corresponding conjunction in the JPD. When the probability of the *if*-clause, *A*, was lower, the participants tended to rate a conditional as less probable. The investigators' explanation of this phenomenon and of the conjunctive estimates was based on their version of Ramsey's test. It calls for individuals to use the ratio procedure, that is, to compare the probabilities of the two conjunctions: p(*A & C*) and p($A \& \neg C$). The investigators argued that those individuals who forgot to use the second

conjunction were affected by the lower probability of *A*, and tended to make conjunctive estimates of the probability of the conditional. Edgington (2003) was skeptical about the ratio procedure. Her skepticism is borne out in our earlier example of the subset procedure: among a number of shapes, there are three triangles, of which one is red. There is no need to compute the ratio – indeed, it cannot be done from this description. Individuals can use the subset procedure to estimate that given a yellow card, the probability that it has circle on it is 1 in 5. The rest of the JPD is irrelevant.

The other pioneering study is due to Oberauer and Wilhelm (2003), and they required their participants to estimate the probabilities of conditionals and the corresponding conditional

probabilities. They gave participants the frequencies of each of the four cases in the JPD for a conditional concerning letters and colors, for example:

If a card has an A then it is red.

In one condition, the JPD had these frequencies:

	red	900
A	¬red	100
¬ A	red	500
¬ A	¬red	500

Over four conditions, the frequency of *A & red* was either high (as above) or low, and the ratio of p(*A & red*) to p($A \& \neg red$), which fixes p(*red|A*), was either high (90%, as above) or medium (50%). In two such experiments, participants – adolescents and adults, respectively – estimated the conditional probability. They also estimated the probability of the conditional based on a random sample of 10 cards from the population as a whole. The reason for the random sample, as the authors explained, is that the presence of a large number of counterexamples "logically entails that the statement cannot be true for the complete deck of cards" (ibid., p. 682). The results showed that estimates for p(*if A then C*) were smaller, and had a larger standard deviation, than those for p(*C|A*). The first of these findings was also corroborated in a between-participants design (Weidenfeld et al., 2005). The individual raw data are no longer available (Oberauer, p.c.), but the paper reports that 23% of adolescents and 52% of adults made estimates that fit the Equation. As the investigators noted, some people made the conjunctive interpretation. But, they concluded: "… the probabilistic interpretation is the modal psychological meaning of *if* among logically untrained adults…" (ibid., p. 690). The slight oddity here is their earlier appeal to the use of a sample of cards, because the JPD shows that the conditional cannot be true. We agree with the authors, but

their action is contrary to the probabilistic interpretation of conditionals: truth value is more important than probability.

A subsequent extensional study with estimates of both sides of the Equation corroborated the two different sorts of estimate, one conjunctive and one matching the Equation (Oberauer, Geiger, et al., 2007). Individuals differed in which interpretation they tended to make. Estimates that fit the Equation related to the participants' judgments of truth and falsity corresponding to the partial truth table (see Table 1), whereas conjunctive estimates related to the participants' judgments corresponding to a conjunction. But, conjunctive estimates, as their third Experiment showed, were not a result of an incomplete ratio procedure in which individuals estimate only $p(A \& C)$. So, these investigators proposed a theory combining a probabilistic account with mental models to explain the conjunctive errors.

Another way to examine extensional estimates is to allow participants to infer their own numerical values from simple descriptions. Girotto and Johnson-Laird et al. (2004) used three such problems, of which one was:

> There are three cards face down on a table: 3, 6, and 8. Paolo takes one card at random, and then he takes another at random. Vittorio says:
>
> If Paolo has the 8 then he has the 3.
>
> What's the probability that Vittorio's assertion is true?

The experiment used this indirect question, because when participants thought aloud about a question such as: *What's the probability that if Paolo has the 8 then he has the 3?* they at once transformed it into a demand for a conditional probability: *If Paolo has the 8, then what's the probability that he has the 3?* Requests for the probability of a conditional tend to be ambiguous in this way (Baron, p.c., March, 2021), and it is a factor likely to enhance estimates matching the Equation. There are only three possible pairs of cards in the problem: 8 & 3, 6 & 3, and 8 & 6. Vittorio's assertion implies that only the first two of them are possible for Paolo's selections, but he cannot have chosen the remaining pair of 8 & 6. Hence, given that he has an 8, the conditional probability that has a 3 equals 1/2. On separate trials with different cards, the participants estimated the probabilities of conditionals and the corresponding conditional probabilities. The experiment investigated two other problems, and one of them was a check that the participants understood the nature of the task, because its conditional probability was equal to 1. Most participants (88%) made this estimate. For the present problem, however, more participants (50%) made a conjunctive estimate than one satisfying the Equation

(42% of participants). Over all the problems, these results were robust. Estimates for conditionals tended to be either those for conjunctions or for conditional probabilities, but most participants violated the Equation. The variation in estimates of the probabilities of conditionals was greater than the variation in estimates of conditional probabilities – a difference attributable to the conjunctive estimates.

In a further extensional investigation, Fugard and his colleagues studied estimates of the probabilities of conditionals concerning the outcomes of throws of a dice (Fugard et al., 2011). The experimenters computed the conditional probability in the Equation rather than requiring the participants to estimate it. Over the trials of the experiment, the proportion of estimated probabilities of conditionals matching the Equation increased from around 40% at the beginning, when the majority of estimates were conjunctive, to around 80% by the end of the experiment. The participants also made the conjunctive estimates faster than they made the conditional probability estimates. How the participants' estimates of conditional probabilities might have changed over the experiment is, of course, impossible to tell. And, alas, these authors take Johnson-Laird and Byrne's theory (2002) to be a fragment of classical logic in which the material conditional holds; cf. our earlier proof to the contrary.

As we showed in the previous section, the model theory made these predictions about extensional estimates of the probabilities of conditionals:

- System 1's intuitive estimates of $p(if\ A\ then\ C)$ will be close to $p(A \& C)$.
- System 2's deliberations use the subset procedure, and so they will yield the larger value, $p(C|A)$, which fits the Equation.
- Intuitive estimates will be faster but more variable than deliberative estimates.

Overall, the extensional estimates bore out these predictions, and its account of conjunctive estimates is more plausible than the alternative accounts. One surprise was the people differ reliably in whether they rely on intuitions or on deliberations. But, the final study implied that with repeated tests individuals switch from system 1 to system 2.

Experiments on intensional estimates These estimates in experiments concern unique events, for which no JPD or its equivalent exists. Likewise, estimates of the intensional probability of a conditional, *p(if A then C)*, or of a conditional probability, $p(C|A)$, are subjective, and therefore neither correct nor incorrect. But, when such an estimate is combined with estimates of $p(A)$ and $p(C)$, they can together yield subadditive JPDs in which the probabilities sum to more than 100% contrary to the probability calculus.

A pioneering intensional study examined factual and counterfactual conditionals (Over et al., 2007). These conditionals concerned causal relations, such as:

> If Adidas gets more superstars to wear their new football boots then the sales of these boots will increase.

The experimenters chose conditionals for which p(A) and p(C) were likely to have high or low probabilities, and they used all four combinations in the conditionals in their study. The participants estimated the probabilities of each of the four cases in a conditional's JPD with the proviso that the four values should sum to 100% – an important constraint that prevents subadditivity. In two experiments, the participants' estimates of the probability of conditionals correlated with conditional probabilities that the investigators computed from the values of the participants' estimated JPDs. The investigators assumed that the probability of a counterfactual conditional is the same as the probability of the parallel factual conditional for which the occurrence of the event in its *if*-clause is still an open question (Adams, 1975; see also Elqayam & Over, 2013). And they argued that their results supported Ramsey's test, and the partial truth table. They also reported a correlation between the probabilities of conditionals and those of conjunctions in the JPDs, but there was no reliable evidence for a particular group of participants making conjunctive estimates (ibid., p. 78).

Individuals have no overt knowledge of how to estimate conditional probabilities. This point was borne out in Girotto and Johnson-Laird's (2004) extensional study in the previous section. It was also corroborated in a subsequent study of intensional estimates as elicited in questions, such as:

> What is the chance that there will be a substantial decrease in terrorist activity in the next 10 years, assuming that a nuclear weapon will be used in a terrorist attack in the next decade?

The participants estimated p(A), p(C|A), p(C), and p(A|C) for 16 different sets of contents (Khemlani et al., 2015). Their first three estimates yielded a subadditive JPD on almost a quarter of the trials, and 42 of the 43 participants made one or more such estimates. As system 1 predicts, their estimates of conditional probabilities, p(C|A), tended to rely on a small adjustment to their estimates of p(C), which occurred on 63% of trials. On the remaining trials, their estimates matched the Equation. The first three estimates above allowed the fourth estimate to be computed from Bayes's theorem, but the correlation between the two was tiny and not reliable. One striking finding was that participants' estimates of p(C|A) were faster after they had estimated p(A) and p(C) than before they had made these estimates. P-theories have no ready explanation for the difference, but the model theory does: system 1 has to estimate p(C) in order to nudge its value to obtain p(C|A).

In a further study to test the model theory, participants were presented with conditionals, such as:

> If the legal drinking age in the United States is reduced, then there will be more traffic accidents.

They judged their joint truth with each case in their partitions, and they also estimated their joint probabilities with each case in their partitions (Goodwin & Johnson-Laird, 2018). They were not instructed that these probabilities should sum to 100%. The two tasks were presented all together on the same webpage. The participants overall estimates of the joint probabilities for various sorts of conditional conjoined with possibilities in their partitions were as follows:

If A then C, and A and C:	joint probability = 81%
If A then C, and not-A & not-C:	joint probability = 60%
If A then C, and not-A & C:	joint probability = 51%

These results show a striking degree of subadditivity. The model theory predicts both this result and the reliable trend over the three probabilities in terms of the availability of models of conditionals (see, e.g., Barrouillet et al., 2000). The participants also tended to judge that each pair could be jointly true, which is a direct rebuttal of the partial truth table but which corroborates the model theory's truth table (see Table 1).

A series of experiments examined pairs of everyday factual and counterfactual conditionals (Byrne & Johnson-Laird, 2019), which the model theory implies have meanings running in parallel with one another. The participants estimated the joint probabilities of a conditional with each of the four cases in its partition. Both factual and counterfactual conditionals yielded similar probabilities. For example, in one experiment the mean estimates of the joint probability of *If A then C* with each case in its partition were as follows, for factual and counterfactual conditionals, respectively:

A	C:	80% and 78% probabilities
not-A	not-C:	64% and 64% probabilities
not-A	C:	35% and 35% probabilities
A	not-C:	28% and 27% probabilities
	Sums of JPD:	207% and 204% probabilities

When participants in another experiment were told that their estimates should sum to 100%, not surprisingly subadditivity disappeared, but the same trend over the four cases still occurred. The results corroborated intensional estimates of conditionals as based on nudges of intuitive estimates

of p(C), a procedure that is liable to be subadditive. Because participants had to use such estimates four times on each trial, the subadditivity was massive.

In a study of reasoning from conditionals, Singmann et al. (2014) also asked their participants to make estimates of the probabilities of various assertions relating to everyday conditionals, such as:

If Greece leaves the Euro then Italy will too.

As our analysis of the raw data shows, the participants' estimates of the probabilities of conditionals over three trials were closer to their estimates of conditional probabilities on a mean of 1.9 trials, and closer to their estimates of conjunctions on a mean of 1.1 trials (both sorts of estimate occurred reliably, Wilcoxon tests, $z > 4.76$ and 4.1, $ps < .0001$).

Theorists supporting the "inferentialist" view, which we described earlier, have argued that the Equation holds only in case A implies a greater likelihood of C than $\neg A$ does, i.e., $p(C|A)$ is greater than $p(C|\neg A)$, and so this *probabilistic relevance* has a positive value (Skovgaard-Olsen et al., 2016; see also Krzyżanowska et al., 2017). When other investigators manipulated probabilistic relevance numerically, varying the frequencies of the cases in the partition, it had no reliable effects on estimates of the extensional probabilities of conditionals (Oberauer, Weidenfeld, & Fischer, 2007). But, Skovgaard-Olsen and his colleagues manipulated it using, not numerical means, but scenarios that should affect probabilistic relevance. The probability of conditionals with a positive relevance tended to match the Equation, those with a negative relevance elicited reliably lower matches, and those with zero relevance had the lowest match of all. The participants did not have to calculate the values of probabilistic relevance, and no evidence showed that they did. Perhaps the scenario itself had a direct effect on the credibility of the conditionals.

As we showed earlier, the model theory made these predictions about intensional estimates of the probabilities of conditionals:
- Intuitive estimates nudge estimates of p(C), which can yield a subadditive JPD.
- Deliberative estimates should tend to equal p(C|A).
- The four joint estimates of the probability of a conditional with each case in its JPD should yield considerable subadditivity.

The experimental evidence corroborates this account, and the judgements that conditionals can be jointly true with each of the cases in their partitions rule out the partial truth table, but corroborate the model theory's truth table (see Table 1). The most surprising result was that the vast subadditivity of joint probabilities of conditionals with each case in their partitions.

General discussion

Our goal is to reach a conclusion about how people estimate the probabilities of conditionals. To get there, however, we first assess the various assumptions underlying probabilistic theories (p-theories). Some of their proponents have adopted Ramsey's test (e.g., Evans & Over, 2004; Oaksford & Chater, 2007, p. 109; Oberauer & Wilhelm, 2003). Some have adopted in addition de Finetti's partial truth table (e.g., Baratgin et al., 2013; Fugard et al., 2011). Some have adopted Adams's probabilistic logic (e.g., Evans, 2012; Oaksford & Chater, 2007), though others reject it (e.g., Douven & Verbrugge, 2010). And some have incorporated an AI nonmonotonic system P, which is not probabilistic (Kraus et al., 1990), into their p-theories (e.g., Gilio, 2002; Pfeifer & Kleiter, 2003).

These probabilistic theories have had the great merit of inspiring research, and they have led to suggested rapprochements with the model theory (e.g., Oaksford & Chater, 2010; Oberauer, Geiger, et al., 2007). Their early developments, such as an application to the selection task – in which participants select evidence pertinent to testing general conditional hypotheses (Oaksford & Chater, 1996) – seemed very promising. But, this development, unlike the model theory, failed to predict some crucial phenomena (see Ragni et al., 2018).

P-theories and the model theory do agree on some fundamental points:
- Much reasoning in daily life is from uncertain premises – probabilities (p-theories), or possibilities (the model theory).
- Subjective probabilities correspond to degrees of belief.
- The Equation is normative: individuals ought to estimate the probability of a conditional as equal to the corresponding conditional probability.

The two theoretical approaches also agree that the percentages of fits to the Equation in experiments are too small to support the hypothesis that the everyday interpretations of conditionals are always conditional probabilities. If less than half the participants in a study make an interpretation, but it increases during the course of the experiment (Fugard et al., 2011), the participants are acquiring the interpretation. The meaning of *if*, in contrast, is what they bring to the experiment.

The two approaches have different foundations. The model theory is founded on possibilities that hold in default of knowledge to the contrary. P-theories are founded on numerical probabilities and the probability calculus. As a result, p-theories have several quite profound problems, which we now enumerate.

1. **P-theories and the prediction of errors.** A common extensional error is conjunctive: individuals estimate p(A & C) instead of p(*if A then C*). Among p-theorists, only two accounts of the error exist. In one, Evans et al. (2003)

postulated a version of Ramsey's test that yields the probability of a conditional from a ratio based on p(A & C) and p(A & ¬ C), and they further proposed that some individuals neglect the second conjunction, so, as a result, they make the conjunctive error. But, as Edgington (2003) remarked, this *ratio* hypothesis has neither a theoretical nor an empirical motivation – beyond, perhaps, harking back to the conjunctive error. In another account, Oberauer and Wilhelm (2003) imported considerations from the model theory to explain the error. In the current model theory, however, conditionals refer to conjunctions of three possibilities that each hold in default of knowledge to the contrary. Only one of these possibilities is unique to a true conditional (see Table 1), and the other two are presuppositions that are also possible if the conditional is false. The model of this unique possibility yields the conjunctive error.

2. **P-theories are not algorithmic.** The various p-theories specify that people compute probabilities, but none of them proposes an algorithm for the underlying mental processes. For example, no algorithm has ever been implemented for a complete Ramsey's test (pace Pearl, 2013). The algorithm should take as input an everyday conditional, determine whether its *if*-clause is true or false according to a body of knowledge, and if it is true, estimate the probability of the conditional according to the degree to which its *then*-clause holds in the same body of knowledge. In case the *if*-clause is false, then all bets are off: the test is void (see Ramsey's footnote quoted earlier). To decide whether the proposition in the *if*-clause is at least consistent with what they know, people have to compare it with their knowledge. But, the process of assessing consistency is computationally intractable: it is NP-complete (Cook, 1971). Hence, with an increasing number of propositions in the body of knowledge that are relevant to an *if*-clause, the test soon defeats any finite system, such as the human brain. If p-theorists had formulated an algorithm for the test, it would have forced them to constrain the number of beliefs that it consults.

3. **P-theories do not apply to deontic conditionals.** Numerical probabilities, unlike possibilities, cannot underlie everyday speech acts that create conditional deontic states (see Austin, 1975). For instance, a museum attendant creates an obligation for you:
   If you have a ticket for noon, then you must enter now.
   The same obligation cannot be created in an assertion referring, not to deontic possibilities, but to probabilities. A reviewer suggested that "must" could have a probabilistic interpretation, but the museum attendant fails to create an obligation, or a speech act of any sort, if she asserts instead its probabilistic paraphrase:

If you have a ticket for noon, then you have a probability of 100% of entering now.
Obligation does not guarantee certainty (Johnson-Laird & Ragni, 2019): people often fail to meet their obligations. Likewise, probabilities cannot paraphrase assertions that create conditional permissions.

4. **P-theories predict inferences that people reject.** Here is one crucial case in which the model theory predicts the rejection of an inference that p-theories allow, but there are others. The partial truth table in Table 1 predicts that conditionals that have a truth-value must have true *if*-clauses (see Table 1). Most people judge that this conditional, for example:

If Mary has the flu then she is ill

is true without the need for any evidence (Quelhas et al., 2017). They know that flu is an illness. So, according to the partial truth table they should infer that the conditional implies:

Mary has the flu.

The partial truth table has this consequence, because it specifies that the only way in which a conditional can have a truth-value is in case its *if*-clause is true. Conversely, a conditional with a truth-value, whether it is true or false, must have a true *if*-clause. The partial truth table is therefore mistaken. And judgments of joint truth of a conditional and cases in which its *if*-clause is false are also contrary to the partial truth table but corroborate the model theory's truth table (see Table 1).

5. **P-theories fail to predict inferences that people make.** Here are two examples:
   If the store sells shoes or sandals, then it follows that it sells shoes.
   and:
   Few customers had sole or lobster.
   Therefore, few customers had sole.
   The model theory's conjunctive analysis of disjunctions predicts the first sort of inference, and people accept it. The theory's analysis of quantifiers allows certain of them to refer to their own proper subsets (Johnson-Laird et al., 2021; Khemlani & Johnson-Laird, 2021). Yet, both these inferences violate p-validity and validity in standard logic.

6. **P-theories fail to predict subadditive estimates of probabilities.** Perhaps the most striking phenomenon that the model theory predicts contrary to p-theories is the subadditivity of certain joint probability distributions

(JPDs). The process of estimating p(C) and then nudging it one way or the other to estimate a unique p(C|A) can lead with estimates of p(A) and p(C) to a predictable subadditivity of the JPD (Khemlani et al., 2015). When such estimates occur four times in a row to assess the joint probabilities of a conditional with each case in its partition, the result is a massive subadditivity (Byrne & Johnson-Laird, 2019; Goodwin & Johnson-Laird, 2018). A tell-tale sign of the model theory is accordingly the occurrence of subadditive judgments of probability in the absence of either heuristic cues (Tversky & Kahneman, 1983) or the unpacking of categories (Tversky & Koehler, 1994).

7. **P-theories and defeasible inferences.** Defeasible reasoning appears to be ubiquitous in daily life, and at least some p-theories, as we mentioned, have sought to allow it by adopting system P (Kraus et al., 1990). However, defeasible reasoning in daily life differs in character from system P: human reasoners tend to explain the origins of an inconsistency rather than just to withdraw tentative conclusions. And they judge such explanations as more probable than minimal edits to premises (Johnson-Laird et al., 2004; Khemlani & Johnson-Laird, 2012).

8. **P-theories and the automatic elicitation of probabilities.** In p-theories, conditionals automatically elicit probabilities. In the model theory, probabilities have no overt consequences unless the contents of a task or its instructions elicit them, and its participants have at least a rudimentary grasp of them. So, p-theories imply that no essential difference should occur between these two conditionals:

If a vaccine is approved then it is safe for human use and:
If a vaccine is approved then it is probably safe for human use.

But, it follows from the model theory that they should elicit different inferences. Goodwin (2014) corroborated the model theory in a series of nine experiments: individuals distinguished between the inferential consequences of the two sorts of conditional. A pertinent finding also occurred in a study of the strategies that individuals develop in order to make simple inferences from premises including conditionals (van der Henst et al., 2002). They thought aloud as they inferred conclusions, and their protocols and the diagrams they drew showed that they developed various strategies focused on the possibilities to which the premises refer. About one in ten of their conclusions used modal terms, such as *may*, *might*, *can*, *could*, and *possibly* (Ibid., p. 548). They drew no conclusions referring to probabilities. This failure is contrary to the automatic elicitation of probabilities.

Table 2 summarizes the preceding analyses.

Proponents of the new paradigm have criticized the model theory on several points (see Baratgin et al., 2015;

Oaksford et al., 2019), and its defenders have replied to them (Johnson-Laird et al., 2015b; Hinterecker, Knauff, & Johnson-Laird, 2019). The critics treat the subadditivity of estimates of probabilities as peripheral; nonetheless they are anomalies that p-theories cannot explain. The critics call for a normative theory of human reasoning that is complete and decidable. A formal calculus can be complete, that is, any valid inference in the semantics of the calculus is provable using the proof theory of the calculus, and decidable, which means that a finite number of steps determines whether or not it is valid (see, e.g., Jeffrey, 1981, p. 185). The model theory has only a semantic basis, and so the question of its completeness is inapplicable. System 2, however, is normative, and it yields a decision about the status of a putative inference, bearing in mind that even sentential reasoning soon becomes computationally intractable with an increase in the number of clauses in premises (Cook, 1971).

When a theory survives crucial comparisons with its rivals, defenders of the latter theories seldom abandon them. Instead, they try to devise experiments that refute the theory. So far, they have not presented evidence of this sort against the model theory. So, their criticisms concern, not what the theory predicts, but whether it has certain desirable properties. By far, the most important of these properties is that the theory is open to an empirical test that could falsify it. The model theory's three basic predictions, which we outlined earlier, are testable in this way:

• Inferences dependent on multiple models should be harder than those dependent on a single model. To determine whether an inference depends on multiple models or a single model, the decision rests on the program implementing the theory, but it is obvious in many simple cases, for example, conjunctions have one model, and conditionals have more than one model.

• Inferences dependent on what is false should yield systematic fallacies ("illusory inferences") in comparison with those that do not. Again, the program predicts them: system 1 accepts conclusions that system 2 rejects.

• Errors arise from the neglect of possibilities – from catastrophes in daily life to erroneous conclusions in experiments, which, apart from illusory inferences, tend to be consistent with models of the premises but not to follow from them.

The underlying principles and their computer implementation made correct predictions about spatial and syllogistic reasoning, and later about reasoning about conditionals and other compounds (see Johnson-Laird & Byrne, 1991).

The auxiliary principles of the model theory concern semantics, such as that the meaning of a conditional refers to a conjunction of default possibilities, in which those corresponding to the falsity of the *if*-clause are presuppositions. Robust results contrary to the resulting truth table and its

Table 2 The main bases for probabilistic theories, and whether (+) or not (-) they and the model theory account correctly for phenomena concerning conditionals. All theoretical bases in the table accept the Equation that p(if A then C) = p(C|A) as a norm

	1. Predicts conjunctive errors in estimates of the Equation	2. Has an algorithm simulating mental processes	3. Applies to creation of permissions and obligations	4. Free of the 'paradox' that truth of *if* A then C implies A	5. Predicts inferences of the sort: *If A or B then it follows that A*	6. Predicts subadditive JPDs	7. Predicts explanations in defeasible reasoning	8. Has no automatic elicitation of probability
Ramsey's test	+*	-	-	+	-	-	-	+
Partial truth table	-	-	-	-	-	-	-	+
Jeffrey table for probabilities	-	-	-	+	-	-	-	-
Adams' probabilistic logic	-	-	-	-	-	-	-	-
de Finetti's coherence and system P	-	-	-	-	-	-	+	-
Model theory of default possibilities	+	+	+	+	+	+	+	+

* Evans et al. (2003) and Oberauer and Wilhelm (2003) offer explanations of conjunctive errors in estimates of the Equation (see text)

predictions about the probabilities of conditionals would show that this aspect of the theory is wrong.

The chief weakness of the theory is, in fact, that it does not predict which individuals rely only on the intuitive models of system 1 and which individuals also use the explicit models of system 2. We were therefore forced to use a free parameter for this decision in fitting the theory to data (e.g., Khemlani et al., 2018). However, people of greater ability or training should be more likely to use system 2 (see Stanovich, 1999), and thus to estimate the probability of a conditional as a conditional probability rather than the probability of a conjunction (see Evans et al., 2007). Likewise, the easier a problem – in terms of its demands on the processing capacity of working memory – the more likely individuals should be to engage system 2. For instance, it is easy to enumerate the possibilities to which a conditional refers, and adults typically list the three default possibilities of the model theory (e.g., Barrouillet et al., 2000). But, it is difficult to estimate the extensional probability of a conditional, and a reliable proportion of participants rely only on intuitive models, and so make erroneous conjunctive estimates. P-theories aim to characterize human reasoning and some of their proponents refer to the approach as the "new paradigm." The accumulation of anomalies such as those in Table 2 suggests at the very least a need for a *new* new paradigm.

Conclusions

People can have good lives without ever estimating a numerical probability. Yet, most of them can assess intuitive probabilities as in, "it's highly likely." This intuitive thinking (in system 1) underlies even numerate individuals' ability to

estimate the numerical probability of a conditional (in system 2). They can infer the probability of:

If the next throw of this fair dice yields an even number then it will be a six.

They enumerate the subset of cases in which such a throw results in a six. This subset yields an estimate that fits the Equation in which the probability of a conditional equals the corresponding conditional probability. Other individuals may focus on the model of the one possibility that holds only if the conditional is true: the fair dice yields an even number and it is six. This conjunctive estimate is wrong. Deliberation takes longer than intuition, but it yields more accurate and less varied estimates. Individuals can also estimate the probability of a conditional about a unique event, such as:

If the Prime Minister loses the next election, then he will resign the party's leadership.

They estimate the probability that the Prime Minister's resignation from the proportion of models of relevant evidence supporting this event. They then nudge this probability a little, one way or the other, depending on the effect of the loss of the election, which they treat as though it were another piece of evidence. Able individuals, however, can deliberate in order to use the subset procedure.

The probability calculus supplements human intelligence rather than underlies its deliberations. That may be a factor in why it took so long to formulate the calculus as the epigraph this paper makes clear. But, the human ability to manipulate models of the world provides a basis for the development of logic, the probability calculus, and their numerous ramifications.

References

Adams, E.W. (1975). *The logic of conditionals*. Reidel.

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

Austin, J. L. (1975). *How to do things with words*. 2nd Ed. Harvard University Press.

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

Baratgin, J., Over, D. E., & Politzer, G. (2013). Uncertainty and the de Finetti tables. *Thinking & Reasoning*, *19*, 308-328.

Bar-Hillel, M. A., & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition*, *11*, 109-122.

Baron, J. (2008) *Thinking and Deciding*. Fourth Ed. Cambridge University Press.

Barrouillet, P., Gauffroy, C., & Lecas, J. F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review*, *115*, 760–771.

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

Bauer, M. I., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science*, *4*, 372-378.

Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford University Press.

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

Byrne, R. M. J., & Johnson-Laird, P. N. (2019). *If* and *or*: real and counterfactual possibilities in their truth and probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Coletti, G., & Scozzafava, R. (2002). *Probabilistic logic in a coherent setting*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

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

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions of the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.

Craik, K. (1943). *The Nature of Explanation*. Cambridge University Press.

de Finetti, B. (1992). Foresight. In Kotz, S. et al. (eds.), *Breakthroughs in Statistics*. Springer, pp. 134-174. (Originally published, 1936.)

de Finetti, B. (1995). The Logic of Probability. *Philosophical Studies*, *77*, 181-190. (Originally published, 1935.)

Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002). The relative contribution of content and context factors on the interpretation of conditionals. *Experimental Psychology*, *49*, 181–195.

Douven, I. (2015). How to account for the oddness of missing-link conditionals. *Synthese*, 1-14.

Douven, I., Elqayam, S., Singmann, H., Wijnbergen-Huitink, J. (2018). Conditional and inferential connections: A hypothetical inferential theory. *Cognitive Psychology, 101*, 50-81.

Douven, I. & Verbrugge, S. (2010). The Adams family. *Cognition, 117*, 302-318.

Edgington, D. (1995). On conditionals. *Mind, 104*, 235-329.

Edgington, D. (2003). What if? Questions about conditionals. *Mind and Language*, *18*, 380–401.

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. *Thinking & Reasoning*, *19*, 249–265.

Evans, J. St. B. T. (1977). Toward a statistical theory of reasoning. *Quarterly Journal of Experimental Psychology*, *29*, 621-635.

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

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

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition*, *35*, 1772–1784.

Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 321-335.

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

Evans, J.ST. B.T. , Over, D. E., & Handley, S. J. (2005). Supposition, extensionality and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, *112*, 1040–1052.

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition*, *43*, 197-223.

Fugard, A. J. B., Pfeifer, N., Mayerhofer, B., & Kleiter, G. D. (2011). How people interpret conditionals: Shifts toward the conditional event. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 635-648.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is relevant for psychology and vice versa. In Wright, G., & Ayton, P. (Eds.), *Subjective probability* (pp. 129–162). Wiley.

Gilio, A. (2002). Probabilistic reasoning under coherence in System P. *Annals of Mathematics and Artificial Intelligence*, 34, 5–34.

Girotto, V. & Johnson-Laird, P. N. (2004). The probability of conditionals. *Psychologica, 47*, 207-225.

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

Goodwin, G.P., & Johnson-Laird, P. N. (2018). The truth of conditional assertions. *Cognitive Science*, *42*, 2502–2533.

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

Hacking, I. (1975). *The emergence of probability*. Cambridge University Press.

Handley, S. J., Evans, J. St. B. T., & Thompson, V. A. (2006). The negated conditional: A litmus test for the suppositional conditional? *Journal of Experimental Psychology: Learning, Memory, Cognition*, *32*, 559–569.

Hempel, C. G. (1945). Studies in the logic of confirmation, Parts I & II. *Mind*, 54, 1–26 & 97–121.

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

Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2019). How to infer possibilities: A reply to Oaksford et al. (2018). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 298-301.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach*. (2nd Ed.) La Salle.

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

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

Jeffrey, R. C. (1991). Matter of fact conditionals. *Proceedings of the Aristotelian Society, Supplementary Volumes, 65*, 161183.

Johnson-Laird, P. N. (1970). The perception and memory of sentences. In Lyons, J. (Ed.) *New Horizons in Linguistics*. (Pp. 261-270). Penguin.

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

Johnson-Laird, P. N. (1978). The meaning of modality. *Cognitive Science, 2*, 17-26.

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

Johnson-Laird, P. N. (2006). *How we reason*. Oxford University Press.

Johnson-Laird, P. N. (2012). Mental models and consistency. In Gawronski, B., & Strack, F. (Eds.) *Cognitive Consistency: A Fundamental Principle in Social Cognition*. Guilford Press. (Pp. 225-244).

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

Johnson-Laird, P. N. & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 646-678.

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

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1994). Why models rather than rules give a better account of propositional reasoning: a reply to Bonatti, and to O'Brien, Braine, and Yang. *Psychological Review, 101*, 734-739.

Johnson-Laird, P.N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*, 640-661.

Johnson-Laird, P.N., Khemlani, S.S., & Goodwin, G.P. (2015a). Logic, probability, and human reasoning. *Trends in Cognitive Sciences, 19*, 201-214.

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015b). Response to Baratgin et al.: Mental models integrate probability and deduction. *Trends in Cognitive Sciences, 19*, 548-549.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science, 288*, 531-532.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106*, 62–88.

Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition, 193*, 130950.

Johnson-Laird, P. N., Rasga, C., & Quelhas, A. C. (2021). Mental models, free choice, and inferences about disjunctions. Under submission.

Johnson-Laird, P. N., & Savary, F. (1996). Illusory inferences about probabilities. *Acta Psychologica, 93*, 69-90.

Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition, 71*, 191-229.

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

Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1, 134-148.

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

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science, 4*, 533–550.

Khemlani, S., Byrne, R. M. J., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive Science, 42*, 1887-1924.

Khemlani, S., & Johnson-Laird, P. N. (2012). Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta Psychologica, 139*, 486-491.

Khemlani, S., & Johnson-Laird, P. N. (2021). Reasoning about properties: A computational theory. (Under submission).

Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2011). What makes intensional estimates of probabilities inconsistent? In Carlson, L., Hölscher, C., & T. Shipley, T. (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. (pp. 1906-1911). Cognitive Science Society.

Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2012). The probability of unique events. *PLOS-ONE, 7*, 1-9.

Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2015). Naïve probability: Model-based estimates of unique events. *Cognitive Science, 39*, 1216-1258.

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2014). The negations of conjunctions, conditionals, and disjunctions. *Acta Psychologica, 151*, 1-7.

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

Knauff, M., & Castañeda, L. E. G. (2021). When nomenclature matters: Is the 'new paradigm' really a new paradigm for the psychology of reasoning? Under submission.

Kolmogorov, A. (1950). *Foundations of the theory of probability*. : Chelsea.

Kratzer, A. (1977). What 'must' and 'can' must and can mean. *Linguistics and Philosophy, 1*, 337–355.

Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence, 44*, 167–207.

Krzyżanowska, K., Collins, P. J., Hahn, U. (2017). Between a conditional's antecedent and its consequent. Discourse coherence vs. probabilistic relevance. *Cognition, 164*, 199-205.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press. Lassiter, D. (2017). *Graded modality: Qualitative and quantitative perspectives*. Oxford University Press.

Lassiter, D. (2017). *Graded modality: Qualitative and quantitative perspectives*. New York: Oxford University Press.

Lewis, D. (1973). *Counterfactuals*. Basil Blackwell.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review, 85*, 297-315.

Manktelow, K. (2021). *Beyond reasoning: The life, times and work of Peter Wason, pioneering psychologist*. Routledge.

Nickerson, R. S. (1996). Ambiguities and unstated assumptions in probabilistic reasoning. *Psychological Bulletin, 120*, 410-433.

Nickerson, R. S. (2015). *Conditional reasoning: The syntactics, semantics, thematics, and pragmatics of 'If'*. : Oxford University Press.

Nicod, J. (1950). *Foundations of geometry and induction*. : The Humanities Press.

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

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

Oaksford, M., & Chater, N. (2010). Conditional inference and constraint satisfaction: Reconciling mental models and the probabilistic approach. In Oaksford, M., & Chater, N. (Eds.), *Cognition and*

conditionals: *Probability and logic in human thinking* (pp. 309–333). : Oxford University Press.

Oaksford, M., Over, D., & Cruz, N. (2019). Paradigms, possibilities, and probabilities: Comment on Hinterecker, Knauff, and Johnson-Laird (2016). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 288–297.

Oberauer, K., Geiger, S. M., Fischer, K., & Weidenfeld, A. (2007). Two meanings of 'if'? Individual differences in the interpretation of conditionals. *The Quarterly Journal of Experimental Psychology, 60*, 790-819.

Oberauer, K. & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 680-693.

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

Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology, 54*, 62-97.

Partee, B.H. (1979). Semantics – mathematics or psychology? In Bäuerle, R., Egli, U., & von Stechow, A. (Eds.) *Semantics from different points of view*. (Pp. 1-14) : Springer-Verlag.

Pearl, J. (2013). Structural counterfactuals: A brief introduction. *Cognitive Science, 37*, 977–985.

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

Pfeifer, N. & Kleiter, G. D. (2003). Nonmonotonicity and human probabilistic reasoning. *Proceedings of the 6th workshop on uncertainty processing*, 221-234.

Pfeifer, N. & Kleiter, G. D. (2007). Human reasoning with imprecise probabilities: Modus ponens and Denying the antecedent. *5th International Symposium on Imprecise Probability: Theories and Applications*. , Czech Republic.

Politzer, G., Over, D. E., & Baratgin, J. (2010). Betting on conditionals. *Thinking & Reasoning, 16*, 172-197.

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

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

Ragni, M., & Johnson-Laird, P. N. (2021). Reasoning about epistemic possibilities. *Acta Psychologica, 208*, 103081.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review, 120*, 561-588.

Ragni, M., Kola, I., & Johnson-Laird, P. N. (2018). On selecting evidence to test hypotheses. *Psychological Bulletin, 144*, 779-796.

Ramsey, F. P. (1990a). Truth and probability. In Mellor, D. H. (Ed.) *F. P. Ramsey: Philosophical Papers* (pp. 52–94). Cambridge, UK: Cambridge University Press. (Originally published in 1926.)

Ramsey, F. P. (1990b). General propositions and causality. In Mellor, D. H. (Ed.) *F. P. Ramsey: Philosophical Papers* (Pp. 145-163). Cambridge, UK: Cambridge University Press. (Originally published in 1929.)

Ramsey, F. P. (1990c). Mathematical logic. In Mellor, D. H. (Ed.) *F. P. Ramsey: Philosophical Papers* (pp. 225-244). Cambridge, UK: Cambridge University Press. (Originally published in 1926.)

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

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

Rips, L. J., & Marcus, S. L. (1977). Suppositions and the analysis of conditional sentences. In Just, M. A., & Carpenter, P. A. (Eds.) *Cognitive processes in comprehension*. (Pp. 185-220.) New York: Wiley.

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

Sanfilippo, G., Gilio, A., Over, D., & Pfeifer, N. (2020). Probabilities of conditionals and previsions of iterated conditionals, *International Journal of Approximate Reasoning, 121*, 150173.

Shpitser, I., & Pearl, J. (2007). What counterfactuals can be tested. *Proceedings of 23rd Conference on uncertainty in artificial intelligence*, 352-359. Vancouver, .

Singmann, H., Klauer, K. C., & Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology, 5*, Article 316.

Skovgaard-Olsen, N., Collins, P., Krzyzanowska, K., Hahn, U., Klauer, K. C. (2019). Cancellation, negation, and rejection. *Cognitive Psychology, 108*, 42-71.

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition, 150*, 26-36.

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

Stalnaker, R. (1968). A theory of conditionals. *American Philosophical Quarterly Monograph Series, 2*, 98-112

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

Strasser, C., & Antonelli, G.A. (2019). Non-monotonic logic. In Zalta, E. N. (Ed.) *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/sum2019/entries/logic-nonmonotonic/>.

Tversky, A., & Kahneman, D. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 292–315.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*, 547–567.

van der Henst, J-B., Yang, Y., & Johnson-Laird, P. N. (2002). Strategies in sentential reasoning. *Cognitive Science, 26*, 425-468.

Venn, J. (1888). *The logic of chance*. (3rd Ed.) : Macmillan.

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

Weidenfeld, A., Oberauer, K., & Hörning, R. (2005). Causal and non-causal conditionals: An integrated model of interpretation and reasoning. *Quarterly Journal of Experimental Psychology, 58A*, 1479-1513.