Journal of Experimental Psychology: Learning, Memory, and Cognition

Modality, Probability, and Mental Models

Thomas Hinterecker, Markus Knauff, and P. N. Johnson-Laird

Online First Publication, February 29, 2016. http://dx.doi.org/10.1037/xlm0000255

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000255

Modality, Probability, and Mental Models

Thomas Hinterecker
Max Planck Institute for Biological Cybernetics, Tübingen, Germany, and Justus Liebig University

Markus Knauff
Justus Liebig University

P. N. Johnson-Laird
Princeton University and New York University

We report 3 experiments investigating novel sorts of inference, such as: A or B or both. Therefore, possibly (A and B). Where the contents were sensible assertions, for example, *Space tourism will achieve widespread popularity in the next 50 years or advances in material science will lead to the development of antigravity materials in the next 50 years, or both*. Most participants accepted the inferences as valid, though they are invalid in modal logic and in probabilistic logic too. But, the theory of mental models predicts that individuals should accept them. In contrast, inferences of this sort—A or B but not both. Therefore, A or B or both—are both logically valid and probabilistically valid. Yet, as the model theory also predicts, most reasoners rejected them. The participants' estimates of probabilities showed that their inferences tended not to be based on probabilistic validity, but that they did rate acceptable conclusions as more probable than unacceptable conclusions. We discuss the implications of the results for current theories of reasoning.

Keywords: modal logic, probabilistic logic, mental models

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

So, is it possible that there's both an ace and a king in the hand?

The inference seems valid, and it is a typical example of *modal* reasoning, that is, reasoning about what is possible or necessary. Psychologists have studied the modal reasoning of children (e.g., Byrnes & Beilin, 1991; Inhelder & Piaget, 1958; Piéraut-Le Bonniec, 1980; Sophian & Somerville, 1988) and of adults (e.g., Bell & Johnson-Laird, 1998; Goldvarg & Johnson-Laird, 2000; Osherson, 1976). But, the focus in psychology has been on deontic reasoning—what is permissible and impermissible—rather than on what is possible and impossible. As a result, no comprehensive theory of human reasoning about modalities exists. One obvious limitation is the dearth of studies; another is the ambiguity of modal terms. Consider, for example, the claim:

It is possible that there are unicorns on Mars.

One interpretation of "possible" concerns what is logically possible (the alethic modality), and, logically, unicorns could exist on Mars. But, another interpretation is epistemic, and from all we know, unicorns could not exist on Mars. Our present concern is everyday assertions that take knowledge into account.

A valid deduction, according to Jeffrey (1981, p. 1), is one whose conclusion is true in every case in which all of its premises are true, that is, if its premises are true then its conclusion is true too. So, consider again our opening inference about the hand of cards. Does the conclusion follow validly from the premise? As we will show, most people think so. It seems obvious. Yet, the inference isn't valid in modal logic, and so its acceptability stands in need of explanation. The main goal of the present paper is therefore to explain modal inferences of this sort.

Three potential explanations for modal reasoning exist. First, reasoners may rely on some sort of modal logic (Osherson, 1976). Second, they may rely, not on logic, but on the balance of probabilities: The conclusion is at least as probable as the premise. We refer to this approach as "p-logic," which is short for probabilistic logic. Its pioneer was Adams (1975), and we focus on his approach, but it does have various versions (e.g., Oaksford & Chater, 2007; Pfeifer & Kleiter, 2005). Third, reasoners may rely neither on logic nor p-logic, but on mental models, that is, iconic representations of possibilities (e.g., Johnson-Laird, 2006; Johnson-

Thomas Hinterecker, Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, and Experimental Psychology and Cognitive Science, Justus Liebig University; Markus Knauff, Experimental Psychology and Cognitive Science, Justus Liebig University; P. N. Johnson-Laird, Emeritus Professor, Psychology, Princeton University, and Visiting Scholar, Psychology, New York University.

The research was carried out during the first author's visit to the Department of Psychology, New York University, in the Spring of 2015, and we thank the Chair, Peter Gollwitzer, and Professor Greg Murphy for their help, and for providing a stimulating intellectual atmosphere and facilities to carry out research. We thank Gernot Kleiter and David Over for helpful discussions of probabilistic validity, Dan Osherson for helping us to show how modal logic proves our opening inference, Mike Oaksford for a magnanimous review, and three anonymous reviewers for their helpful criticisms. The research in this article was partially supported within the Priority Program "New Frameworks of Rationality" and by Grant KN465/6–2 and KN465/11–1 from the Deutsche Forschungsgemeinschaft (DFG) to Markus Knauff.

Correspondence concerning this article should be addressed to Thomas Hinterecker, Max Planck Institute for Biological Cybernetics, Human Perception, Cognition and Action, Spemannstr. 38, D-72076 Tübingen, Germany. E-mail: thomas.hinterecker@tuebingen.mpg.de

Laird, Khemlani, & Goodwin, 2015). The present article aims to sort out these potential explanations. Its plan is straightforward. It begins with a logical analysis of the inferences under investigation, and with a psychological theory based on modal logic. It outlines p-logic and then the theory of mental models. Next, it describes three experiments designed to test the contrasting predictions of the three approaches. Finally, it discusses the implications of the experimental results.

Modal Logic

Logicians from Aristotle onward have studied modal deductions about possibility and necessity (see, e.g., Kneale & Kneale, 1962). Early in the 20th century, logicians axiomatized various modal logics, of which there are many (Hughes & Cresswell, 1996). Some are based on the sentential calculus, which concerns negation and idealized sentential connectives, such as *and*, *if* and *or*, to which logicians add the modal operators of *possibility* and *necessit*y, and for which they provided a semantics in terms of "possible worlds" (Kripke, 1963). As far as we know, only Osherson (1976) has formulated a set of modal rules of inference intended to account for human reasoning. He used rules as an alternative to axioms in a method of formalizing logic known as "natural deduction." And he used the system to explain the inferences of naive individuals, that is, those who have not mastered logic, the probability calculus, or their cognates. His effort was pioneering, but, as he remarked, not altogether successful (Osherson, 1976, p. 232). However, his system illuminates why our opening inference is not valid in logic. We abbreviate the inference as follows:

A or B.

Therefore, possibly (A and B).

The inference cannot be proved in Osherson's system, or indeed in any modal logic, because, if *A* implies *not B*, then the conclusion would assert that a self-contradiction is possible:

Possibly (not-B and B).

But, a self-contradiction is impossible. The proof of the inference in any modal logic calls for an additional premise from background knowledge to rule out such self-contradictions. Logicians refer to inferences that depend on missing premises as "enthymemes." Quite what the additional premise should be is problematic. Our immediate intuition was that it should be that it is not the case that *A* implies that *not-B*:

Not (A $\rightarrow$ not-B).

The arrow denotes the logical connective known as the material conditional, which is similar to a conditional, such as: *If there's an ace then there's not a king*. But, there are divergences between the two. For example, the preceding negation is equivalent to the following assertion in logic:

A and B.

So, our candidate for the missing premise is stronger than the original conclusion that it was supposed to prove: *possibly (A and B*). We consulted an expert, who suggested that the missing premise should be:

Possibly (not (A $\rightarrow$ not B)).

But, this premise is likewise equivalent in logic to:

Possibly (A and B).

Hence, it assumes the conclusion to be proved, and the original premise, (*A or B*), plays no role in the inference.

A noncircular argument, which we formulated with help from Dan Osherson (personal communications, June 29, 2015), is based instead on the additional premise:

Possibly (A $\leftrightarrow$ B)

where "$\leftrightarrow$" abbreviates the conjunction of two material implications: $A \rightarrow B$ and $B \rightarrow A$. This premise together with the original one, (*A or B*), allows one to prove the required conclusion. However, no algorithm exists for finding additional premises of the appropriate logical form for enthymemes, and our difficulty in finding one casts doubt on whether naïve reasoners proceed in this way. It suggests that we should consider alternative explanations for why reasoners should tend to accept the inference.

Probabilistic Logic

Given the following two premises:

If Paul goes fishing he has a fish supper.

Paul goes fishing.

most individuals draw the conclusion:

Paul has a fish supper.

But, in seminal research, Byrne (1989) showed that with the addition of a further premise:

If Paul catches some fish he has a fish supper.

most individuals no longer draw the preceding conclusion. This result, which is robust, illustrates what is known as "nonmonotonic" reasoning: Further premises may lead to the weakening or to the withdrawal of the conclusion of a valid deduction. This provision allows reasoners to make tentative or defeasible inferences, which are commonplace in daily life. By contrast, inferences in orthodox logic are monotonic. As further premises accrue, new valid deductions can be made, and it is never necessary to withdraw an earlier conclusion.

Adams (e.g., 1975, 1998) was dissatisfied with the monotonicity of orthodox logic and with its treatment of conditionals as material implications. His solution was to turn to the probability calculus, at least in the case of conditionals, and to formulate an account of probabilistic logic (p-logic) in which to a first approximation "at least as probable" is substituted for "valid." He allowed that p-logic departs from classical logic only for conditionals (Adams, 1998, p. 189). But, he regarded p-logic as elucidating human reasoning, and his approach was one of the inspirations for the present studies. Hence, we outline the theory of p-logic.

The lower the probability of a proposition, *A*, the more uncertain it is, where "uncertainty" is perhaps better thought of as informativeness (Bar-Hillel, 1964; Johnson-Laird, 1983; Suppes, 1966). In the case of immediate inferences from a single premise to a single conclusion, as in our opening example, the principle governing a probabilistically valid (p-valid) inference is simple. For any coherent assignment of probabilities, that is, one that does not violate the probability calculus, the probability of the conclusion cannot be less than the probability of the premise:

probability (premise) $\leq$ probability (conclusion)

In other words, p-validity demands that the conclusion is not more informative than the premises. Adams's other step is a radical treatment of conditionals in which he argued that they do not correspond to material implications. They are neither true nor false, but have a probability equal to the conditional probability of the *then*-clause given the *if*-clause. We say no more here about conditionals, because our studies did not investigate them.

P-logic is a cornerstone of an avowed "new paradigm," which seeks to replace logic with probability, and to replace studies of deduction with studies of probabilities (e.g., Evans, 2012; Oaksford & Chater, 2007; Over, 2009). The new paradigm presupposes that degrees of belief are, in essence, subjective probabilities—a point that is relevant to the well-known finding that beliefs bias reasoning. Proponents of the new paradigm tend to accept p-logic, though not all of them accept p-validity (Pfeifer & Kleiter, 2009; for a review, see Johnson-Laird et al., 2015). Likewise, p-logic is monotonic (Chater & Oaksford, 2009; Over, 2009) contrary to Adams's avowed purpose in developing it (Adams, 1998, p. 3), though it has inspired nonmonotonic systems of reasoning (Kraus, Lehmann, & Magidor, 1990; Pearl, 1988, Ch. 10).

Consider the following sort of inference about the cards in a hand:

There's an ace or there's a king but not both.

Therefore, there's an ace or there's a king or both.

It is valid in logic, because the conclusion is true in every case in which the premise is true. It is also p-valid, because the conclusion is at least as probable as the premise if not more so. The converse inference:

There's an ace or there's a king or both.

Therefore, there's an ace or there's a king but not both.

is likewise both invalid and p-invalid. If there's both an ace and a king, the premise is true, but the conclusion is false, and so the conclusion can be less probable than the conclusion.

According to p-logic, deductions that are valid in logic (conditionals excepted) are always p-valid, because their conclusions cannot be less probable than their premises. Consider again, our opening inference about the hand of cards, which is of the sort:

There's an ace or there's a king or both.

Therefore, possibly (there's an ace and there's a king).

As in logic, if the presence of an ace doesn't rule out the presence of a king, then the inference is p-valid. In this case, the conclusion is consistent with any case in the *partition* of events:

ace and king
ace and not-king
not-ace and king
not-ace and not-king

We can therefore formulate a normative account of the probability of conclusions referring to possibilities, such as the conclusion to the present inference. Because it refers to the first conjunct as merely a possibility, it does not exclude the occurrence of any of the other cases in the partition. Its probability is 100%. The disjunctive premise, however, refers only to the first three of the cases in the partition. Its probability could be less than 100%. Hence, the conclusion is at least as probable than the premise. The inference is therefore p-valid, and if p-logic guides human reasoners, then, granted that *ace* doesn't imply *not-king*, they should accept it. They should also judge that the conclusion is at least as probable as the premise.

The approach presupposes that individuals' estimates conform to the probability calculus. When they can rely on simple additive computations, their estimates can be accurate (Juslin, Nilsson, & Winman, 2009), though the next section presents results to the contrary. And a classic finding is that when they rely on heuristics such as "representativeness," their estimates are often subadditive (Tversky & Kahneman, 1983), that is, they yield probabilities in the cases in the partition that sum to a probability greater than 100%. For example, they estimate the probability of a conjunction, *A and B*, as greater than the probability of one of its conjuncts, that is, they commit the "conjunction fallacy." They also produce subadditive estimates if the probability of an event must be unpacked into estimates of its constituents, for example, when unpacking the probability of death into the probability of death from illness and the probability of death from other causes (Rottenstreich & Tversky, 1997). Proponents of the new paradigm have not yet proposed any independent account of subadditivity, or any account of where the numbers come from in estimates of probabilities. Their focus is on what is computed rather than on how it is computed. However, as the next section shows, the theory of mental models proposes a computational process that explains the provenance of numbers in estimates of probability and the causes of subadditivity.

Mental Models

The theory of mental models—the "model theory" for short—accounts for what is computed in reasoning: to deduce is to maintain semantic information, to simplify, and to reach a new conclusion (Johnson-Laird & Byrne, 1991, Ch. 2). It also accounts for how the computations are carried out, postulating that inferences are based on models of sets of possibilities (e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). A disjunction of the sort: *There's an ace or there's a king or both*, refers to a set of three possibilities:

ace	
	king
ace	king

This diagram depicts the *mental* models of the possibilities, but the *fully explicit* models of the disjunction also represent what is false in each possibility, using negation to do so:

ace	not-king
not-ace	king
ace	king

Intuitive reasoning is based on mental models, typically a single mental model of sentential premises (e.g., Johnson-Laird & Byrne, 1991), but in easy tasks, such as the listing of possibilities to which assertions refer, deliberation yields fully explicit models. This account applies to any sort of assertion, but we focus on those that occur in our present studies—conjunctions and disjunctions of specific events. Fully explicit models determine whether or not an inference is valid—whether its conclusion holds in every possibility to which its premises refer (Jeffrey, 1981, p. 1). One divergence from logic, however, is that there must be at least one possibility to which the premises refer: Inferences in daily life are not treated as valid because their premises are contradictory.

The model theory postulates a process of modulation in which the meanings of assertions, their referents, and general knowledge, can block possibilities to which compounds would otherwise refer (Johnson-Laird & Byrne, 2002). It works by conjoining models of these cases with models of the premises. For instance, the assertion:

It rained or it poured.

refers to only these fully explicit possibilities:

rained	not-poured
rained	poured.

The meaning of *poured* rules out the possibility in which poured but didn't rain, because pouring is raining. It follows from the disjunction that it rained, and all that is uncertain is whether or not it poured. Modulation can also add relations between clauses (see Juhos, Quelhas, & Johnson-Laird, 2012).

If one assertion, *A*, contradicts another, *not-A*, their conjunction, such as *there's a king and there isn't a king*, yields the null model, akin to the empty set. And the null model does not refer to any possibility. Hence, when individuals know that *A* implies *not-B*, their interpretation of *A or B or both*, yields only models of two possibilities:

A
B

Otherwise, the set includes a model of the third inclusive possibility:

A B

It therefore follows at once in this second case:

Possibly (A and B).

The model theory accordingly predicts that naive reasoners should make this inference, and that they should reject it in case the disjunctive premise is exclusive. Theories based on logic and p-logic treat this inference as invalid unless it can be proved that *A* does not imply not-*B*.

The model theory contrasts with both logic and p-logic in its account of inferences from a disjunctive premise to a disjunctive conclusion. As we showed earlier, an inference of this sort:

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

is both logically valid and p-valid. But, according to the model theory, naive reasoners should tend to reject the inference because the premise does not establish the possibility of *A and B* to which the conclusion refers. The inference from an inclusive to an exclusive disjunction:

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

is neither logically valid nor p-valid. We explain presently the model theory's predictions about this inference.

Possibilities introduce uncertainty into reasoning, and according to the model theory they provide the foundations of probabilistic reasoning. Like probabilities, possibilities in daily life even come in degrees, as in the following intuitive scale, which lies outside modal logics:

impossible—hardly possible—possible—highly possible—impossible not to occur

Many probabilities follow from the proportions of models of possibilities in which they hold, or from the frequencies of these possibilities (Johnson-Laird et al., 1999). When inferences are about unique events, however, they depend on the proportions of possibilities in models of relevant evidence. The theory explains how individuals adduce evidence in order to infer the probabilities of conjunctions and disjunctions, and conditional probabilities (Khemlani, Lotstein, & Johnson-Laird, 2012, 2015). In order to assess, say, the probability that Apple is profitable next year, the model theory postulates that reasoners adduce evidence, such as:

Most companies that are profitable one year are profitable the next year. A mental model of this proposition yields a relevant proportion based on the quantifier, "most:"

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

Given Apple's profitability this year, a simple intuitive system transforms the proportion into a special sort of iconic model:

|—— |

where the left-hand end represents *impossibility*, the right-hand end represents *certainty*, and the length of the icon represents a probability. Models of this sort are often proposed as the way in which non-numerate individuals represent magnitudes (e.g., Barth et al., 2006; Carey, 2009; Gordon, 2004). Reasoners can translate this model into a simple non-numerical claim:

It's highly likely.

The model is kinematic, because the magnitude it represents can be pushed one way or another by other evidence, and it can represent the probability of conjunctions, disjunctions, and conditional probabilities.

Suppose individuals have to estimate the probability of a compound assertion, such as the disjunction or conjunction of two events. How are they to cope when their estimates of the two events differ—one has a high probability and the other a low probability? Naïve individuals who have not mastered the probability calculus are likely to compromise. The result is a rough average of the two probabilities both for conjunctions and disjunctions.

This account has been implemented in a computer program, *mReasoner*, which integrates deduction and probabilities (Khemlani et al., 2015). It uses loops of a small fixed number of iterations to form the rough means of two pointers on an icon, by moving each pointer toward the other until they meet. Deliberations, which are computationally more powerful, can map the resulting iconic model into a numerical probability. The results of the compromise, of course, violate the probability calculus.

When estimates of the probability of *A* and the probability of *B* are combined with those of the probabilities of conjunctions, *A and B*, or disjunctions, *A or B or both*, or conditional probabilities of *B given A*, they fix the joint probability distribution (JPD), that is, the probabilities of each of the conjunctions in the partition: *A and B*, *A and not-B*, *not-A and B*, and *not-A and not-B*. But, the compromise procedure for compounds yields subadditive values for the JPD, that is, it sums to more than 100% (Khemlani et al., 2015). Deliberations, however, can use the partition to figure out more appropriate procedures, such as multiplying the probabilities of conjuncts to estimate the probability of their conjunction, and adding them to estimate the probability of their disjunctions. Experiments have corroborated the model theory's prediction that estimates that fix the values of the JPD tend to be subadditive even when they do not rely on heuristics (Khemlani et al., 2015).

We have surveyed three main theoretical approaches to modal reasoning: modal logic, p-logic, and mental models. We now turn to a series of experiments designed to test their predictions.

Experiment 1

The main aim of our first two experiments was to determine whether naïve individuals tend to make inferences of the sort:

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

and to reject inferences of the sort:

A or B but not both.
Therefore, possibly (A and B).

where the contents of *A* and *B* are sensible everyday assertions.

The model theory predicts these evaluations. Logic and p-logic also predict them granted that *A* does not imply *not-B*. In the case of p-logic, the propensity to accept or to reject them should be reflected in participants' estimates of the respective probabilities of premises and conclusions. Indeed, only these probabilities should matter. That is, individuals should make those inferences in which they judge that the conclusion is at least as probable as the premise, and reject those inferences in which they judge that the conclusion is less probable than the premise. To elicit the participants' intuitive sense of "possibility," the experiment prefaced conclusions with the phrase, "It is possible that . . .," and offered no explanation for how it was to be interpreted. If *A* tends to raise the probability of *B*, as opposed to lower it, then reasoners should be more likely to accept such inferences, because the conclusion *A and B* is more believable. Nevertheless, such effects have no bearing on the status of the inferences in modal logic and in the model theory.

A second aim of Experiment 1 was to examine two different inferences about disjunctive assertions. The first sort, again with sensible everyday contents, was:

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

As we showed in the preceding section, the inference is both logically valid and p-valid, but the model theory predicts that reasoners should tend to reject it, because the premise does not establish the possibility of *A and B*. The second sort of inference was the converse of the first sort:

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

The inference is neither logically valid nor p-valid. But, although the model theory also predicts that it should tend to be rejected, inexpert reasoning might sometimes lead to its acceptance on the grounds that the conclusion refers only to possibilities consistent with the premise. These two sorts of inference were included in part to test these predictions, but also to add some variety to the different sorts of inference in the experiment.

The participants estimated the probabilities of each of the following four assertions, which related to the partition for the premise and conclusion:

It is possible that A and B.
It is possible that A and not-B.
It is possible that not-A and B.
It is possible that not-A and not-B.

These probability estimates allowed us to examine the relations between p-validity and the participants' individual inferences. Normatively, as we showed earlier, each of these probabilities should be 100%.

Method

Participants. The experiment tested 30 participants, who were recruited via the Prolific Academic United Kingdom web service (18 males; 12 females; mean age 28.4, *SD* 9.3). We accepted only individuals who registered as native English-speakers. The majority of them also registered as United Kingdom citizens. They were all paid £1.5 for the roughly 15 to 20 min of the experiment.

Design. The participants acted as their own controls. They evaluated six inferences from inclusive disjunctions to possible conjunctions, four inferences from exclusive disjunctions to possible conjunctions, one inference from an exclusive disjunction to an inclusive disjunction, and one inference from an inclusive disjunction to an exclusive disjunction. There were therefore 12 trials in the experiment proper, preceded by one practice trial.

For half the disjunctive premises in inferences about possible conjunctions, *A* increased the probability of *B*, and for the other half of inclusive disjunctive premises *A* decreased the probability of *B* (according to a norming study in Khemlani et al., 2012).

After the participants had evaluated each inference, they estimated the probability of the possibility of each of the four conjunctions of events in the partition for the inference, including the conjunction that occurred in the conclusion. They then estimated the probability of two disjunctions of the same contents, one inclusive and the other exclusive, so that one member of each pair corresponded to the premise. The order of the trials was random for each participant, the order of the four conjunctive possibilities in the probability judgments was likewise random on each trial, and so too was the order of the two disjunctions in the probability judgments. The set of 12 pairs of sentences were assigned to the sorts of inferences using a Latin square, and so each pair occurred roughly equally often in an inclusive disjunction and in an exclusive disjunction in the experiment as a whole.

Materials. We selected 12 contents from the materials used in two studies of the probabilities of unique events (Khemlani et al., 2012, 2015). The selection process was based on the topicality of the events. The appendix presents these contents, distinguishing between those for which an independent panel of judges had rated *A* as increasing the probability of *B* and those for which the judges had rated *A* as decreasing the probability of *B*. Some of these ratings may not generalize to a United Kingdom population, but the difference was not a major manipulation in the experiment, and the investigators' judgments concurred with them at the time in which the experiments were carried out (Spring, 2015).

Procedure. The experiment was carried out on the Prolific Academic web site, which takes the usual precautions to ensure that individuals carried out the experiment only once. The instructions warned the participants that they would not be recompensed if they carried out the experiment improperly, for example, by guessing or by responding too quickly. They were told that the task was not a test of intelligence or personality. The instructions then explained that the participants had to evaluate inferences, and to assess probabilities of related assertions. The key instruction was: "You will assess 13 inferences in total. While doing so, you have to decide whether or not the conclusion follows of necessity from the premise, that is, given the truth of the premise must the conclusion be true." The experiment began with a single practice trial, though the participants were not told that it was for practice. It called for the participants to evaluate whether a possible conjunction followed from a conditional.

The first page on each trial called for the evaluation of an inference, such as:

> Premise: Scientists will discover a cure for Parkinson's disease in 10 years OR the number of patients who suffer from Parkinson's disease will triple by 2050, OR both.

Conclusion: It is possible that scientists will discover a cure for Parkinson's disease in 10 years AND the number of patients who suffer from Parkinson's disease will triple by 2050.

Does the premise imply that the conclusion is true?

Yes No

The participants made their response by clicking on the "Yes" button or the "No" button. The second page on each trial presented the possibilities from the partition for which they had to estimate probabilities:

What are the chances out of 100 that each of the following assertions (1 to 6) is true?

Choose a number from 0 (*no chance at all*) to 100 (*completely certain*) for each assertion by using the sliders. If you cannot see a complete page on your system, please scroll down to the rest of the page.

1. It is possible that scientists will discover a cure for Parkinson's disease in 10 years AND the number of patients who suffer from Parkinson's disease will triple by 2050.

2. It is possible that scientists will discover a cure for Parkinson's disease in 10 years AND the number of patients who suffer from Parkinson's disease will NOT triple by 2050.

3. It is possible that scientists will NOT discover a cure for Parkinson's disease in 10 years AND the number of patients who suffer from Parkinson's disease will triple by 2050.

4. It is possible that scientists will NOT discover a cure for Parkinson's disease in 10 years AND the number of patients who suffer from Parkinson's disease will NOT triple by 2050.

5. It is possible that scientists will discover a cure for Parkinson's disease in 10 years OR the number of patients who suffer from Parkinson's disease will triple by 2050, OR BOTH.

6. It is possible that scientists will discover a cure for Parkinson's disease in 10 years OR the number of patients who suffer from Parkinson's disease will triple by 2050, BUT NOT BOTH.

Under each of these assertions, there was a slider, which participants could set anywhere from 0% to 100%. It also displayed the exact percentage setting of the slider. Its initial position was at 50%, but there was a warning: "50% chance—Slider not modified yet." If a participant failed to move the slider, another warning appeared advising the participant to set its value, and so it was necessary to move the slider away from and then back to 50% in order to choose that value. Once the participant had completed a trial, and clicked on the "continue" button, the program presented the next trial. After the completion of the experiment, a further page thanked the participant, and explained how to obtain payment.

Results and Discussion

Table 1 presents the overall acceptances for the four sorts of inference. One participant accepted every inference, suggesting an improper grasp of the problems, and so we dropped this participant's data from the analyses. We report first the analysis of the two sorts of inference from disjunctions to possible conjunctions, then the analysis of the two sorts of inference from a disjunctive premise to a disjunctive conclusion, and finally an analysis of individual differences.

A total of 26 out of the 29 participants made a higher proportion of acceptances of possible conjunctions from inclusive disjunctions than from exclusive disjunctions, and there was one tie (Binomial test, $p < .00001$). Hence, the participants were not evaluating inferences at random, and their performance corroborated the predictions of the model theory. There was no difference in the mean times to evaluate inferences from inclusive disjunctions (24.1 s, *SE* 3.0) and those from an exclusive disjunction (26.8 s, *SE* 5.6; Wilcoxon's test, $z = 0.10$, $p = .92$, two-tail).

Table 2 presents the overall estimates of the probabilities of the four conjunctive possibilities depending on whether *A* increased or decreased the probability of *B*. The difference between these two conditions was not significant (Wilcoxon's test, $z = 1.22$, $p = .22$), but there was a reliable interaction in the estimates of the four conjunctions of possibilities (Wilcoxon's test, $z = 2.78$, $p < .005$, Cliff's $d = -.45$). This interaction is attributable to two effects. On the one hand, when *A* increased the probability of *B*, the participants estimated the probability of *possibly A and B* as lower than the probability of *possibly A and not-B*, but to a lesser degree than when *A* decreased the probability of *B* (Wilcoxon's test, $z = 2.73$, $p < .005$, Cliff's $d = -.40$). On the other hand, there was no reliable effect on the estimates for the other two contingencies: *possibly not-A and B* and *possibly not-A and not-B* (Wilcoxon's test, $z = 1.74$, $p = .08$, two-tail). As Table 2 shows, the probabilities of the four contingencies summed to more than 100%. This result does not violate the probability calculus. The participants were not estimating the probability of each conjunction in the JPD, but instead the probability of their possibilities. For example, *A and B* is possible in any of the contingencies in the JPD, and so its probability could be as high as 100%. The minimum sum of the four probabilities, however, must be at least equal to 100%. There

Table 1
The Four Sorts of Inference in Experiment 1, Their Status in Logic and p-Logic Granted That It Is Possible That A Mutually Implies B, the Model Theory's Predictions, and the Percentages of Acceptances in the Experiment

The four sorts of inference	Status in logic and p-logic	The model theory's predictions	Percentages of acceptances
1. A or B or both. Therefore, possibly A and B.	Valid	Accept	82
2. A or B but not both. Therefore, possibly A and B.	Invalid	Reject	10
3. A or B but not both. Therefore, A or B or both.	Valid	Reject	3
4. A or B or both. Therefore, A or B but not both.	Invalid	Reject	24

Table 2
The Mean Estimates of the Percentage Probabilities of the Four Conjunctive Possibilities in Experiment 1 for the Inferences From Disjunctions to Possible Conjunctions, Depending on Whether A Increases, or Decreases the Probability of B (as Assessed in a Prior Norming Study)

Effect of A on p(B)	Conjunctions prefaced with "It is possible that . . ."				
	A and B	A and not-B	not-A and B	not-A and not-B	Sum
A increases p(B)	52	58	58	55	222
A decreases p(B)	42	63	59	51	215
Overall	47	61	59	53	220

was only one trial in the whole experiment that violated this minimum, and there were only three trials on which the probabilities summed to exactly 100%. Twenty-six participants (out of 29) made estimates that summed to more than 100% on every trial, with values approaching 400% for one participant. We explain these estimates and those of the next experiment in the General Discussion.

In general, if individuals are assessing p-validity for each inference from a disjunction to a possible conjunction, then they should tend to accept those inferences for which they estimated that the conjunctive conclusion was at least as probable as the disjunctive premise, and to reject those inferences for which they estimated that the conjunctive conclusion was less probable than the disjunctive premise. The former are p-valid, and the latter are p-invalid. Table 3 shows the relevant results, which do not include the data from three participants, because they did not make p-valid and p-invalid estimates for both sorts of disjunction. Most inferences turned out to be p-invalid (67%), and the relative probabilities of premise and conclusion had only a marginal effect on whether the participants accepted or rejected the corresponding inference (Wilcoxon's test, $z = 1.55$, $p = .06$). The participants simply tended to accept the possible conclusions from inclusive disjunctions, and to reject those from exclusive disjunctions.

Given that p-validity did not yield robust effects, we examined whether the participants were sensitive merely to the subjective probability of the conclusions. The correlation between participants' probability estimates of a conclusion and their acceptances or rejections was small but highly reliable ($R = 0.16$), $F(1, 288) = 56.74$, $p < .0000001$. One interpretation of this relation is that it reflects belief bias: individuals tend to accept likely conclusions and to reject unlikely conclusions (cf., Evans, Barston, & Pollard, 1983; Oakhill & Johnson-Laird, 1985). But, the participants' particular evaluations of the inferences could have influenced their probability estimates.

Turning to the inferences from disjunctive premises to disjunctive conclusions, 22 out of the 29 participants rejected both inferences (Binomial test, $p < .000001$[2] given a prior probability of .25), and only one participant accepted both inferences. Of the remaining six participants who rejected just one of the two inferences, all of them accepted the inference from the inclusive to the exclusive disjunction (Binomial test, $p < .025$ given a prior probability of .5). The pattern of results is incompatible with both logic and p-logic, which predict that reasoners should accept the inference from the exclusive to the inclusive disjunction, and reject the inference from the inclusive to the exclusive disjunction. The model theory predicts the rejection of both inferences. But, as we explained earlier, it allows that inexpert reasoners might make the inference from inclusive to exclusive disjunctions: the premise establishes each of the two possibilities to which the conclusion refers. Hence, as the results showed, they should be more likely to make this second inference than the first one—even though the second inference is flawed, because the premise refers to a possibility, *A and B*, in which the conclusion is false.

Overall, the participants differed in the degree to which they drew the inferences that are valid according to the model theory (Friedman test, $\chi^2 = 106.36$, $p < .0000001$). Some participants made more responses in accord with its predictions than others did. Of course, the model theory's predictions for inferences to possible conjunctions concur with those from logic and p-logic granted an additional premise ruling out self-contradictory conclusions. The participants' ratings of the probabilities of the 12 different conjunctions, *possibly A and B*, showed a small but reliable concordance (Kendall's $W = .09$, $p < .01$). For example, they rated the possibility of the conjunction for Contents 5 as improbable, but the conjunction for Contents 7 as probable (see the Appendix). Readers should bear in mind that the participants rated the probabilities of the conjunction in the context of different sorts of inference. Hence, the concordance is likely to have been higher if every participant had rated the conjunctions for the same sorts of inference (cf. the ratings in Khemlani et al., 2015).

Experiment 2

Experiment 2 examined inferences from the two sorts of disjunction, inclusive and exclusive, to possible conjunctions, in order to corroborate the previous experiment. But, unlike the previous study, it called for estimates of the actual conjunctions in the JPD in order to test the model theory's prediction that these estimates should tend to be subadditive. If participants estimate the probability of each conjunction subadditively, the sum of their four estimates of conjunctions should yield a JPD reliably greater than 100%.

Table 3
The Percentages (and Actual Frequencies) of the Acceptances and Rejections of the Inferences From Disjunctions to Possible Conjunctions in Experiment 1 and Their p-Validity or p-Invalidity From the Participants' Subsequent Individual Estimates of the Probabilities of Premises and Conclusions ($n = 26$)

P-Validity	Inferences from inclusive disjunctions		Inferences from exclusive disjunctions	
	Acceptances	Rejections	Acceptances	Rejections
p-valid	29 (45)	6 (10)	7 (7)	24 (25)
p-invalid	51 (80)	13 (21)	3 (3)	66 (69)

Note. p-valid = participants' estimates of the probability of the conclusion was at least as high as their estimates of the probability of the premise; p-invalid = participants' estimates of the probability of the conclusion was less than their estimates of the probability of the premise.

Method

Participants. The experiment tested a new sample of 30 participants from the same population as before (nine males; 21 females; mean age 27.3, *SD* 8.6). They were each paid £1.5 for the roughly 15 to 20 min of the experiment.

Design. The participants evaluated four inferences from inclusive disjunctions to possible conjunctions, and four inferences from exclusive disjunctions to possible conjunctions. There were therefore eight trials in the experiment proper, preceded by two practice trials. For half of both sorts of disjunctive premise, *A* increased the probability of *B*, and for the other half *A* decreased the probability of *B* (according to the norming study in Khemlani et al., 2012).

After the participants had evaluated each inference, they estimated the probability of the four conjunctions in the JPD, and then the probabilities of the conclusion and of the premise. The order of the trials was random for each participant. After each inference, the order of the four conjunctions in the probability judgments was random on each trial, and so was the order of probability judgments of the premise and conclusion. The set of eight pairs of sentences were assigned to the two sorts of disjunction for each pair using a Latin square, and so each pair occurred roughly equally often in an inclusive disjunction and in an exclusive disjunction in the experiment as a whole.

Materials and procedure. We selected eight contents from those used in Experiment 1 (see the Appendix), rejecting those with the largest and smallest effects on the participants' performance in making inferences according to the model theory. As before, the experiment was carried out on the Prolific Academic web site, using the same instructions and procedure. One practice trial called for the participants to determine whether a possible conjunction followed from a conditional, and the other practice trial called for them to determine whether an actual conjunction followed from a conditional.

Results and Discussion

The participants accepted the inferences of the sort: *A or B or both*; therefore, *possibly A and B* on 91% of trials, but the inferences of the sort: *A or B but not both*; therefore, *possibly A and B* on only 4% of trials. A total of 29 out of the 30 participants made a higher proportion of acceptances of possible conjunctions from inclusive disjunctions than from exclusive disjunctions, and there was one tie, which yielded a vastly significant difference (Binomial test, $p = .5^{29}$). The results accordingly corroborated the model theory, replicating the results of the previous experiment. The difference between the mean response times for inferences from an inclusive disjunction (28.3 s, *SE* 3.5) and those from an exclusive disjunction (32.5s, *SE* 5.3) was not reliable (Wilcoxon's test, $z = 0.24$, $p = .81$, two-tail).

Table 4 shows the percentages of acceptances and rejections of the inferences depending on whether the participants' estimates of the probabilities of the particular premises and conclusions corresponded to p-valid or p-invalid inferences. Nine participants did not make both p-valid and p-invalid estimates for the two sorts of inference, and so they could not be included in the analysis. As in the previous experiment, most inferences were p-invalid (64%), and whether an inference was from an inclusive disjunction (89% acceptance) or from an exclusive disjunction (only 2% acceptance) had a much greater effect than whether it was p-valid (25% acceptance) or p-invalid (33% acceptance). As a result, inferences from inclusive disjunctions were not reliably affected by p-validity (Wilcoxon's test, $z = 0.9$, $p = .18$), but inferences from exclusive disjunctions were reliably affected by p-invalidity (Wilcoxon's test, $z = 3.45$, $p < .0005$, Cliff's $d = -.85$). The correlation between participants' acceptances and rejections of the two inferences and their probability estimates of the conclusions was moderate but highly reliable ($R = .33$), $F(1, 298) = 148.70$, $p < .0000001$. As in the previous study, this result may reflect belief bias or the influence of inferences on the probability estimates.

Table 4
The Percentages (and Actual Frequencies) of the Acceptances and Rejections of the Inferences to Disjunctions in Experiment 2 Depending on Whether the Participants' Estimates of the Probabilities of the Particular Premises and Conclusions Corresponded to p-Valid Inferences or to p-Invalid Inferences (n = 21)

	Inferences from inclusive disjunctions		Inferences from exclusive disjunctions	
P-Validity	Acceptances	Rejections	Acceptances	Rejections
p-valid	49 (41)	4 (3)	1 (1)	19 (16)
p-invalid	40 (34)	7 (6)	1 (1)	79 (66)

Note. p-valid = participants' estimates of the probability of the conclusion was at least as high as their estimates of the probability of the premise; p-invalid = participants' estimates of the probability of the conclusion was less than their estimates of the probability of the premise.

Table 5 presents the overall probability estimates of the four conjunctions in the JPD depending on whether *A* increased or decreased the probability of *B*. There was no reliable difference between these two conditions (Wilcoxon's test, $z = .20$, $p = .85$), but a reliable interaction over the four conjunctions, akin to the interaction in the previous experiment (Wilcoxon's test, $z = 3.07$, $p < .005$, Cliff's $d = -.52$). This interaction is attributable to two effects. On the one hand, when *A* increased the probability of *B*, the participants estimated the probability of *A and B* as lower than the probability of *A and not-B*, but to a lesser degree than when *A* decreased the probability of *B* (Wilcoxon's test, $z = 3.22$, $p < .0005$, Cliff's $d = -.38$). On the other hand, there was no reliable effect on the estimates for the other two contingencies: *not-A and B* and *not-A and not-B* (Wilcoxon's test, $z = 1.53$, $p > .12$, two-tail).

As we mentioned earlier, previous studies have shown that naïve individuals tend to make estimates of probabilities yielding subadditive JPDs of two events, that is, they summed to more than 100% (Khemlani et al., 2012, 2015). In the present experiment, the estimates were massively subadditive, with an overall mean of 191%. Twenty-five out of the 30 participants made subadditive estimates on all eight experimental trials (Binomial, $p < .0002$). Three participants made estimates that summed exactly to 100% on all eight trials; perhaps they had acquired some knowledge of the probability calculus. And only five trials yielded superadditive JPDs, summing to less than 100%. According to the model theory, the massive subadditivity occurred because the participants in the present study estimated the probabilities of each of four conjunctions, whereas the participants in the earlier studies estimated the

Table 5
The Mean Estimates of the Percentage Probabilities of the Four Conjunctions in the Joint Probability Distribution (JPD) in Experiment 2, Depending on Whether A Increased, or Decreased, the Probability of B (as Assessed in a Norming Study)

	Conjunctions in JPD				
Effect of A on p(B)	A and B	A and not-B	not-A and B	not-A and not-B	Sum
A increases p(B)	41	54	57	38	191
A decreases p(B)	34	62	60	35	191
Overall	37	58	59	36	191

probabilities of only one conjunction (or one other compound, such as a disjunction or a conditional probability). If each conjunction in the JPD tends to yield subadditivity, then four such conjunctions should increase the size of the JPD's subadditivity. The estimates in the present study were reliably lower than those in Experiment 1 (Mann–Whitney *U* test, W = 271.5, $z = 2.47$, $p <$ .01, Cliff's $d = -.38$), which suggests that the participants in Experiment 1 took into account that they were estimating probabilities of the *possibility* of each of the four conjunctions.

The participants differed in their degree of competence in the inferences (Friedman test, $\chi^2 = 60.18$, $p < .0001$). They also differed in the extent to which their estimates of the probabilities in the JPD were subadditive (Friedman test, $\chi^2 = 124.08$, $p <$ .001). Perhaps because the contents no longer contained those that yielded the biggest and the smallest effects on the competence of the participants in Experiment 1, the estimates over the eight different contents for *A and B* did not show a reliable concordance (Kendall's W = .051, $p = .14$). In general, the results bore out the predictions of the model theory, but did not confirm the predictions of p-validity in the case of inclusive disjunctions. They also corroborated the model theory's prediction that naïve individuals tend to make subadditive estimates of the probabilities in the JPD.

Experiment 3

The experiment examined immediate inferences from three sorts of premise: *A or B or both*; *A and B*; and *not both A and B*. The predictions of modal logic and p-logic run in parallel, though p-logic predicts in addition that, for those inferences that participants accept, they should rate the probability of the conclusion as at least as probable as the premise. As we show, the model theory diverges from both logic and p-logic in the case of certain inferences.

The first sort of inferences in the experiment were from inclusive disjunctions, *A or B or both*, for which reasoners should build three mental models:

A
    B
A B

The two previous experiments investigated the putative conclusion:

Possibly (A and B).

The present experiment added two new inferences that the model theory predicts that individuals should accept:

Possibly A
and
Possibly B.

The inferences are p-valid given that neither *A* nor *B* is self-contradictory, and so their validity in logic and p-logic calls for an additional premise ruling out these cases. And, as in the case of possible conjunctions, it is difficult to determine what the missing premise should be. An obvious candidate for the first inference is that *A* is possible, that is, it is not a self-contradiction, but this premise is identical to the conclusion to be proved, and so the actual premise plays no role in the deduction. We leave the discovery of the correct missing premise as an exercise for readers, because we have been unable to find it. Those participants who are guided by p-validity should estimate that the conclusions are at least as probable as the premises. The experiment also added a fourth inference:

Possibly (not-A and not-B).

The disjunctive premise implies that *not-A and not-B* is impossible, and participants should reject this inference.

The second sort of inferences in the experiment were from conjunctions of the sort:

A and B.

to four sorts of conclusion:

Therefore, A.
Therefore, B.
Therefore, A or B or both.
Therefore, Not-A and not-B.

The first three inferences are valid and p-valid, and the fourth inference is neither valid nor p-valid. In contrast, the model theory predicts that the first two inferences should be accepted, but the second two inferences should be rejected. In particular, the truth of the conjunction does not establish truth of the possibility, *not-A and B*, or the possibility *A and not-B*, to which the conclusion, *A or B or both* refers, and so reasoners should tend to reject it.

The third sort of inferences in the experiment were from premises of the sort: *not both A and B*, to four sorts of conclusion:

Therefore, possibly A.
Therefore, possibly B.
Therefore, possibly (not-A and not-B).
Therefore, A and B.

In logic and p-logic, the first three of these inferences are valid provided that, respectively, *A*, *B*, and *not-A and not-B* are not self-contradictions. The fourth inference is invalid and p-invalid. The model theory makes the same predictions, but it predicts a trend: participants should accept *possibly not-A and not-B* more often than *possibly A,* which they should accept more often than *possibly B*, because individuals should flesh out their models of the premise in this order (see Khemlani et al., 2015):

Not-A and not-B.
Not-A and B.
A and not-B.

The experiment also called for the participants to estimate the probabilities of premises and conclusions in order to examine the predictions of p-validity.

Method

Participants. The experiment tested a new sample of 30 participants from the same population as before (20 males; 10 females; mean age 29.2, *SD* 8.5). They were each paid £1.85 for an experiment of approximately 20 min.

Design. The participants evaluated sets of four inferences on each trial. They carried out eight trials of the four inferences from inclusive disjunctions of the sort:

Premise: *A or B or both.*

Conclusions: *Possibly A, possibly B, possibly A and B, possibly not-A and not-B*.

They carried out two trials of the four inferences from conjunctions of the sort:

Premise: *A and B.*

Conclusions: *A, B, not-A and not-B, A or B or both.*

And they carried out two trials of the four inferences from negative conjunctions of the sort:

Premise: *not both A and B.*

Conclusions: *Possibly A*, *possibly B*, *possibly not-A and not-B*, *A and B*.

There were accordingly 12 trials in the experiment proper, and one initial practice trial. For each sort of premise, half the contents were such that *A* increased the probability of *B*, and half the contents were such that *A* decreased the probability of *B* (based on the norming study in Khemlani et al., 2012). The conclusions were in a random order on each trial. After the participants had evaluated the set of inferences in a trial, they estimated the probabilities of the premise and of each of the conclusions, which were presented in a random order. We used a Latin square to assign contents to sorts of inference.

Materials and procedure. We used the same materials as in Experiment 1, and carried out the experiment on the Prolific Academic web site. The instructions were similar to those in Experiment 1, except for the key instruction: "In each problem you first have to decide whether or not assertions follow of necessity from a premise, that is, given the truth of the premise must the assertions be true?" The first page on each trial called for the evaluation of four inferences from the premise presented at the top of the page. Below the premise, the page showed the first conclusion, and the participants evaluated whether or not the premise implied that the conclusion was true, making their response by clicking either a "Yes" or a "No" button. After they clicked one of these buttons, the next conclusion appeared below. This procedure repeated until the participants had evaluated the fourth conclusion. When they clicked a "continue" button, the second page appeared. But, until they clicked this button, they could rethink their answers and change their evaluations for any of the four inferences. On the second page, the premise and conclusions were presented (in a random order) with a slider below each assertion on which the participants estimated their probabilities. The practice trial presented an exclusive disjunction, *A or B but not both*, and the participants evaluated four conclusions: *A*, *B*, *A and B*, and *A or B or both*.

Results and Discussion

Table 6 shows the overall acceptance and rejection of conclusions according to the model theory's predictions, and the relation between acceptance and rejection and the participants' estimates of the probabilities of the conclusions, which we have dichotomized into those that they estimated as having a probability greater than or equal to 50% and those that they estimates as having a probability less than 50%. We rejected the data from one participant, because of a malfunction in the program. Overall, the participants accepted 81% of conclusions that the model theory predicts that they should accept, and they rejected 79% of conclusions that the model theory predicts that they should reject. Its predictions were more often correct than incorrect (Wilcoxon's test, $z = 4.60$, $p < .000005$, Cliffs' $d = -.99$). The participants' acceptances and rejections correlated with their dichotomized probability estimates (Wilcoxon's test, $z = 4.62$, $p < .000005$, Cliffs' $d = -.99$). They accepted 83% of conclusions that they estimated as having a probability greater or equal to 50%, and they rejected 76% of conclusions that they estimated as having a probability less than 50%. However, a reliable interaction showed that the tendency to fit the model theory's predictions was greater than the tendency to fit the probabilities (Wilcoxon's test, $z = 2.49$, $p < .01$, Cliffs' $d = -.25$). The model theory accordingly outperformed probability in predicting the participants' evaluations of the inferences.

Table 6
The Percentages of Participants' Acceptances and Rejections of All Inferences in Experiment 3 Depending on Whether the Model Theory Predicted Acceptance or Rejection, and on Whether the Participants' Individual Estimates of the Probabilities of Conclusions Was at Least Equal to 50% or Else Lower Than 50% (n = 29)

The model theory's predictions	Participants accept		Participants reject	
	>= 50%	<50%	>= 50%	<50%
Model theory predicts accept	69	12	8	11
Model theory predicts reject	15	6	15	64

Table 7 presents the mean percentage estimates of the probabilities of premises and conclusions, and the overall percentages of acceptances for each of the inferences in Experiment 3.

For the inferences from *A or B or both*, as the table suggests, the mean acceptance of the three conclusions:

Possibly A, possibly B, possibly A and B

was greater than the mean acceptance of the conclusion:

Possibly not-A and not-B.

The difference was reliable (Wilcoxon's test, $z = 4.62$, $p < .00001$, Cliffs' $d = -.90$). In addition, the mean acceptance of the conclusions *possibly A* and *possibly B* was reliably greater than the mean acceptance of *possibly A and B* (Wilcoxon's test, $z = 2.43$, $p < .01$, Cliffs' $d = -.16$). This difference suggests that individuals were aware that the conjunction of the events *A* and *B* was less probable than the probability of either one of them alone.

For the inferences from *A and B*, the mean acceptance of the two conclusions:

A

and

B

was greater than the mean acceptance of the two conclusions:

A or B or both.

Not-A and not-B.

Table 7
The 12 Sorts of Inference in Experiment 3, the Model Theory's Predictions, the Participants' Mean Percentage Estimates of the Probabilities of Premises and Conclusions, and the Percentages of Their Acceptances of the Inferences (n = 29)

Sorts of inferences	The model theory's predictions	Participants' mean percentage estimates of probabilities	Participants' percentages of acceptances of inferences
1. A or B or both		76	
∴ Possibly A	Accept	61	91
∴ Possibly B	Accept	62	94
∴ Possibly A and B	Accept	56	88
∴ Possibly not-A and not-B	Reject	23	18
2. A and B		73	
∴ A	Accept	74	88
∴ B	Accept	71	88
∴ A or B or both	Reject	56	45
∴ Not-A and not-B	Reject	23	14
3. Not both A and B		70	
∴ Possibly A	Accept	50	69
∴ Possibly B	Accept	56	72
∴ Possibly not-A and not-B	Accept	51	55
∴ A and B	Reject	22	16

The difference corroborates the model theory's prediction, and it was reliable (Wilcoxon's test, $z = 3.70$, $p < .0005$, Cliffs' $d = -.58$).

For the inferences from *not both A and B*, the mean acceptance of the three conclusions:

Possibly A, possibly B, possibly not-A and not-B

was reliably greater than the mean acceptance of the conclusion:

A and B.

The difference was again highly reliable (Wilcoxon's $z = 4.07$, $p < .00005$, Cliffs' $d = -.75$).

The model theory predicts a declining trend in the acceptances of the three inferences. The conclusion *possibly (not-A and not-B)* should be accepted more often than *possibly B,* which should be accepted more often than *possibly B,* but the rank-order trend was only marginal (Page's L = 361, $z = 1.64$, $p = .05$).

Only the probability estimates for inferences from the conjunction, *A and B*, fix the JPD. They were subadditive. The means in Table 7 show that participants tended to estimate the probability of *A and B* as falling between their estimates of the probabilities of its conjuncts, and, most strikingly, their estimate of the probability of the disjunction, *A or B or both*, was lowest of all. These results replicate those of Khemlani, Lotstein, and Johnson-Laird (2015), and so we spare readers the statistical details. These estimates, of course, violate the probability calculus.

Table 8 shows the percentages of acceptances and rejections of all the inferences depending on whether the participants' probabilities estimates showed that the inferences were p-valid or p-invalid. Overall, p-valid inferences were accepted reliably more often than p-invalid inferences (Wilcoxon's test, $z = 3.18$, $p < .001$, Cliffs' $d = -.70$, $n = 28$). The difference was reliable for inferences from the premise: *A and B,* and from the premise: *not both A and B* (Wilcoxon's tests, $z = 3.95$, $p < .0005$, Cliffs' $d = -.94$, $n = 24$; $z = 2.61$, $p < .005$, Cliffs' $d = -.72$, $n = 17$, respectively), but it was not reliable from the premise: *A or B or both* (Wilcoxon's test, $z = .94$, $p = .17$, $n = 22$). Finally, as in the previous studies, the participants differed in their degree of ability in making inferences (Friedman test, $\chi^2 = 129.70$, $p < .0000001$).

General Discussion

Consider the following immediate inference from a single premise to a conclusion:

Intelligent alien life is found outside the solar system in the next 10 years, or world governments dedicate more resources to contacting extraterrestrials, or both.

Therefore, it is possible that intelligent alien life is found outside the solar system in the next 10 years and world governments dedicate more resources to contacting extraterrestrials.

As far as we know, psychologists have not studied these sorts of inference before. They have the following grammatical form:

A or B or both.

Therefore, it is possible that A and B.

The premise has the mental models:

```
A
     B
A    B
```

If knowledge shows that *A* and *B* are incompatible, then it blocks the construction of the third model. But, it should not be

Table 8
The Percentages (and Actual Frequencies) of the Acceptances and Rejections for Inferences From All Sorts of Premises in Experiment 3 Depending on Their p-Validity or p-Invalidity in the Participants' Individual Estimates of the Probabilities of Premises and Conclusions (n = 28)

	Overall inferences	
P-Validity	Acceptances	Rejections
p-valid	35 (470)	10 (135)
p-invalid	27 (363)	28 (376)

Note. p-valid = participants' estimates of the probability of the conclusion was at least as high as their estimates of the probability of the premise; p-invalid = participants' estimates of the probability of the conclusion was less than their estimates of the probability of the premise.

blocked for the contents in our experiments. Hence, the model theory predicts that in the absence of such blocking individuals should accept these inferences. In our studies, most participants did accept them. We tested only modest numbers of participants, but replicated the result three times, each with high significance: 82% in Experiment 1 (n = 29), 91% in Experiment 2 (n = 30), and 88% in Experiment 3 (n = 30). The results accordingly corroborated the model theory.

Our studies also examined inferences from one sort of disjunction to another (e.g.):

Intellectual property law in the U.S. will be updated to a reflect advances in technology by the year 2040, or Russia will become the world center for software development by 2040, but not both.

Therefore, intellectual property law in the U.S. will be updated to a reflect advances in technology by the year 2040, or Russia will become the world center for software development by 2040, or both.

In Experiment 1, nearly everyone responded, "No" (97%). The inference is from an exclusive disjunction to an inclusive disjunction:

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

It is logically valid, because the conclusion is true in every case in which the premise is true. Likewise, it is probabilistically valid (Adams, 1998), because for any consistent assignment of probabilities the conclusion is at least as probable as the premise, if not more probable. So, why did reasoners baulk? According to the model theory, the premise does not imply the set of possibilities to which the conclusion refers. In particular, nothing in the premise implies the possibility of *A and B*, and so naïve reasoners reject the inference. Participants in Experiment 1 also rejected the converse inference from an inclusive to an exclusive disjunction, but reliably fewer responded, "No" (76%). The model theory explains this response on the grounds that poorer reasoners may judge that the inference follows because the conclusion refers only to possibilities consistent with the premise.

One potential explanation of our main results is that the participants evaluated inferences using a matching strategy: they accepted just those conclusions that matched the polarity of their premises, whether affirmative or negative (Mike Oaksford, personal communication, September 21, 2015). They therefore accepted the inference:

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

Both premise and conclusion are affirmative. And they rejected the inference:

A or B or both.
Therefore, possibly not-A and not-B.

The premise is affirmative but the conclusion is negative. When participants rely on intuitions alone, such factors may affect their judgments (see, e.g., Oaksford & Stenning, 1992). But, at least two results cast doubt on its general use in our studies. First, it does not explain why participants were more likely to accept the inference in Experiment 1 (see Table 1):

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

than the converse inference. Second, a difference in Experiment 3 runs counter to the matching strategy (see Table 7). From the premise, *not both A and B*, participants accepted mismatching conclusions, such as *possibly A* (on 69% of trials) and *possibly B* (on 72% of trials) more often than they accepted a matching conclusion: *not-A and not-B* (on 55% of trials). A test for the future is to examine inferences of the sort:

A or B or both.
Is possible that A and not B?

Matching predicts that individuals should reject the inference, whereas the model theory predicts that they should accept it.

Granted that individuals reason more often than they match, the two salient alternatives to the model theory are modal logic (Osherson, 1976) and probabilistic logic (Adams, 1998). In probabilistic logic, the concept of probabilistic validity (p-validity) replaces logical validity, and an inference is p-valid provided that given any consistent assignment of probabilities its conclusion is at least as probable as its premises (see, e.g., Cruz, Baratgin, Oaksford, & Over, 2015; Evans, 2012, p. 11; cf. Johnson-Laird et al., 2015). Logic and p-logic have a common treatment of inferences, such as:

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

They are valid and p-valid only if *A* does not imply *not-B*. Otherwise, the conclusion would then be equivalent to: *possibly(B and not-B*), and self-contradictions are not possible and have a probability of zero. A plausible assumption is that the force of "or both" in the statement of the premise rules out this possibility. Yet, the proof of the conclusion in modal logic is not obvious. It calls for reasoners to adduce a hitherto missing premise. A plausible candidate is: It's possible that *A* does not imply *not-B*. But, as our account of modal logic showed, this premise is equivalent to the very conclusion to be proved. A noncircular argument depends on the additional premise that it is possible that *A* implies *B* and that *B* implies *A*. The profound difficulty for theories based on modal logic, however, is that no algorithm exists for formulating missing premises of the appropriate logical form, that is, those that do not merely assert the conclusion to be proved, but that make it possible to prove this conclusion from the given premise. In contrast, for the model theory, it is the lack of information—which would otherwise block the model of *A and B*—that allows the inference to be made. This feature of "modulation" is part of the computer program implementing the model theory.

A second challenge to logic and p-logic from our results is that inferences from an exclusive disjunction to an inclusive disjunction are valid, and yet most reasoners reject them:

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

As we mentioned earlier, the model theory predicts their rejection on the grounds that the premise fails to establish the possibility of *A and B*. Proponents of logic and p-logic could invoke Grice's (1989) maxim of quantity, that is, that speakers be informative. The preceding inference violates the maxim. It throws semantic information away by adding a disjunctive alternative to the premise in its conclusion. But, how do individuals determine that the inference throws away information? They could consider the respective possibilities to which the premises and conclusion refer. That leads directly back to the model theory. Moreover, because an inference from the inclusive to an exclusive disjunction does not violate the maxim of quantity, a Gricean explanation fails to explain why reasoners tend to reject it.

Our results present further challenges to probabilistic logic. The participants' estimates of the probabilities of the premises and conclusions did not determine their acceptance or rejection of inferences. Most of the participants' inferences in all three experiments were p-invalid. However, these studies gathered estimates of probabilities after the participants had made inferences. A recent study called for estimates in assessments of inferences, and yielded stronger support for p-validity (Evans, Thompson, & Over, 2015). But, in our experiments, its predictions were less successful than those of the model theory.

Yet, the participants were sensitive to probabilities. Their estimates of the probabilities of *A and B* and of *A and not-B* reflected the difference between those contents in which *A* increased the probability of *B* and those in which *A* decreased the probability of *B* (according to an earlier norming study, Khemlani et al., 2012). Likewise, in all three of our experiments, participants' acceptance of conclusions correlated with their estimates of the conclusions' probabilities. The results could merely reflect a residual effect from the evaluation of inferences to estimates of the probabilities of their conclusions. But, they could instead reflect the participants' beliefs about the conclusions in a typical case of belief bias (Evans et al., 1983; Oakhill & Johnson-Laird, 1985).

The model theory accounts for how individuals estimate probabilities, both those based on proportions in possibilities (Johnson-Laird et al., 1999), and those based on evidence that fixes the subjective probabilities of unique events of the sort in the present studies (Khemlani et al., 2015). Its mechanisms predict that estimates of the probabilities of conjunctions, or of disjunctions, and of their constituent events should tend to be subadditive, because naive individuals tend to compromise and to take a rough average of the probabilities of their conjuncts (for an account of the underlying mechanism embodied in the *mReasoner* program, see the section above on the model theory).

In Experiment 1, the participants estimated the probabilities of possibilities, such as the probability of the possibility of *A and B*. As the section on the model theory explained, the normative response should be 100%, but the participants' estimates were much lower. Such massive *superadditivity* is very rare in estimates of probability. An anonymous reviewer argued that the model theory has no explanation for these results. In fact, it does explain them albeit post hoc. The participants were using the same mechanism that they used to estimate the probabilities of actual contingencies in Experiment 2. One sign is that the effect of whether *A* increased or decreased the probability of *B* occurred in both experiments (see Tables 2 and 5). Given that the correct estimate for the probabilities of these possibilities is 100%, this factor should have had no effect in Experiment 1. The participants made one correction in this experiment: they added an additional probability for the joint occurrence or nonoccurrence of *A and B* in comparison with the participants' estimates in Experiment 2.

Subadditivity is typical in estimates of the probabilities of actual contingencies (Khemlani et al., 2012, 2015). But, the degree of subadditivity was striking when individuals estimated the probabilities of all four conjunctions in the JPD in Experiment 2. The participants' mean estimates of the probabilities over all the contents were as follows for the four cases in the partition:

A and B: 37.5%
A and not-B: 58%
Not-A and B: 59%
Not-A and not-B: 36.5%

They sum to 191% instead of the 100% that the probability calculus demands. The means here are typical of individual estimates: 83% of the participants made only subadditive estimates on each trial in the experiment. What causes such massive subadditivity? One cause, as we mentioned earlier, is the unpacking of the probability of an event into the probabilities of its constituent events (e.g., Rottenstreich & Tversky, 1997). But, proponents of the probabilistic view of reasoning have yet to provide an independent account of subadditivity or of the processes yielding numerical estimates of probabilities. According to the model theory, estimates of a single conjunction and its conjuncts should tend to be subadditive. Hence, when individuals estimate the probabilities for four conjunctions, the result should be massive subadditivity. Recent findings showed that negations of events tend to elicit underestimates of probabilities (Evans et al., 2015). If the participants were also prone to this bias in our studies, it would decrease subadditivity. So, without this bias, the subadditivity might have been still greater.

The picture that emerges from our studies is clear. On the one hand, naive individuals cope with only rudimentary aspects of probabilities. They are biased against inferring improbable conclusions, and they often infer conclusions that are less probable than the premises, thereby violating p-validity. Their estimates of compound assertions violate the norms of the probability calculus in a systematic way. On the other hand, their reasoning depends on models of possibilities. They infer what's possible from a disjunction, because they can envisage the relevant possibility to which the disjunction refers. Hence, they infer that an inclusive disjunction, *A or B*, implies that *A* is possible (91% of responses in Experiment 3), that *B* is possible (94% of responses), and that *A and B* is possible (88% of responses). These inferences corroborate the model theory's principle that compound assertions, such as disjunctions, refer to conjunctions of possibilities, which mental models and fully explicit models represent (Johnson-Laird et al., 2015). As a corollary, individuals reject the logically valid and probabilistically valid conclusions of immediate inferences in those cases in which nothing in the premise implies one of the possibilities to which the conclusion refers. They therefore reject inferences of the sort: *A or B but not both*; therefore, *A or B or both*, because nothing in the premise implies the possibility of *A and B* to which the conclusion refers. If the present account is on the right lines, then it should be feasible to extend the theory of mental models to give a comprehensive account of modal reasoning in daily life.

References

Adams, E. W. (1975). *The logic of conditionals*. Dordrecht, the Netherlands: Reidel. http://dx.doi.org/10.1007/978-94-015-7622-2

Adams, E. W. (1998). *A primer of probability logic*. Stanford, CA: CSLI Publications.

Bar-Hillel, Y. (1964). *Language and information: Selected essays on their theory and application*. Reading, MA: Addison Wesley.

Barth, H. L. A., Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition, 98,* 199–222. http://dx.doi.org/10.1016/j.cognition.2004.09.011

Bell, V., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22,* 25–51. http://dx.doi.org/10.1207/s15516709cog2201_2

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31,* 61–83. http://dx.doi.org/10.1016/0010-0277(89)90018-8

Byrnes, J. P., & Beilin, H. (1991). The cognitive basis of uncertainty. *Human Development, 34,* 189–203. http://dx.doi.org/10.1159/000277054

Carey, S. (2009). *The origin of concepts*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195367638.001.0001

Chater, N., & Oaksford, M. (2009). Local and global inferential relations: Response to Over (2009). *Thinking & Reasoning, 15,* 439–446. http://dx.doi.org/10.1080/13546780903361765

Cruz, N., Baratgin, J., Oaksford, M., & Over, D. E. (2015). Bayesian reasoning with ifs and ands and ors. *Frontiers in Psychology, 6,* 1–9.

Evans, J. St. B. T. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning, 18,* 5–31. http://dx.doi.org/10.1080/13546783.2011.637674

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11,* 295–306. http://dx.doi.org/10.3758/BF03196976

Evans, J. St. B. T., Thompson, V. A., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology, 6,* 398. http://dx.doi.org/10.3389/fpsyg.2015.00398

Goldvarg, Y., & Johnson-Laird, P. N. (2000). Illusions in modal reasoning. *Memory & Cognition, 28,* 282–294. http://dx.doi.org/10.3758/BF03213806

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306,* 496–499. http://dx.doi.org/10.1126/science.1094492

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

Hughes, G. E., & Cresswell, M. J. (1996). *A new introduction to modal logic*. London, UK: Routledge. http://dx.doi.org/10.4324/9780203290644

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London, UK: Routledge & Kegan Paul. http://dx.doi.org/10.1037/10034-000

Jeffrey, R. C. (1981). *Formal logic*. New York, NY: McGraw-Hill.

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

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

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

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109,* 646–678. http://dx.doi.org/10.1037/0033-295X.109.4.646

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences, 19,* 201–214. http://dx.doi.org/10.1016/j.tics.2015.02.006

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62–88. http://dx.doi.org/10.1037/0033-295X.106.1.62

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

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological Review, 116,* 856–874. http://dx.doi.org/10.1037/a0016979

Khemlani, S. S., Lotstein, M., & Johnson-Laird, P. (2012). The probabilities of unique events. *PLoS ONE, 7,* e45975. http://dx.doi.org/10.1371/journal.pone.0045975

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

Kneale, W., & Kneale, M. (1962). *The development of logic*. Oxford, UK: Oxford University Press.

Kraus, K., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence, 44,* 167–207. http://dx.doi.org/10.1016/0004-3702(90)90101-5

Kripke, S. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica, 16,* 83–94.

Oakhill, J. V., & Johnson-Laird, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *The Quarterly Journal of Experimental Psychology, 37A,* 553–569. http://dx.doi.org/10.1080/14640748508400919

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198524496.001.0001

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 835–854. http://dx.doi.org/10.1037/0278-7393.18.4.835

Osherson, D. N. (1976). *Logical abilities in children Vol. 4. Reasoning and concepts*. Hillsdale, NJ: Erlbaum.

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning, 15,* 431–438. http://dx.doi.org/10.1080/13546780903266188

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Palo Alto, CA: Morgan Kaufmann.

Pfeifer, N., & Kleiter, G. D. (2005). Towards a mental probability logic. *Psychologica Belgica, 45,* 71–99. http://dx.doi.org/10.5334/pb-45-1-71

Pfeifer, N., & Kleiter, G. D. (2009). Framing human inference by coherence based probability logic. *Journal of Applied Logic, 7,* 206–217. http://dx.doi.org/10.1016/j.jal.2007.11.005

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

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 104,* 406–415. http://dx.doi.org/10.1037/0033-295X.104.2.406

Sophian, C., & Somerville, S. C. (1988). Early developments in logical reasoning: Considering alternative possibilities. *Cognitive Development, 3,* 183–222. http://dx.doi.org/10.1016/0885-2014(88)90018-4

Suppes, P. (1966). Probabilistic inference and the concept of total evidence. In J. Hintikka & P. Suppes (Eds.), *Aspects of inductive logic* (pp. 49–65). Amsterdam, the Netherlands: North-Holland. http://dx.doi.org/10.1016/S0049-237X(08)71662-8

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

(*Appendix follows*)

Appendix

The 12 Pairs of Assertions Used in Experiments 1 and 3, and the Eight Pairs Used in Experiment 2 as Shown by the Asterisk(*)

Event *A*	Event *B*
Event A decreases likelihood of event B	
1*. The United States will sign the Kyoto Protocol and commit to reducing CO2 emissions.	Global temperatures reach a theoretical point of no return in the next 100 years.
2*. Intellectual property law in the U.S. will be updated to a reflect advances in technology by the year 2040.	Russia will become the world center for software development by 2040.
3. A nuclear weapon will be used in a terrorist attack in the next decade.	There will be a substantial decrease in terrorist activity in the next 10 years.
4*. The United States adopts an open border policy of universal acceptance.	English is legally declared the official language of the United States.
5. Greece will make a full economic recovery in the next 10 years.	Greece will be forced to leave the EU in the next 10 years.
6*. Scientists will discover a cure for Parkinson's disease in 10 years.	The number of patients who suffer from Parkinson's disease will triple by 2050.
Event A increases likelihood of event B	
7*. A new illegal but synthetic drug becomes popular in the USA over the next 2 years.	The movement to decriminalize drugs doubles its numbers by 2017.
8*. 3-dimensional graphics will be required to contain explicit markers to indicate their unreal nature by 2020.	Competitive video game playing will achieve mainstream acceptance by 2020.
9*. The Supreme Court rules on the constitutionality of gay marriage in the next 5 years.	A gay person will be elected as president in the next 50 years.
10. In less than 15 years, millions of people will live past 100.	Advances in genetics will end the shortage of replacement organs in the next 15 years.
11. Space tourism will achieve widespread popularity in the next 50 years.	Advances in material science will lead to the development of antigravity materials in the next 50 years.
12*. Intelligent alien life is found outside the solar system in the next 10 years.	World governments dedicate more resources to contacting extra-terrestrials.

Note. For the first six pairs, a norming study showed that event A decreases the estimated likelihood of event B, and for the second six pairs, the study showed that event A increases the estimated likelihood of event B.

Received July 20, 2015
Revision received January 20, 2016
Accepted January 24, 2016 ■