Available online at www.sciencedirect.com
SCIENCE DIRECT
Cognitive Psychology xxx (2006) xxx–xxx
Cognitive Psychology
www.elsevier.com/locate/cogpsych

The probability of causal conditionals

David E. Over [a,*], Constantinos Hadjichristidis [b], Jonathan St. B.T. Evans [c], Simon J. Handley [c], Steven A. Sloman [d]

[a] *Psychology Department, University of Sunderland, Sunderland SR6 0DD, UK*
[b] *University of Leeds, UK*
[c] *University of Plymouth, UK*
[d] *Brown University, USA*

Accepted 6 May 2006

Abstract

Conditionals in natural language are central to reasoning and decision making. A theoretical proposal called the Ramsey test implies the conditional probability hypothesis: that the subjective probability of a natural language conditional, P(if p then q), is the conditional subjective probability, P(q|p). We report three experiments on causal indicative conditionals and related counterfactuals that support this hypothesis. We measured the probabilities people assigned to truth table cases, P(pq), P(p¬q), P(¬pq) and P(¬p¬q). From these ratings, we computed three independent predictors, P(p), P(q|p) and P(q|¬p), that we then entered into a regression equation with judged P(if p then q) as the dependent variable. In line with the conditional probability hypothesis, P(q|p) was by far the strongest predictor in our experiments. This result is inconsistent with the claim that causal conditionals are the material conditionals of elementary logic. Instead, it supports the Ramsey test hypothesis, implying that common processes underlie the use of conditionals in reasoning and judgments of conditional probability in decision making.

Keywords: Causal conditionals; Counterfactuals; Conditional probability; Ramsey test; Covariation; Causation; Delta-p rule

* Corresponding author. Fax: +44 0191 515 2308.
E-mail address: david.over@sunderland.ac.uk (D.E. Over).

0010-0285/$ - see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.cogpsych.2006.05.002

1. Introduction

Conditional statements are ubiquitous in both ordinary and scientific discourse. They are used for many purposes, from laying down rules for guiding behaviour to expressing scientific hypotheses (Evans & Over, 2004). One basic use of conditionals is to express uncertainty. We are unsure about the weather, and so we say that we will have an alfresco lunch *if* it is sunny. We are unconvinced by our colleagues' arguments, but conclude that their theory will be confirmed *if* there is a significant result in an experiment. Uncertainty is always with us in human affairs, and indicative conditionals are of great importance for this reason alone. It is unsurprising that so much research has been done on them since the ancient Greeks (Sanford, 1989).

Though people often use a conditional to express uncertainty, they can of course be uncertain about the conditional itself. They can have high or low confidence in it, judging it to have high or low probability. For a Saturday in the summer, our friends can be fairly confident that, if it is sunny, we will have an alfresco lunch. Our colleagues would be less confident that, if we are given a deadline for finishing our marking, then we will meet it.

1.1. The subjective probability of a conditional

Consider an ordinary indicative conditional, of the form 'if p then q,' in natural language:

(1) If the cost of petrol increases (p), then traffic congestion will improve (q).

Suppose (1) is asserted about a possible increase in the cost of petrol (or gasoline) at a particular time in the UK (or the USA). People make subjective probability judgments about conditionals like (1) all the time in ordinary affairs. The question is how they do it. Ramsey (1931/1990, p. 247) hypothesized that people could judge 'if p then q' by '...adding p hypothetically to their stock of knowledge and arguing on that basis about q...' By these means, they would then fix '...their degrees of belief in q given p...,' which would be their degrees of conditional subjective probability, P(q|p). This suggested procedure for making a probability judgment about a conditional came to be known as the *Ramsey test*. It implies that common processes underlie judgments about conditionals and conditional probability. At one extreme, people can deduce q from p with valid inferences in a Ramsey test, and then P(if p then q) and P(q|p) will be 1. At the other extreme, people will judge p and q to be inconsistent, and then P(if p then q) and P(q|p) will be 0. In many more cases, people will use inductive inferences, heuristics, or causal models in a Ramsey test, and then P(if p then q) and P(q|p) will be between 0 and 1. Explaining how the Ramsey test is actually implemented—by means of deduction, induction, heuristics, causal models, and other processes—is a major challenge, in our view, in the psychology of reasoning.

The consequence of the Ramsey test, that P(if p then q) is P(q|p), has been very influential in philosophical logic. Leading philosophical logicians (Adams, 1975, 1998; Bennett, 2003; Edgington, 1995, 2003; Ramsey, 1931; Stalnaker, 1968) have long argued for this consequence. A famous proof in philosophical logic by Lewis (1976) established that P(if p then q) cannot be identical with P(q|p) if we accept some prominent analyses of conditionals, e.g., that of Stalnaker (1968), though it can if we accept others, e.g., that of Adams (1975, 1998). Whether or not P(if p then q) should be P(q|p) has implications in philosophical logic about whether conditionals always express propositions that are objectively true or false (Bennett, 2003, has an account of Lewis's proof and what it implies).

In this paper, we will be concerned with experimental tests of the consequence of the Ramsey test: the *conditional probability hypothesis* that the subjective probability of a conditional, P(if p then q), is the conditional subjective probability of q given p, P(q|p). Rips and Marcus (1977) were the first psychologists to state a theory that implied the conditional probability hypothesis. They were influenced by Stalnaker (1968) and described a kind of Ramsey test for making judgments about conditionals. However, they did not support their theory with experimental studies of the relation between the probability of conditionals and conditional probability (see Evans & Over, 2004, Ch. 4, for an extended account and evaluation of Rips & Marcus's theory). Much more recently, other psychologists and some linguists have become interested in the conditional probability hypothesis (Kaufmann, 2004, 2005; Liu, Lo, & Wu, 1996; Oaksford & Chater, 1998, 2003; Oaksford, Chater, & Larkin, 2000; Politzer & Bourmand, 2002; Stevenson & Over, 1995, 2001). The hypothesis has been supported in a number of psychological experiments (Evans, Handley, & Over, 2003; Evans, Over, & Handley, 2005; Hadjichristidis et al., 2001; Oberauer, Geiger, Fischer, & Weidenfeld, in press; Oberauer & Wilhelm, 2003; Ohm & Thompson, in press; Over & Evans, 2003). On the other hand, Girotto and Johnson-Laird (2004, p. 224) argued that a result in Oberauer and Wilhelm (2003) does not support the conditional probability hypothesis. Girotto and Johnson-Laird also found that a minority of participants conformed to the hypothesis. They concluded (p. 224), '. . .*not* that the conditional probability hypothesis is false, but rather that its use is only *one* of the strategies that naive individuals use in estimating the probability of conditional extensionally' (emphasis in the original).

One limitation of the above experiments (with the exception to some extent of Hadjichristidis et al., 2001 & Over & Evans, 2003) is that they only concern conditionals about frequency distributions that are assumed to be understood by the participants. An example is a conditional about a specific card randomly drawn from a pack of cards of different colours and shapes:

(2) If the card is yellow, then it has a circle on it.

Conditionals like (2) are sometimes called *basic conditionals*. They have a '. . .neutral content that is as independent as possible from context and background knowledge. . .' (Johnson-Laird & Byrne, 2002, p. 648). In most early experiments, minimal background information was given on the relevant frequency distribution, e.g., on the number of cards in the pack of different colours and shapes. People do sometimes make judgments about cases like these in ordinary affairs, a simple gamble being an obvious example. Studying these examples might also reveal how people try to use relevant frequency distributions in more complex or difficult cases.

However, people use many conditionals in ordinary reasoning and decision making when there is no known relevant frequency distribution. The conditional probability hypothesis is about *subjective* probability. It is the claim that people's subjective judgment about P(if p then q) will be the same as their subjective judgment about P(q|p). People can base their subjective confidence in a conditional on relevant frequency information. But in the absence of this, they can use inductive or categorical inferences (Hadjichristidis et al., 2001), heuristics (Evans & Over, 2004), or causal models (Sloman, 2005). The Ramsey test can be seen as a special case of the simulation heuristic of Kahneman and Tversky (1982) applied to conditionals. When running a simulation from p, people can apply other heuristics, such as availability, and inductive inferences and causal models to try to find a relation between p and q (Keren & Teigen, 2004; Lagnado & Sloman, 2004).

Almost all conditionals that are central to reasoning and decision making, from ordinary affairs to scientific research, are non-basic. The assertion of these conditionals depends very much on their 'non-neutral' content and on context and background knowledge. Previous experiments on basic conditionals about simple frequency distributions yield no results about people's probability judgments under the influence of content, context, and background knowledge. Our aim here is to change the emphasis of previous experimental research, by concentrating on non-basic conditionals of a sort that are ubiquitous in real-world reasoning and decision making. We will focus on indicative conditionals like (1) and closely related counterfactuals. Such conditionals can be called *causal conditionals* in the weak sense that they can be justified by evidence about a possible causal relation or mechanism. One question we will ask about these conditionals is whether they are properly called 'causal' in the strong sense that they actually express a causal relation.

We can easily ask people to rate their confidence in (1). However, Girotto and Johnson-Laird (2004) have argued that some care is necessary in how this question is formulated. They contend (p. 208) that a question

(Q1) What is the probability that if p then q?

can be easily re-interpreted as the question

(Q2) If p, what is the probability of q?

They assume that the probability operator in (Q2) is applied only to what they call (p. 208) the main clause, i.e., to the consequent q, rather than to the whole conditional, as it clearly seems to be, to us, in (Q1). They claim that a question of the form (Q2) is '…a direct request for the conditional probability' (p. 223). It is, however, a logical fallacy to claim that (Q2), with the probability operator applied to the main clause, is a request for the conditional probability. This *modal fallacy*, as it is called, implies the absurd consequence that all subjective probabilities collapse to 1 or 0 (Edgington, 1995, p. 269). Another instance of the fallacy is to hold that in ordinary conditionals of the form, 'if p then necessarily q,' the necessity attaches to the main clause. This implies the absurd consequence that all propositions are either necessarily true or necessarily false (see Swartz, 1999, for a general introduction to the modal fallacy that covers its probabilistic instance).

The general view in philosophical logic is that the ambiguity is in (Q2) and not in (Q1). The probability operator in (Q1) is clearly applied to the conditional as a whole, but in (Q2) it might be applied to the whole conditional or only to its main clause. According to Kaufmann (personal communication), the general view among linguists is that people will interpret questions like (Q2) as (Q1), at least where the operator is epistemic. People who responded to (Q2) in an experiment with the conditional probability could be interpreting (Q2) as (Q1). By the conditional probability hypothesis, it is the answer to (Q1), with the probability operator applied to *the whole conditional*, that is $P(q|p)$, the conditional probability judgment. (See again Edgington, 1995, on why the conditional probability hypothesis takes this form.)

When evaluating a causal conditional, 'if p then q,' we can ask whether the antecedent event p causes the consequent event q, in order to justify a conditional probability judgement, P(q|p). But if we attached subjective probability to the main clause q, then our question would become whether the p event caused in us a mental state of confidence in q, and *this* question is not about P(q|p). Consider running a Ramsey test on the conditional, 'if the salesman is a brilliant con man then he is deceiving us.' In the test for this example, we subjectively see a causal relation between his being a brilliant con man and his deceiving us. That is why we have high confidence in the conditional *as a whole* and the conditional probability, P(deceiving us|brilliant con man), is high. If the confidence attached to the main clause, then the question would be whether his being a brilliant con man causes us to have high confidence that he is deceiving us. But of course his being a brilliant con man does not cause us to have high confidence that he is deceiving us, but rather high confidence that he is *not* deceiving us. That illustrates why we should not commit the modal fallacy of imagining that the subjective probability is attached to the main clause: that the con man is deceiving us.

Logicians and linguists call the type of ambiguity that we are illustrating in (Q2) a *scope ambiguity*, as it concerns the scope of application of the probability operator. Questions like (Q2) should not be asked in experiments because of the ambiguity. In fact, to forestall all questions about ambiguity, we have avoided both (Q1) and (Q2) in all our work on conditionals. We have followed a suggestion of Johnson-Laird (personal communication) of asking participants how likely a *claim* or *sentence* like (1) or (2) is to be *true*. In Hadjichristidis et al. (2001) we attributed the claim to a person (called 'Peter'), as suggested later by Girotto and Johnson-Laird (2004). But in other work, we have got similar results just by using phrases like 'the following statement' (Evans et al., 2003; Over & Evans, 2003). We adopted this practice in the experiments reported here, using the following general form.

> For each of the following, could you please rate the probability that the statement is TRUE:
> If the cost of petrol increases, then traffic congestion will improve.
> ...
> ...
> ...

As shown above, we placed the probability operator before the truth operator, and followed that with a list of the conditionals to be rated on separate lines. The purpose of this procedure was to make it as unlikely as possible that the participants would 'jump' the probability operator over the truth operator and then all the way down to the consequent, or the main clause, of each conditional in the list. We note that Girotto and Johnson-Laird (2004, p. 223) report that they '...tried to block the re-interpretation of the question about the probability of a conditional as the conditional probability' but '...individuals still have a tendency to re-interpret the question as a question about a conditional probability.' For some reason that we do not understand, they do not see their difficulty in blocking the conditional probability interpretation as evidence for the conditional probability hypothesis. We do.

We could also directly ask people for their judgments about the conditional probability of q given p. The problem with such a direct question is that ordinary people might not interpret it in the way the experimenters intend. The 'given that' construction is common in technical language. But according to the conditional probability hypothesis itself, the natural way to express a conditional probability judgment in natural language is to express a degree of confidence in a conditional. To use the 'given that' construction in natural language could be confusing and hard for the participants to interpret. In this paper, we will use a new technique (introduced briefly in Over & Evans, 2003) to obtain *implicit* conditional probability judgments. (Manktelow & Over, 1995; Over, Manktelow, & Hadjichristidis, 2004, proposed and used a version of this technique to study deontic conditionals, and Oaksford & Moussakowski, 2004; Oaksford & Wakefield, 2003, used another version of it to study the selection task.)

Using our technique, we ask the participants for their probability judgments about four natural language conjunctions that would often be expressed in ordinary contexts. These conjunctions correspond to the four possible states of affairs that might be relevant to the probability of a conditional of the form 'if p then q' like (1). First, it could true (T) that the cost of petrol will increase and true (T) that traffic congestion will improve. This is a 'p & q,' or TT, possibility. Second, it could be true (T) that the cost of petrol will increase but false (F) that traffic congestion will improve. This is 'p & not-q,' or TF, possibility. Third, it could be false (F) that the cost of petrol will increase but true (T) that traffic congestion will improve. This is a 'not-p & q,' or FT, possibility. And fourth, it could be false (F) that the cost of petrol will increase and false (F) that traffic congestion will improve. This is a 'not-p & not-q,' or FF, possibility.

The participants do not have to understand what 'conditional probability' and the 'given that' construction mean in the technical sense. They only have to give us their degrees of confidence in the four natural language conjunctions: 'p & q' (TT), 'p & not-q' (TF), 'not-p & q' (FT), and 'not-p & not-q' (FF). We can use these probabilities to derive an implicit conditional probability judgment, P(q|p), that is,

$$\frac{\mathrm{P(p\ \&\ q)}}{\mathrm{P(p\ \&\ q) + P(p\ \&\ not\text{-}q)}},$$

for each participant. Finally, we can compare this implicit conditional probability judgment with each participant's explicit probability judgment about (1), 'if p then q'. Notice how strong a test this is of the conditional probability hypothesis. Any computational limitations or biases that people possess that apply differentially to the different kinds of probability judgment work against the prediction that people's explicit judgments about P(if p then q) will equal their implicit P(q|p).

1.2. The material conditional

There is another point of view that makes a different prediction about how the participants' explicit judgments will be related to their implicit judgments. If conditionals like (1) were the material, truth functional conditional of elementary extensional logic, then 'if p then q' would be logically equivalent to 'not-p or q' and be true in TT, FT, and FF, but false in TF. That would mean that P(if p then q) would be P(not-p or q), which is P(p & q) + P(not-p & q) + P(not-p & not-q). Some philosophers (see Bennett, 2003, for a review) would predict this equality, as they have held that conditionals like (1) are simply material conditionals.

In psychology, Johnson-Laird and Byrne (1991) originally claimed that conditionals like (1) are material conditionals. According to their account, people would represent

(1) with an initial mental model of the conjunction, 'p & q,' plus some indication that there are other, implicit models. People would also, in some circumstances, make the implicit mental models fully explicit. These explicit mental models would reveal the conditional to be a material conditional, with models corresponding to those for 'p & q,' 'not-p & q,' and 'not-p & not-q.' Thus two competing predictions can be derived from the mental model theory of Johnson-Laird and Byrne (1991; see also Johnson-Laird, Legrenzi, Girotto, Legrenzi and Caverni, 1999). Many people, with only an initial model for 'if p then q,' would judge P(if p then q) to be P(p & q). We will call this claim the *conjunctive probability hypothesis*. Other people, with fully explicit models, would judge P(if p then q) to be P(p & q) + P(not-p & q) + P(not-p & not-q). We will call this claim the *material conditional hypothesis*.

Johnson-Laird and Byrne (2002), in the most recent version of mental model theory for conditionals, continue to argue that basic conditionals have a 'core meaning' that corresponds to the material conditional. They do not, however, make a precise prediction about people's probability judgments for non-basic conditionals like (1). They do continue to imply the conjunctive probability hypothesis: that the initial mental model of a conditional, whether basic or non-basic, is 'p & q,' with an indication that there are other, implicit models. But they also hold that the specific meaning of a non-basic conditional can sometimes 'modulate' its core meaning, in a process they term 'semantic modulation,' and make it about a relation of some kind, such as a temporal or spatial relation (Johnson-Laird & Byrne, 2002, p. 673). They do not say what the subjective probability of a non-basic conditional will be when it is about a temporal, spatial, or other relation (Evans et al., 2005). Of course, such a conditional would be stronger than the mere disjunction 'not-p or q.' It would be about a relation between p and q, and its probability would not be P(not-p or q).

Johnson-Laird and Byrne (2002, p.655) do discuss an example of a causal conditional, 'if a patient has malaria then she has a fever.' But the fully explicit models that they present for it make it equivalent to the disjunction, 'the patient does not have malaria or she has fever,' which is the material conditional. Since people have the ability, according to Johnson-Laird and Byrne (2002), to make mental models fully explicit, their position seems to be that people's probability judgments about some causal conditionals will conform to the material conditional hypothesis. In other cases, semantic modulation would make the conditional about a relation and so stronger than the material conditional. Our experimental technique can be used to test whether some causal conditionals are treated as material conditionals. In general, it enables us to test the conditional probability, conjunctive probability, and material conditional hypotheses for non-basic conditionals.

1.3. Correlation, causation, and the delta-p rule

There are two additional proposals that could make more definite predictions about the results of our experiments. These are suggested by some philosophical analyses (Bennett, 2003), and by what Johnson-Laird and Byrne (2002) say about semantic modulation. The first is that conditionals like (1) are stronger than an expression that P(q|p) is high: what is additionally expressed is the belief that p *raises* the probability of q. To illustrate, consider:

(3) A petrol price increase will raise the probability of an improvement in traffic congestion.

The first proposal is that (1) has the same meaning as (3). The probability of (1) would then be determined by more than the conditional probability. For (1) would state outright that p will raise the probability of q. The extent to which p raises the probability of q, or the degree of covariation between p and q, can be measured by the *delta-p rule* (Allan, 1980). In our terminology, the delta-p rule states that this degree is given by P(q|p) – P(q|not-p). Of course, when P(q|p) – P(q|not-p) is positive, P(q|p) is greater than P(q). Just as we can derive an implicit P(q|p) for each participant from judgments about P(p & q) and P(p & not-q), we can derive an implicit P(q|not-p) from judgments about P(not-p & q) and P(not-p & not-q). From the implicit P(q|p) and P(q|not-p), we get an implicit P(q|p) – P(q|not-p) and can see whether this relates to an explicit judgment about P(if p then q). A relation to the delta-p rule would be predicted if a conditional like (1) had the same meaning as (3). It could also be predicted on the basis of Johnson-Laird and Byrne (2002) if the relation introduced by semantic modulation was one that implied a correlation between p and q. We will call this prediction the *covariation proposal*.

The second possible proposal is stronger still, and we will call it the *causal proposal*. According to it, conditionals like (1) state outright that p causes q. As we have already noted, the conditionals studied in this paper could be called 'causal conditionals.' We use this label for conditionals that are at least justified by evidence of a causal relation between p and q. These conditionals could be justified in this way even if they were only material conditionals or just expressed conditional probability judgments. It is an open question, to be tested experimentally, whether they should be called 'causal' in a stronger sense. One stronger sense would be that they express a correlation judgment, as in the covariation proposal. But our causal proposal is stronger still: that these conditionals are semantically equivalent to the statement that p causes q. Consider:

(4) A petrol price increase will cause traffic congestion to improve.

Under the causal proposal, (1) is not only justified by evidence for (4), but (1) has the same meaning as (4). This proposal could also be made on the basis of Johnson-Laird and Byrne (2002) if the relation introduced by semantic modulation was one of causation between p and q. Another possibility is that semantic modulation introduces a model of a causal mechanism connecting p and q (Johnson-Laird, personal communication). Under the causal proposal, (1) is stronger than an expression that P(q|p) is high *and* stronger than the statement that p raises the probability of q. For (4) can be false when both P(q|p) is high and P(q|p) – P(q|not-p) is high. There could, for example, be a spurious correlation between p and q. Nevertheless, there cannot be a causal relation between p and q if P(q|p) = P(q|not-p), for p and q would then be statistically independent. In recent analyses of causal strength (Cheng, 1997; Pearl, 2000), P(q|p) – P(q|not-p) > 0 is a necessary condition for p to be a cause of q.

Both the covariation proposal and the causal proposal imply that the probability of a conditional like (1) will be related to the delta-p rule. Both proposals imply that P(if p then q) will be positively influenced by P(q|p) but negatively influenced by P(q|not-p). We will call this prediction the *delta-p rule hypothesis*.

1.4. The experiments

The antecedents and consequents of the conditionals in Experiments 1 and 2 could be used to make predictions about current events. Participants were instructed to estimate the probability that the events named would occur in the UK in the next 10 years. For experimental completeness within our type of conditional, 'if p then q,' we sought to cover all four cases in which the antecedent probabilities, P(p), and the consequent probabilities, P(q), were either high or low. We will refer to these four permutations using HH, HL, LH, and LL, which are, respectively, cases in which P(p) is high and P(q) is high, P(p) is high and P(q) is low, P(p) is low and P(q) is high, and P(p) is low and P(q) is low. (See Oaksford, Chater, & Grainger, 1999; Oaksford et al., 2000, on the first use of the method of getting prior ratings of P(p) and P(q), and on why these four cases are also of theoretical interest.) Conditionals were selected in each of the cases on the basis of pre-test ratings from an initially larger set. The participants who did the pre-test ratings ($N = 21$) were different from those used in the experiment proper, but were drawn from the same population. They were never shown conditional statements as such, but rated 42 antecedent probabilities, P(p), and 42 consequent probabilities, P(q), drawn from 42 conditional statements devised by the authors. These were randomly ordered with the proviso that an antecedent and a consequent from the same conditional were presented at different points in the list. On the basis of this pre-test, 32 conditionals were selected for use in the experiment proper (see Table 1). The experiments were conducted before May 2002 at which time none of the specified antecedent events had occurred.

In Experiment 3, we evaluate our hypotheses using a type of counterfactual conditional that is closely related to the causal indicative conditionals we study in Experiments 1 and 2, differing perhaps only in time orientation. The counterfactuals referred to events that could have, but did not in fact, occur in the last 5 years. There is an extensive literature on counterfactuals in judgment and decision making and social psychology (Byrne, 2005; Mandel, Catellani, & Hilton, 2005; Roese, 2005). But this does not include analyses of people's subjective probability judgments about conditionals. To our knowledge, our experiments are the first on people's probability judgments about counterfactuals.

2. Experiment 1

The object of Experiment 1 was to investigate four hypotheses about the probability of a causal conditional: as conjunctive probability, as the material conditional, as conditional probability, and as conforming to the delta-p rule. As explained above, Johnson-Laird and Byrne (1991, 2002) imply the conjunctive probability hypothesis. Johnson-Laird and Byrne (1991), and some parts of Johnson-Laird and Byrne (2002), as well as some philosophers, imply the material conditional hypothesis. Rips and Marcus (1977) and a number of recent psychologists, along with some other philosophers, imply the conditional probability hypothesis. The covariation and causal proposals imply conforming to the delta-p rule. To simplify our symbolism, let pq be 'p & q,' p¬q be 'p & not-q,' and so on. Then the four hypotheses make different predictions about what will affect the probability of a conditional like (1) in natural language:

Conjunctive probability: $P(pq)$
Material conditional: $P(MC) = 1 - P(p\neg q)$

Table 1
Indicative conditionals used in Experiments 1 and 2, together with pre-test ratings

	P(p)	P(q)
HH conditionals		
If Adidas get more superstars to wear their new football boots then the sales of these boots will increase	0.72	0.59
If the government takes a stricter approach towards individuals refusing employment then unemployment in this country will increase	0.66	0.49
≠If the American economy continues to grow then the European economy will continue to grow	0.71	0.50
If Labour wins the next election then proportional representation will be introduced for parliamentary elections	0.61	0.49
If nurses' salaries are improved then the recruitment of nurses will increase	0.68	0.64
If fertility treatment improves then the world population will rise	0.71	0.83
If car ownership increases then traffic congestion will get worse	0.66	0.84
If computers become more powerful then economic forecasting will improve	0.91	0.56
HL conditionals		
If the cost of petrol increases then traffic congestion will improve	0.89	0.20
If more doctors are trained then NHS waiting lists will increase	0.57	0.38
If immigration laws are made stricter then the number of immigrants in the UK will increase	0.64	0.40
If computer technology continues to develop then unemployment will increase	0.92	0.42
If more people use protective sun creams when sunbathing then the number of deaths due to skin cancer will be reduced	0.67	0.46
If student fees are increased then applications for university places will drop	0.58	0.27
If global warming continues then London will be flooded	0.75	0.18
If the Conservatives change their leader then they will win the next election	0.68	0.35
LH conditionals		
If NHS waiting lists increase then Labour will win the next election	0.38	0.61
If student grants are brought back then university entries will increase	0.24	0.63
If divorce is made more difficult then the number of marriages will increase	0.29	0.63
If European countries maintain current quotas of fish from the North Sea then many species of fish will be endangered	0.45	0.63
If primary school class sizes are reduced then national literacy will improve	0.46	0.58
If unemployment drops then Labour will win the next election	0.49	0.61
If more people take up smoking then tax on cigarettes will be increased	0.42	0.84
If trial by jury is abandoned then convictions of criminals will increase	0.21	0.58
LL conditionals		
If capital punishment is brought back then the murder rate in the UK will increase	0.08	0.32
If governments across the world prosecute tobacco companies then smoking related diseases will increase	0.39	0.38
If less violence is shown on television then the amount of violent crime will be reduced	0.23	0.38

(*continued on next page*)

Table 1 (*continued*)

	P(p)	P(q)
If nationalism in the world declines then the number of conflicts in the world will be reduced	0.33	0.27
If the Conservatives win the next election then Britain will leave the European Community	0.35	0.24
If a cure for AIDS is found then sales of condoms will drop	0.45	0.22
If Saddam Hussein is assassinated then Iraq will become a democracy	0.36	0.32
If Tim Henman wins Wimbledon then he will receive a knighthood	0.41	0.21

Note. ≠In Experiment 2 this statement was substituted by 'If the American economy recovers then the European economy will recover.'

Conditional probability: P(q|p)
Delta-p rule: P(q|p) − P(q|¬p)

In order to derive the above probabilities, we determined each participant's judged probability of the four truth table cases, TT, TF, FT, and FF, for each conditional. The participants in Experiment 1 were required to fill out three booklets with probability ratings for the 32 conditionals shown in Table 1. One booklet required judgments of probability for the four truth table cases, one to judge the probability that the whole conditional was true, and one to judge the probability that the whole conditional was false. Experiment 1 consequently provides the opportunity to examine whether judgments of the probability of truth and the probability of falsity are complementary (sum to 1).

2.1. Method

2.1.1. Participants

Forty students of the University of Plymouth volunteered to take part in return for a small fee. Participants were tested in small groups.

2.1.2. Materials and procedure

A repeated measures design was employed with each participant completing the three booklets. Each of these booklets contained items derived from the same 32 conditionals (see Table 1). Direct estimations of probabilities as percentages were used throughout. In the truth table task, participants were asked to judge the probability that the specified events would occur in the UK within the next 10 years. The cases were grouped together and a further instruction was that the scores given must total 100. For example, corresponding to conditional (1) above, participants were asked to rate four conjunctions corresponding to the four rows of a truth table, TT, TF, FT, and FF:

The cost of petrol increases and traffic congestion improves. ____
The cost of petrol increases and traffic congestion does not improve. ____
The cost of petrol does not increase and traffic congestion improves. ____
The cost of petrol does not increase and traffic congestion does not improve. ____
100%

The other two booklets asked participants to rate between 0 and 100% the probability that the conditional statements given were true or false. Across the booklets, different random presentation orders of items were employed. Order of presentation of the three booklets was randomized.

2.2. Results and discussion

2.2.1. Truth table task

The mean probability estimates for truth table cases are shown in Table 2 and are broken down by the prior classification of conditionals in terms of component probabilities: HH, HL, LH, and LL. The purpose of this task was to provide relevant probability calculations to test our contrasting theories about the perceived probability of conditional statements, reported below. However, as a manipulation check, we first performed analyses of variance to see whether the effect of our prior probability judgments of the antecedents and consequents was reflected in the judged probability of the four truth table cases. For example, when assessing the probability of a ‘p & not-q’ conjunction, a TF case, we would expect a higher estimate when the antecedent of the associated conditional had been given a high prior probability, P(p), and when the consequent of the associated conditional had been given a low prior probability, P(q). Here and throughout, ANOVAs were computed both across participants (reported as F_1) and across sentences (reported as F_2). Note

Table 2
Mean (*SD*) ratings of the four truth table cases in Experiment 1

TT	Consequent probability	
	High	Low
Antecedent probability		
High	45.3 (13.3)	39.2 (15.5)
Low	33.3 (11.6)	24.4 (12.3)
TF	Consequent probability	
	High	Low
Antecedent probability		
High	18.3 (8.7)	24.8 (10.2)
Low	14.6 (5.3)	20.4 (6.3)
FT	Consequent probability	
	High	Low
Antecedent probability		
High	13.5 (7.2)	10.4 (5.4)
Low	23.8 (8.2)	16.0 (7.2)
FF	Consequent probability	
	High	Low
Antecedent probability		
High	20.7 (9.2)	23.0 (10.2)
Low	27.7 (10.7)	38.7 (12.8)

that the analyses of the four cases are not independent, as participants were required to make estimates sum to 100%.

All ANOVAs found the expected main effects, confirming our prior classification of sentences. TT ratings were higher when antecedent probability was high ($F_1(1,39) = 46.5$, MSE = 154, $p < .001$; $F_2(1,28) = 20.92$, MSE = 75.90, $p < .001$) and when the consequent probability was high ($F_1(1,39) = 22.2$, MSE = 100, $p < .001$; $F_2(1,28) = 6.63$, MSE = 75.90, $p < .05$). TF ratings were higher when antecedent probability was high ($F_1(1,39) = 10.1$, MSE = 65.1, $p < .005$; $F_2(1,28) = 4.03$, MSE = 45.21, $p = .054$) and when consequent probability was low ($F_1(1,39) = 43.1$, MSE = 34.7, $p < .001$; $F_2(1,28) = 7.13$, MSE = 45.21, $p < .05$). FT ratings were higher when antecedent probability was low ($F_1(1,39) = 79.3$, MSE = 32.1, $p < .001$; $F_2(1,28) = 22.81$, MSE = 21.20, $p < .001$) and when consequent probability was high ($F_1(1,39) = 25.1$, MSE = 47.1, $p < .001$; $F_2(1,28) = 12.63$, MSE = 21.20, $p < .001$). Finally, FF ratings were higher when antecedent probability was low ($F_1(1,39) = 39.9$, MSE = 129, $p < .001$; $F_2(1,28) = 34.52$, MSE = 29.01, $p < .001$) and when consequent probability was low ($F_1(1,39) = 29.3$, MSE = 60.1, $p < .001$; $F_2(1,28) = 12.03$, MSE = 29.01, $p < .005$).

Significant interactions between the two factors were observed for FF ratings ($F_1(1,39) = 12.8$, MSE = 58.6, $p < .001$; $F_2(1,28) = 5.54$, MSE = 29.01, $p < .05$) and for FT ratings but only across participants ($F_1(1,39) = 7.6$, MSE = 28.8, $p < .01$; $F_2(1,28) = 2.17$, MSE = 21.20, $p = .15$). In both cases, the effect of consequent probability was more marked when the antecedent probability was low.

2.2.2. *Probability of conditional task*

The mean probability estimates for each of the probability of conditional tasks are shown in Table 3 and are broken down by the prior classification of conditionals in terms of component probabilities: HH, HL, LH, and LL. For each case, we performed two 2 (Antecedent probability) by 2 (Consequent probability) analyses of variance, one across participants and one across items, to examine the effect of antecedent and consequent probabilities. Participants rated conditionals as more probably true when the antecedent probability was high ($F_1(1,39) = 13.9$, MSE = 103, $p < .001$; $F_2(1,28) = 2.21$, MSE = 58.96, $p = .148$) and when the consequent probability was high ($F_1(1,39) = 35.5$, MSE = 103, $p < .001$; $F_2(1,28) = 5.09$, MSE = 58.96, $p < .05$). Similarly, they rated conditionals as more probably false when the antecedent probability was low ($F_1(1,39) = 9.3$, MSE = 103, $p < .005$; $F_2(1,28) = 1.99$, MSE = 105.75, $p = .17$) and when

Table 3
Mean (*SD*) TRUE and FALSE probability of conditional ratings in Experiment 1

TRUE	Consequent probability	
	High	Low
Antecedent probability		
High	60.7 (16.4)	52.2 (16.4)
Low	55.7 (15.5)	45.2 (15.2)
FALSE	Consequent probability	
	High	Low
Antecedent probability		
High	34.1 (17.0)	37.7 (15.2)
Low	36.0 (13.4)	45.5 (13.0)

the consequent probability was low ($F_1(1,39) = 14.1$, MSE = 122, $p < .001$; $F_2(1,28) = 3.75$, MSE = 105.75, $p = .063$). The interaction between these two factors was not significant for true ratings and only marginally significant for false ratings across participants ($F_1(1,39) = 4.0$, MSE = 85.2, $p = .053$; $F_2(1,28) < 1$). The trend is for conditionals low in both antecedent and consequent probability to be rated as less probable than other kinds of conditionals. We defer interpretation of these analyses in Section 5.

2.2.3. Correlation and regression analyses

The analyses of most theoretical interest concern the relation between probabilities computed from the truth table task and the perceived probability of the whole conditional statement. We considered four alternative hypotheses: the material conditional, conditional probability, conjunctive probability, and the delta-p rule. We computed their predictions from the truth table case estimates, as follows.

Conjunctive probability: P(pq) = P(TT)
Material conditional: P(MC) = P(TT) + P(FT) + P(FF)
Conditional probability: P(q|p) = P(TT)/[P(TT) + P(TF)]
Delta-p rule: P(q|p) − P(q|¬p) = P(q|p) − [P(FT)/[P(FT) + P(FF)]

We analyzed the relation between these probabilities and the probability of conditionals as measured across problem materials (conditional sentences). We first did this by computing the mean of all participants for each sentence and then calculating correlations and regressions on these mean scores (see Table 4). Looking first at the correlations for the four predictions in Table 4(a), we see that all are statistically significant, but the highest correlations are with conditional probability and the lowest with the probability of the material conditional. (In general, judgments of probable falsity reflect those of probable truth.) We then tested for differences among these critical correlations by using the procedure outlined in Meng, Rosenthal, and Rubin (1992). The correlation with conditional probability was significantly higher than those with any of the other three predictors

Table 4
Correlations and regressions computed across sentences using mean ratings of all participants in Experiment 1

	TRUE	FALSE
(a) Correlations between judged probability of conditionals as true or false and various probabilities computed from the truth table task		
Hypotheses		
Conjunctive probability	.74*	−.71*
Material conditional	.65*	−.68*
Conditional probability	.89*	−.89*
Delta-p rule	.72*	−.74*
Other		
P(p)	.29	−.25
P(q)	.73*	−.71*
P(q/¬p)	.10	−.08
(b) Multiple regressions using three statistically independent predictors (β weights)		
P(p)	.05	.00
P(q|p)	.90*	−.93*
P(q|¬p)	−.11	.14

(for all comparisons $p < .05$). That is, the conditional probability hypothesis captured the data significantly better than any of the other three hypotheses.

A problem with our four hypotheses is that they are themselves highly correlated (there is a high degree of multicollinearity) making it hard to detect whether their contribution is significant. To get a better idea about the relevant merits of each hypothesis we performed a multiple linear regression analysis with judged probability of the conditional as the dependent measure and three *statistically independently measured* predictors: P(p), P(q|p) and P(q|¬p).[1] The observed multicollinearity between these predictors was minimal. Across our three experiments tolerance values always exceeded .78, which means that the multiple R^2 of a predictor regressed on the other two never exceeded .22. The four hypotheses make different predictions for the probable truth of the conditional:

Conjunctive probability: Both P(p) and P(q|p) should be positive predictors since their product equals P(pq).
Material conditional: P(p) should be a negative predictor since the material conditional is true when we have not-p.
Conditional probability: P(q|p) should be the only predictor of judgments.
Delta-p rule: P(q|p) should be a positive predictor and P(q|¬p) an equally large negative predictor as their difference measures the degree of correlation between p and q.

The results of this analysis are presented in Table 4(b) and provide further strong support for the conditional probability hypothesis, since P(q|p) have β weights approaching 1 (minus 1 for false judgments) and neither of the other predictors differs significantly from 0. However, we are aware that averaging across participants might conceal important individual differences. Hence we repeated these regression analyses for each individual participant. The results are summarized in Table 5 where we show the mean regression weights across individual participants together with a one-sample t-test to decide whether these differ from zero (test suggested by Lorch & Myers, 1990). These findings confirm strong support for the conditional probability hypothesis since the mean β weights for P(q|p) are very substantial (.42 true, −.38 false) compared with those for the other predictors. However, a small effect of P(q|¬p) emerged (−.08 true, .07 false) in the direction expected by the delta-p rule hypothesis. This is significant for both true and false cases if a one-tailed probability is computed.

[1] An anonymous reviewer asked whether the data in our studies meet the assumptions of linear regression—*linearity* of the relationship between dependent and independent variables, *homeoscedasticity* (constant variance) of the error terms, *normality* of the error terms and *independence* (no serial correlation) of the error terms (see, e.g., Hair, Black, Babin, Anderson, & Tatham, 2006). They do. Visual inspection of studentized residual versus standardized predicted value plots for each study showed symmetrical distribution along the horizontal line. Visual inspection of related partial regression plots showed either a linear relation or no relation. Both of these observations support linearity. Visual inspection of studentized residual versus standardized predicted value plots showed homeoscedasticity of the error terms (nice cloudy shapes). Levene's tests for homogeneity of variance were not significant (for all $p > .10$), further supporting homeoscedasticity. We tested for normality by inspecting normal probability plots. In each case the residuals fell close to the diagonal line. For each study, we also carried out tests to examine whether the kurtosis and skeweness were significantly different than 0 – they were not. As for the independence assumption, this was not a concern in the present studies because participants received the various items in different random orders. Moreover, in each study the *Durbin-Watson statistic* was close to 2, which means no serial correlation in the residuals.

Table 5
Multiple regression analyses carried out on individual participants across sentences for three predictors in Experiments 1 and 2

	TRUE			FALSE		
	Mean (*SD*) β weights	*t*(38)	*p*	Mean (*SD*) β weights	*t*(38)	*p*
(a) Experiment 1						
P(p)	.02 (.19)	.59	ns	.02 (.20)	.61	ns
P(q|p)	.42 (.23)	11.30	<.001	−.38 (.28)	−.8.29	<.001
P(q|¬p)	−.08(.18)	−2.84	.007	.07(.23)	1.95	.059
	Probability of conditional			Causal strength		
	Mean (*SD*) β weights	*t*(40)	*p*	Mean (*SD*) β weights	*t*(40)	*p*
(b) Experiment 2						
P(p)	.16 (.18)	5.51	<.001	.14 (.17)	5.11	<.001
P(q|p)	.51 (.21)	15.77	<.001	.50 (.24)	13.24	<.001
P(q|¬p)	−.015 (.20)	−.51	ns	−.03 (.25)	−.83	ns
	Probability of conditional			Causal strength		
	Mean (*SD*) β weights	*t*(25)	*p*	Mean (*SD*) β weights	*t*(25)	*p*
(c) Experiment 3						
P(p)	−.04 (.20)	−1.07	ns	.05 (.25)	.94	ns
P(q|p)	.42 (.23)	9.33	<.001	.39 (.26)	7.79	<.001
P(q|¬p)	−.04 (.18)	−1.21	ns	−.11 (.21)	−2.65	<.05

2.2.4. Conditionals always true or false?

Some philosophical logicians have argued that indicative conditionals are neither true nor false when their antecedents are false (Adams, 1975, 1998; Bennett, 2003; Edgington, 1995). The claim is that a conditional 'if p then q' with a false antecedent p creates a 'truth value gap.' The antecedent p and consequent q can be 'factual' statements that are objectively true or objectively false, but the conditional itself is not 'factual' (Adams, 1998). When the antecedent is false, the conditional is held to be neither true nor false objectively, but people have a subjective confidence in it that equals P(q|p). A conditional of this type can be called an Adams conditional (Evans & Over, 2004).

Fig. 1 displays stacked bar charts of true and false ratings of conditionals. Judgments summed to around 93% overall. This was shown to be significantly less than 100%, using a

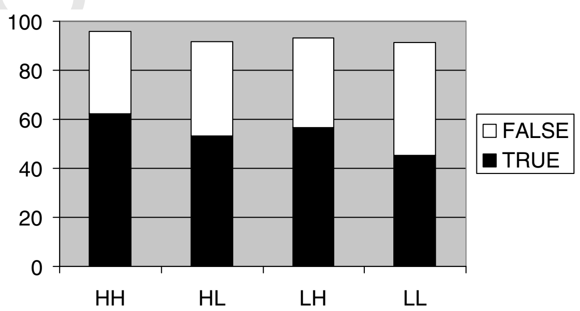

Fig. 1. Stacked TRUE and FALSE ratings of conditionals in Experiment 1.

one-sample *t*-test to compare mean true plus false judgments for each participant with the null hypothesis ($t(39) = 2.50$, $p < .01$). An analysis of variance carried out on total true plus false ratings produced a significant main effect for consequent probability ($F(1,39) = 4.28$, MSE = 85.2, $p < .05$). High consequent probability items (HH and LH) received higher estimates than low consequent probability items (HL and LL). This violates the prediction of binary complementarity in support theory (Tversky & Koehler, 1994) for judgments of event probability (for further violations of this prediction see Macchi, Osherson, & Krantz, 1999). The failure of complementarity is driven by a weak trend for scores to be lower when the consequent probabilities were low (HL and LL cases; see Fig. 1). Supposing there are truth value gaps for conditionals, why is there not a more general failure of complementarity?

The question is difficult to answer if one holds *both* that there is a truth value gap when a conditional has false antecedent, *and* that 'true' and 'false' always have the same strict meanings in natural language. In that case, there should be a significant 'gap' between 1 and the sum of the probability that a conditional with false antecedent is true and of the probability that it is false. This would have the effect of generally reducing true plus false ratings. However, supporters of the view that conditionals in natural language are Adams conditionals could simply take the result as empirical evidence that 'true' and 'false' are ambiguous in natural language. They could argue that 'true' and 'false' do not have the same meaning when applied to a conditional and to its antecedent or consequent in our examples. In their view, to say that the antecedent or consequent is 'probably true' is to say that it probably corresponds to an objective fact, but to say that the conditional itself is 'probably true' is just to say it is subjectively probable. The latter is a 'pleonastic' use of 'true,' and there does not have to be a truth value gap for that (Edgington, 2003). Followers of Adams could argue that this ambiguity in the ordinary use of 'true' makes truth value gaps for conditionals hard to discover in experiments. We will return to the Adams conditional below, in Section 5.

2.2.5. Conclusions

Experiment 1 has provided powerful evidence for the conditional probability hypothesis. In previously published studies by Evans et al. (2003), Oberauer and Wilhelm (2003), and Over and Evans (2003), it was shown that the judged probability of conditionals was determined by P(q|p) for many participants when abstract, basic conditionals were used with specified probability distributions. The findings of Experiment 1 are stronger than these, in that the conditional probabilities were derived from prior beliefs about relevant possibilities without presenting frequencies, and predict very closely the judged probability of non-basic indicative conditionals. As in earlier studies, there was no evidence at all for the material conditional hypothesis. Moreover, unlike the earlier studies, the conditional probability is the overwhelming factor in these judgments. There was no evidence of a subgroup of people who base their judgments on the conjunctive probability. We will consider what our results imply about the delta-p rule in Section 5.

3. Experiment 2

One objective of Experiment 2 was to replicate the findings of Experiment 1 with respect to the conditional probability hypothesis. To that end, we again used the truth table task and the true conditional task (the task that asked participants to judge the probability that a conditional is true). Items in Experiment 2 were derived from the same conditionals as items in Experiment 1 (see Table 1), with one exception. The conditional, 'If the American economy continues to grow then the European economy will improve,' was replaced by the conditional, 'If the American economy recovers then the European economy will recover.' At the time we conducted Experiment 2, the American economy was in a recession.

A second objective of Experiment 2 was to compare people's judgments of the probability of a conditional with their judgments of the strength of a causal relation between p and q. To that end, Experiment 2 included a causal strength task. Participants were presented with each one of the 32 conditionals and were asked to rate the causal strength of the relation between p and q in each case using a 5-point scale. Higher ratings indicated higher causal strength estimates. The negative effects of P(q|¬p) were very limited in Experiment 1. It could be argued that this is because participants were asked to judge the probability that the conditionals were true rather than the strength of the causal relation. The delta-p rule for measuring the degree of correlation between p and q takes the difference between P(q|p) and P(q|¬p), P(q|p) – P(q|¬p). A positive difference here is necessary if p is a cause of q (Cheng, 1997; Pearl, 2000). It might be that a stronger link would be found with the delta-p rule if participants were asked to rate the causal strength between p and q. Experiment 2 provides a test of this hypothesis.

3.1. Method

3.1.1. Participants

Forty-two students of the University of Plymouth volunteered to take part in return for a small fee. Participants were tested in small groups.

3.1.2. Materials and procedure

The materials and procedure were similar to those of Experiment 1, with the exception that the probability that a conditional is false task was substituted by the causal strength task. A repeated measures design was employed with each participant completing the three similar booklets. The presentation order of the statements within each booklet was randomized. The presentation order of the three booklets was counterbalanced, such that seven participants were assigned to each of the six possible presentation orders.

3.2. Results and discussion

3.2.1. Truth table task

The mean probability estimates for truth table cases are shown in Table 6 and are broken down by the prior classification of conditionals in terms of component probabilities, HH, HL, LH, and LL. For each case, we performed two 2 (Antecedent probability) by 2 (Consequent probability) analyses of variance, one across participants and one across items, to examine the effect of antecedent and consequent probabilities. Table 7 shows that, as in Experiment 1, in each case the trends essentially confirm the prior rating of the pre-test. TT ratings were higher when antecedent probability was high ($F_1(1,41) = 62.11$, $p < .001$, MSE = 41; $F_2(1,28) = 15.12$, $p < .01$, MSE = 93.98) and when the consequent probability was high ($F_1(1,41) = 43.88$,

Table 6
Mean (*SD*) ratings of the four truth table cases in Experiment 2

TT	Consequent probability	
	High	Low
Antecedent probability		
High	47.87 (12.26)	43.64 (10.86)
Low	38.01 (14.23)	26.85 (12.45)
TF	**Consequent probability**	
	High	Low
Antecedent probability		
High	17.30 (6.91)	24.90 (9.05)
Low	14.90 (5.97)	24.56 (10.97)
FT	**Consequent probability**	
	High	Low
Antecedent probability		
High	14.52 (5.24)	9.93 (5.38)
Low	22.36 (10.94)	12.29 (6.88)
FF	**Consequent probability**	
	High	Low
Antecedent probability		
High	20.31 (8.65)	21.57 (8.72)
Low	24.74 (10.36)	36.30 (15.55)

Table 7

(a) Mean (SD) probability of conditional ratings in Experiment 2	Consequent probability	
	High	Low
Antecedent probability		
High	61.81 (11.31)	55.75 (14.00)
Low	58.16 (13.95)	45.06 (13.60)
(b) Mean (SD) causal strength ratings in Experiment 2	Consequent probability	
	High	Low
Antecedent probability		
High	3.54 (.46)	3.43 (.45)
Low	3.43 (.40)	2.82 (.54)

$p < .001$, MSE = 56.59; $F_2(1,28) = 5.03$, $p < .05$, MSE = 93.98). TF ratings were higher when the consequent probability was low ($F_1(1,41) = 60.41$, $p < .001$, MSE = 51.79; $F_2(1,28) = 10.97$, $p < .005$, MSE = 54.31), but they were unaffected by antecedent probability (both Fs < 1). FT ratings were higher when antecedent probability was low ($F_1(1,41) = 24.95$, $p < .001$, MSE = 43.69; $F_2(1,28) = 9.56$, $p < .005$, MSE = 21.73) and when the consequent probability was high ($F_1(1,41) = 54.12$, $p < .001$, MSE = 41.71; $F_2(1,28) = 19.79$, $p < .001$, MSE = 21.73). FF ratings were higher when the antecedent probability was low ($F_1(1,41) = 30.14$, $p < .001$, MSE = 127.83; $F_2(1,28) = 24.53$, $p < .001$, MSE = 29.86) and when the consequent probability was low ($F_1(1,41) = 28.42$, $p < .001$, MSE = 60.77; $F_2(1,28) = 11.05$, $p < .005$, MSE = 29.86).

Significant interactions between these two factors were observed for TT, FT, and FF ratings across participants (TT: $F_1(1,41) = 11.68$, MSE = 43.09, $p < .001$; FT: $F_1(1,41) = 5.14$, MSE = 61.39, $p < .05$; FF: $F_1(1,41) = 16.61$, MSE = 67.03, $p < .001$). Across items, however, the interaction was significant only for FF ratings (TT: $F_2(1,28) = 1.02$; FT: $F_2(1,28) = 2.76$, MSE = 93.98, $p > .10$; FF: $F_2(1,28) = 7.08$, MSE = 29.86, $p < .05$). In all cases, the effect of consequent probability was more marked when antecedent probability was low.

3.2.2. Probability of conditional task

Order did not influence judgments ($F_1(5,36) = 1.02$, ns) and therefore was dropped from subsequent analyses. The mean probability estimates for the conditionals are shown in Table 7(a) and are broken down by the classification of conditionals in terms of component probabilities. Probability ratings were higher when the antecedent probability was high ($F_1(1,41) = 68.93$, MSE = 31.30, $p < .001$; $F_2(1,28) = 2.31$, $p = .14$) and when the consequent probability was high ($F_1(1,41) = 49.56$, MSE = 41.00, $p < .001$; $F_2(1,28) = 4.13$, MSE = 177, $p < .06$). An interaction was shown such that consequent probability affected more judgments for low antecedent probability conditionals but only for participants ($F_1(1,41) = 5.92$, $p < .05$, MSE = 67.03; $F_2 < 1$).

3.2.3. Causal strength task

Order did not influence causal strength judgments (Fs < 1) and was dropped from subsequent analyses. The mean causal strength estimates for the relations suggested by the conditional statements are shown in Table 7(b) and are broken down by the classification of conditionals in terms of component probabilities. Causal strength ratings were higher when the antecedent probability was high ($F_1(1,41) = 52.62$, MSE = .10, $p < .001$; $F_2(1,28) = 3.20$, MSE = .32, $p < .10$) and when the consequent probability was high ($F_1(1,41) = 48.20$, MSE = .11, $p < .001$; $F_2(1,28) = 3.31$, MSE = .32, $p < .10$). An interaction was shown such that consequent probability affected more judgments for low antecedent probability conditionals but only for participants ($F_1(1,41) = 15.49$, $p < .001$, MSE = .17; but $F_2(1,28) = 1.56$, ns). An inspection of Table 7(b) suggests that these effects are due to that LL conditionals were judged to involve weaker causal links than the other types of conditionals. These findings are very similar to those for the probability judgments reported above.

3.2.4. Correlation and regression analyses

As in Experiment 1, we considered four alternative hypotheses: the material conditional, conditional probability, conjunctive probability, and the delta-p rule. Scores for each of these predictions were derived in the same way as in Experiment 1. These scores when then correlated across sentences with mean probability ratings of the whole conditional and with mean causal strength ratings (Table 8(a)). As in Experiment 1, all the correlations

Table 8
Correlations and regressions computed across sentences using mean ratings of all participants in Experiment 2

	Probability of conditional	Causal strength
(a) Correlations between judged probability of conditionals and judged causal strength with various probabilities computed from the truth table task		
Hypotheses		
Conjunctive probability	.88*	.86*
Material conditional	.77*	.71*
Conditional probability	.91*	.87*
Delta-p rule	.72*	.73*
Other		
P(p)	.46*	.49*
P(q)	.77*	.73*
P(q/¬p)	.11	.07
(b) Multiple regressions using three statistically independent predictors (β weights)		
P(p)	.14*	.19*
P(q|p)	.93*	.88*
P(q|¬p)	−.20*	−.23*

are statistically significant, but the highest correlations are with conditional probability. The correlations with conditional probability were significantly higher than those with the material conditional or with the delta-p rule but no different than those with the conjunctive probability. We also obtained a near perfect correlation between probability of the whole conditional and causal strength ($r = .98$, $p < .001$)[2] We will consider what this result implies in Section 5.

We again performed a multiple linear regression with three independent predictors as shown in Table 8(b). For both probability and causal strength judgments there was a large and significant effect of conditional probability P(q|p). In contrast to Experiment 1, there was also a significant negative influence of P(q|¬p), as predicted by the delta-p rule hypothesis. However, the size of the β weights was a lot smaller than for P(q|p).

We ran regression analyses, as in Experiment 1, on the data of individual participants and this is summarized in Table 5(b). As in Experiment 1, we obtained strong support for the conditional probability hypothesis but, in contrast to that experiment, the secondary effect that shows up is of P(p) rather than P(q|¬p). This finding provides only weak support for the conjunctive probability hypothesis since it was not supported in the corresponding regression analysis on item means in Experiment 2 nor in any regression analysis in Experiment 1. There was yet again no support at all for the material conditional hypothesis.

[2] An anonymous reviewer suggested that this high correlation could be due to carry-over effects because of the close proximity of the two judgments. The design of Experiment 2 allowed us to test this possibility. In that study we counterbalanced the order of presentation of the three tasks (truth table; probability of conditional; causal strength). There were seven participants assigned in each of the six possible orders of these tasks. In two of these orders the probability of conditional task came first, and in another two the causal strength task came first. In the former case probability of conditional estimates are 'clean'; in the latter case causal strength estimates are 'clean.' Therefore, we calculated the correlation (on item means) between 'clean' probability of conditional judgments and 'clean' causal strength judgments. The resulting correlation was highly significant; $r = .87$, $p < .0001$, a finding that rules out the carry-over effect hypothesis.

3.2.5. Conclusions

Experiment 2 has replicated the major finding of Experiment 1 in support of the conditional probability hypothesis. Judgments of the probability of non-basic indicative conditionals were most strongly related to the conditional probability P(q|p), derived from participants' beliefs about relevant possibilities. It is striking that in all our analyses judgments of causal strength are affected by the same factors that affect probability judgment. But the evidence shows only weak support for the delta-p rule (significant if small effect of P(q|¬p)) in regressions on mean ratings (see Table 8(b)). We will consider this point in Section 5 as well as the strong correlation between judgments of the probability of conditionals and of causal strength.

4. Experiment 3

A strong theoretical case can be made that the causal indicative conditionals we studied in Experiments 1 and 2 differ from certain counterfactuals primarily in their time perspective (Bennett, 2003; Dudman, 1984; Edgington, 1995; Evans & Over, 2004). Suppose that we asserted the indicative (1) last year, about the coming year, but that over the course of that time the cost of petrol did not increase and traffic congestion did not improve. We now know that the cost of petrol did not increase, but our beliefs about the effect of the cost of petrol on traffic congestion are unchanged. Consider the counterfactual:

(5) If the cost of petrol had increased, then traffic congestion would have improved.

How do we make a probability judgment about (5)? Stalnaker (1968) extended the Ramsey test to cases in which the antecedent p of the conditional is strongly disbelieved, as it is in (5) in our example. This extended test has been much discussed in the philosophical literature (Bennett, 2003). In the extended test, we make the smallest possible change to our current belief state in order to accommodate the supposition of p. We then make the conditional probability judgment about q given p. Using this procedure, we can make a probability judgment about (5). We start out, a year ago, not knowing whether the cost of petrol will increase, but we suppose that it will and confidently infer, under that supposition and using the original Ramsey test, that traffic congestion will improve. That gave us confidence in (1), as a statement about the coming year. The year passes, and we notice that the cost of petrol does not increase and traffic congestion does not improve. But we can now use the Ramsey test as extended by Stalnaker. We make the smallest change to our belief state to be consistent in supposing that the cost of petrol did increase during the year, and infer confidently that an improvement in traffic congestion follows. This Ramsey test gives us the same confidence in (5) today as we had in (1) one year ago.

Our justification for (5) today is the same as the justification we had for (1) last year. All that has changed is that the antecedent and consequent of (1) turned out to be false. We expressed a belief last year with (1), and express the same belief this year with (5). This close relation between conditionals like (1) and (5) suggests that people will make similar probability judgments about these indicative and counterfactual conditionals, on the basis of similar psychological processes. The prediction that these similar judgments will be made has never before been empirically tested.

There is much debate on the general relation between indicative conditionals, whether causal, epistemic, or other type, and counterfactuals (Bennett, 2003; Edgington, 1995;

Evans & Over, 2004; Sloman, 2005). Here we consider only the possibility that causal indicative conditionals like (1) and counterfactuals like (5) make common reference in the way just described. We can test this possibility by applying the four hypotheses to counterfactuals like (5). However, note that no one has ever held, as far as we know, the conjunctive probability hypothesis about counterfactuals, and almost no one has held the material conditional hypothesis about them (Bennett, 2003).

We can extend the conditional probability hypothesis to counterfactuals of the form, 'if p had been the case, then q would have been the case,' in the following way. We can propose that the subjective probability of the counterfactual at the present time is the same as the conditional probability P(q|p) at an earlier time (as given by the context). In our example, our confidence in (5) today is predicted to equal the conditional subjective probability judgment we made last year, i.e., our confidence at that time that traffic congestion was going to improve given that the cost of petrol increased.

There is a long literature in philosophy that relates counterfactuals to causation (Woodward, 2003). There has been a great deal of interest in this relationship recently in the cognitive sciences (Pearl, 2000; Sloman, 2005). There is also much work in judgment and decision making and social psychology on counterfactuals and causation (Byrne, 2005; Mandel et al., 2005; Roese, 2005). In light of this long tradition, and of the close relationship between (1) and (5), counterfactuals can also be called 'causal conditionals' in some sense. We ask in what sense this is true. Are counterfactuals 'causal' because they express conditional probability judgments that can be supported by evidence about a causal relation? Or are they 'causal' in the stronger sense of having the same meaning as a statement that a correlation exists, or the still stronger sense of having the same meaning as a statement that a causal relation exists? A proposal that they are 'causal' in one of the stronger senses implies the delta-p rule hypothesis for the reasons we gave above. As we have explained, a correlation between p and q, or a causal relation between p and q, implies that P(q|p) – P(q|¬p) is positive.

In Experiment 3, we used a similar methodology to that of Experiment 2. However, we did not manipulate antecedent and consequent probability, but instead varied the polarity of these components. We constructed 32 counterfactual conditionals (Table 9) broken down into four groups designated as AA (affirmative antecedent, affirmative consequent), AN (affirmative antecedent, negated consequent), NA and NN. We suspect that people can easily process simple negation in real world conditionals. We can just as easily imagine actions taken or not taken, events that occur or do not occur and so on (see Evans & Over, 2004, Chapter 1). Hence, Experiment 3 tests the four hypotheses across conditionals that may contain negated components.

We used statements referring to real world events with plausible causal links similar to those used in Experiments 1 and 2 (see Table 9). However, these statements now referred to events that might have occurred within the last 5 years. In all cases, the event had not actually occurred at the time of the study (or in the case of negated components, it had occurred).

4.1. Method

4.1.1. Participants

Twenty-six undergraduate students of the University of Plymouth volunteered to take part in return for a small fee.

Table 9
Counterfactual conditionals used in Experiment 3

AA
If Queen Elizabeth had died last year, then Prince Charles would have become king.
If Michael Owen had played for Arsenal last season, then Arsenal would have won the 2002/03 Premiership.
If a vaccine for AIDS had been available two years ago, then a major health crisis in Africa would have been avoided.
If Turkey had improved its human right practices in 2000, then Turkey would have been awarded membership in the European Union.
If Robbie Williams had advertized Nike last year, then Nike sales would have increased.
If an economist had come up with a theory that correctly predicted stock market fluctuations two years ago, then that economist would have been awarded the Nobel Prize.
If the Conservatives had been elected to government in the 2001 elections, then taxes would have been increased.
If car theft had been penalized with life-sentence in 2001, then companies that sell car theft systems would have gone bankrupt.

AN
If the Conservatives had won the 2001 elections, then UK would not have sent troops to Iraq.
If university admission had been easier two years ago, then secondary school students would not have studied hard for A-levels.
If more people had practiced safe sex in 2000, then there would not have been an increase in AIDS fatalities.
If David Beckham had been involved in a tax scandal two years ago, then he would not have received an OBE.
If a method to successfully predict earthquakes had been found two years ago, then there would not have been victims due to earthquakes.
If alternative fuels had become widely available five years ago, then the price of petrol would not have remained high.
If the UK had joined euro in 1999, then the UK's relations with France would not have been difficult.
If Sylvester Stallone had run for Mayor of California in 2000, then he would not have continued his career as an actor.

NA
If David Beckham had not played for Manchester United last season, then Manchester United would have been relegated.
If Michael Schumacher had not won the GP title last year, then Ferrari would have fired him.
If the British climate had not become warmer in the last two years, then British energy consumption would have increased.
If Space Shuttle Columbia had not crashed in 2003, then NASA would have sent more manned missions to the moon.
If Cherie Blair had not been married to Tony Blair, then she would have been appointed as a judge.
If England had not played against Brazil in the 2002 FIFA World Cup quarterfinals, then they would have gone through to the next round.
If Geri Hallowell had not left the Spice Girls, then the Spice Girls would still be together.
If Michael Jackson had not given an interview to Martin Bashir, then he would have produced the best-selling album in 2003.

NN
If New York had not been attacked by terrorists in 2001, then the US would not have attacked Iraq.
If the UK had not invaded Iraq in 2003, then Robin Cook would not have resigned.
If railways had not become privatized, then there would not have been an increase in train delays.
If SARS had not been deadly, then the World Health Organization would not have introduced travel sanctions in infected areas.
If London had not introduced congestion charging, then traffic congestion in central London would not have decreased.
If France continued not to support US foreign policy last year, then US companies would not have renewed contracts with France.
If unemployment had not been reduced in 1997–2001, then Labour would not have won the last elections.
If Saddam Hussein had not been Iraq's President in 2003, then the US would not have attacked Iraq.

4.1.2. Design

The experimental design was identical to that of Experiment 2.

4.1.3. Materials and procedure

The materials were constructed and presented in a similar manner to that of Experiment 2. In the probability of counterfactuals task, we asked, as before, people to judge the probability of truth of such statements. An example from our actual materials is:

> If New York had not been attacked by terrorists in 2001, then the US would not have attacked Iraq.

The other tasks needed some modification. In the causal strength and truth table tasks, we asked people to make judgments about a point in time 5 years in the past. In the causal strength task, instead of presenting the conditional sentences, we asked people to rate the causal strength of the relation between antecedent and consequent events. For example, we asked for the causal strength between the statement that New York will not be attacked by terrorists and the statement that the US will not attack Iraq.

In the truth table task, we again used the perspective of a point 5 years previous to the current time with future tense from that point. So, for example, participants were asked to judge the probability of the following cases using future tense from a past point in time:

> New York will not be attacked by terrorists and the US will not attack Iraq.
> New York will not be attacked by terrorists and the US will attack Iraq.
> New York will be attacked by terrorists and the US will not attack Iraq.
> New York will be attacked by terrorists and the US will attack Iraq.

These judgments were used to predict the probability of the conditional and the causal strength judgments.

4.2. Results and discussion

4.2.1. Probability of conditional and causal strength tasks

The mean judged probability of conditionals and mean causal strength ratings are shown in Table 10, broken down by the polarity of components. For the probability ratings, there was no significant effect of antecedent polarity (F_1 and $F_2 < 1$) but an effect of consequent polarity ($F_1(1,25) = 36.76$, $p < .001$, MSE = 1921; $F_2(1,25) = 4.74$, $p < .05$, MSE = 591). There was also an interaction between the two factors ($F_1(1,25) = 64.41$, $p < 0.001$, MSE = 4362; $F_2(1,25) = 10.75$, $p < .005$, MSE = 1342). The causal strength ratings showed similar trends: no antecedent effect (F_1 and $F_2 < 1$), a significant effect of consequent polarity ($F_1(1,25) = 9.38$, $p < .005$, MSE = 2.27; $F_2(1,25) = 3.45$, $p = .074$, MSE = .70) and a significant interaction ($F_1(1,25) = 98.43$, $p < .001$, MSE = 8.44; $F_2(1,25) = 12.80$, $p < .001$, MSE = 2.60). Inspecting the means, it appears that doubly negated conditionals were judged as more probable (and more causally strong) than others, resulting in both the main effect and interactions. The correlation between the two dependent measures, across sentences, was .87.

We do not attach much theoretical significance to the effects of polarity, as the experimenters arbitrarily chose sentences in each class. The presence of negated components

Table 10

(a) Mean (SD) probability of conditional estimates in Experiment 3

	Consequent polarity	
	Affirmative	Negative
Antecedent polarity		
Affirmative	56.5 (12.9)	52.1 (13.8)
Negative	42.5 (13.8)	64.1 (11.6)

(b) Mean (SD) causal strength ratings in Experiment 3

	Consequent polarity	
	Affirmative	Negative
Antecedent polarity		
Affirmative	3.46 (.54)	3.18 (.63)
Negative	2.82 (.41)	3.69 (.46)

might have increased probability judgments merely because affirmative statements tend to refer to sets that are 'rarer' than their contrast classes. The pattern of correlations of probability judgments with our key measures was broadly similar for each group, so we will not break subsequent analyses down by polarity.

4.2.2. Correlation and regression analyses

Table 11 shows both correlations and regressions across the whole set of 32 sentences. We have changed the notation in this table so that A represents antecedent and C consequent regardless of negated components. For example, consider the counterfactual 'If alternative fuels had become widely available 5 years ago, then the price of petrol would

Table 11
Correlations and regressions computed across sentences using mean ratings for all participants in Experiment 3

	Probability of conditional	Causal strength
(a) Correlations between judged probability of conditionals and judged causal strength with various probabilities computed from the truth table task		
Hypotheses		
Conjunctive probability	.62*	.65*
Material conditional	.69*	.67*
Conditional probability	.83*	.80*
Delta-p rule	.69*	.72*
Other		
P(A)	.09	.14
P(C|A)	.40*	.38*
P(C|¬A)	−.08	−.14
(b) Multiple regressions using three independent predictors (β weights)		
P(A)	.06	.13
P(C|A)	.87*	.85*
P(C|¬A)	−.26*	−.34*

Note: A and C represent true antecedent cases and ¬A and ¬C false antecedent cases, regardless of the presence of negated components in the conditionals.

not have remained high.' In that case, P(C|A) would represent the probability that the price of petrol would *not* have remained high given that alternative fuels had become available and so on. The correlations presented in Table 11 are thus directly comparable to those for the previous experiments.

Table 11(a) shows a similar pattern of correlations to those observed in the previous experiments. We then tested for differences between these correlations. The correlation with conditional probability was higher than that with the material conditional and with the conjunction probability. This time, the correlation with conditional probability was not significantly higher than that with the delta-p rule. Subsequently, we performed multiple linear regressions on mean ratings across the 32 sentences with the same predictors as for Experiment 1, for both probability of conditional and causal strength judgments (see Table 11(b)). The results are strikingly similar to those for the indicative conditionals investigated in Experiments 1 and 2. Conditional probability remains the best predictor of both probability and causal strength judgments. There is also a negative but much smaller effect of P(C|¬A), which corresponds to P(q|¬p) for the previous experiments. As previously, we also computed regressions for individual participants, with the results shown in Table 5(c). These analyses confirm strong support for the conditional probability hypothesis, as in Experiments 1 and 2. There is a striking similarity between the results on the probability of counterfactuals and those on the probability of causal indicative conditionals. This similarity is itself the first experimental support for the theoretical argument that these conditional constructions are closely related to each other.

5. General discussion

In a series of three experiments, we investigated the subjective probability of causal indicative conditionals (Experiments 1 and 2) and closely related counterfactuals (Experiment 3). We examined four hypotheses about the subjective probability of one of these conditional. First, it is the subjective conjunctive probability of its antecedent and consequent: the conjunctive probability hypothesis. Second, it is the subjective probability of the material conditional: the material conditional hypothesis. Third, it is the conditional subjective probability of its consequent given its antecedent: the conditional probability hypothesis. Fourth, it depends on the delta-p rule: the delta-p rule hypothesis.

The support for the conditional probability hypothesis was extremely strong in all our experiments. There was no support at all for the material conditional hypothesis, and very little for the conjunctive probability hypothesis. The delta-p rule hypothesis received very modest support, as shown in the weak negative effect of P(q|¬p). The conditional probability hypothesis is a consequence of the Ramsey test of conditionals, and our results are consistent with psychological accounts grounded in the Ramsey test. The first of these was that of Rips and Marcus (1977), who were also influenced by the kind of conditional found in Stalnaker (1968). Much more recently, there has been a growing trend to explain thinking and reasoning about conditionals in terms of conditional probability. (See Evans & Over, 2004, for a review, and Oaksford, 2006, for comment, and Evans et al., 2005; Kaufmann, 2005; Oaksford, 2005; Oberauer et al., in press, for later work.)

The rest of the paper is organized as follows. We will begin by discussing our findings in more detail. We will then consider the implications of our findings for the most influential psychological account of conditionals: the mental model theory of Johnson-Laird and Byrne (1991, 2002). This theory implies the conjunctive probability hypothesis and, in some of its features, the material conditional hypothesis. Opposing theories are based on the Ramsey test and the conditional probability hypothesis. In these theories, the causal conditionals and counterfactuals we studied are Adams conditionals or Stalnaker conditionals, and so we will discuss these conditionals. Finally, we will turn our attention to the covariation and causal proposals, which imply that causal conditionals and counterfactuals make explicit statements about covariation or causation. We will conclude with a summary of our findings and their importance.

5.1. Findings

Previous work has demonstrated a close relation between the probability of abstract, basic indicative conditionals and conditional probability, supporting the conditional probability hypothesis (Evans et al., 2003; Hadjichristidis et al., 2001; Oberauer & Wilhelm, 2003; Over & Evans, 2003; but see also Girotto & Johnson-Laird, 2004). The present studies extended tests of this hypothesis to non-basic conditionals that, like almost all natural language conditionals, are not independent of content and background knowledge. These were causal indicative conditionals (Experiments 1 and 2) and related counterfactuals (Experiment 3) that are central to everyday discourse and decision making.

In Experiments 1 and 2, we used four categories of indicative conditionals defined (on the basis of a prior rating study) to be high in antecedent and consequent probability (HH), high in antecedent but low in consequent probability (HL) and, analogously, the low–high (LH) and low–low (LL) cases. This was mostly done in order to ensure a spread of probabilities across materials to facilitate our correlation and regression analyses. This manipulation allowed for an indirect test of the conditional probability hypothesis. This hypothesis predicts that ratings will increase with high consequent probability because, all else being equal, a high P(q) should lead to higher estimates of P(q|p). They did. In fact, ratings were significantly higher when antecedent as well as consequent probabilities were high. The antecedent effect cannot be predicted by the conditional probability hypothesis, but it may nevertheless be related to its foundation in the Ramsey test. It may be that when the antecedent p is unlikely, as it is in LL cases, some people find it hard to add p hypothetically to their beliefs. This may occur particularly if q is also improbable so that ‘p & q’ is a difficult possibility to imagine. In such cases, a shallow response might simply be to reject the conditional as unbelievable. Where the Ramsey test is fully implemented, however, one would expect the overwhelming evaluation of the indicative conditional to be based on the conditional probability, P(q|p).

We have much more sensitive data analyses available in our experiments here than in earlier experiments to test the conditional probability hypothesis. We can compare judgments made by the same participants to the same set of materials of both component probabilities and probability of the conditional. These analyses do not rely on the wide initial classification of conditional statements at all, but on multiple regression analyses performed across the whole set of sentences as actually rated by this group of participants. These analyses—see Tables 4(b), 8(b) and 11(b)—show that by far the best predictor of P(if p then q) is P(q|p), in line with the conditional probability hypothesis. This is also the case if the regressions are carried out on individual participants and then collated (Table 5).

Previous research on basic conditionals (see Evans et al., 2003; Oberauer & Wilhelm, 2003) showed a strong minority tendency in line with the conjunctive probability hypothesis, in that some participants equated P(if p then q) with P(p & q). What evidence is there for this hypothesis in the current study of non-basic conditionals? One consequence is that people should find these conditionals more believable when P(p) is high. This was the case in that people gave higher ratings to HH and HL than to LH and LL conditionals, although further inspection suggests that this finding was mostly driven by low ratings to LL conditionals rather than high ratings to HH conditionals. In the more sensitive regression analyses, P(p) had no measurable effect in Experiments 1 and 3 but showed up as a weak predictor in Experiment 2. On balance, the influence of conjunctive probability is a lot weaker for these realistic conditionals.

Experiment 3 supports the existence of a relation between counterfactuals and conditional probability, like Experiments 1 and 2 did for causal indicative conditionals. A finding of Experiments 2 and 3 was that people's judgments of the probability of indicative or counterfactual conditionals and of the causal strength between the antecedent and consequent events were very highly correlated. This might seem to provide support for the causal proposal. However, for conditionals and causal strength, by far the strongest predictor of judgments was the conditional probability, P(q|p), rather than the delta-p rule. The latter implies that P(q|¬p) should be an equally strong negative predictor. Where P(q|¬p) had an effect, it was very much weaker. We will now consider the implications of our results for theoretical accounts of conditionals in natural language.

5.2. Implications for mental model theory

Rips and Marcus (1977) were the first to propose what came to be known as mental model representations, of conjunctive and disjunctive propositions, in their version of the Ramsey test (Evans & Over, 2004, pp. 53–56). Mental models of conjunctions can represent conditional probability judgments, provided that some of these models are represented as more probable than others (Evans & Over, 2004, Ch. 9; Over & Evans, 2003). This representation of conditional probability is possible because P(q|p) is high when P(p & q) is higher than P(p & not-q), and low when P(p & not-q) is higher than P(p & q). There are also possible ways of indicating mental links between mental models that could represent conditional probability judgments (Evans & Over, 2004, Ch. 9). Our results certainly do not disconfirm the use of a general mental models framework to explain thinking and reasoning about conditionals, but rather imply that a modification and extension of the framework is necessary. This implication applies especially to the mental models theory of conditionals in Johnson-Laird and Byrne (2002).

Johnson-Laird and Byrne have long held that people's initial representation of an indicative conditional is a mental model of the conjunction, 'p & q,' plus some indication that other models are implicit. They would predict a conjunctive probability response, and there is a tendency to give this response in experiments on basic conditionals (Evans et al., 2003; Girotto & Johnson-Laird, 2004; Oberauer & Wilhelm, 2003). However, we found only very weak signs of it in our results on non-basic conditionals. Johnson-Laird and Byrne (1991) would also have predicted a material conditional pattern. Johnson-Laird and Byrne (2002) still represent some causal conditionals, as we would classify them, as material conditionals. As we pointed out above, they do this (p. 655) with their example, 'if a patient has malaria then she has a fever.' Yet the material conditional pattern was entirely absent from our data.

Johnson-Laird and Byrne (2002) helpfully introduce the distinction between basic and non-basic conditionals. Background knowledge affects the interpretation of non-basic conditionals, and by relying on this fact, they should be able to modify their theory to explain why there is not a conjunctive probability response to non-basic conditionals. They also claim (p. 673) that 'semantic modulation' can make conditionals about temporal or spatial relations and possibly causal relations or mechanisms. As we have suggested above, one might use semantic modulation to argue for a covariation proposal or a causal proposal for at least some conditionals, with the right semantic content. But both of these proposals imply the delta-p rule hypothesis, and we found only very modest support for that. Schroyens and Schaeken (2004) argue that Johnson-Laird and Byrne could relatively easily modify their theory to explain why there is a tendency to give the conditional probability response as the probability of basic conditionals. However, it is unclear how Johnson-Laird and Byrne (2002) could revise their particular theory to imply the conditional probability hypothesis for non-basic conditionals.

To make predictions about the probability of non-basic conditionals, Johnson-Laird and Byrne will finally have to break completely with the material conditional hypothesis. They cannot continue to claim (as on p. 651) that it is valid to infer 'if p then q' from 'not-p' alone or from q alone. These inferences, the so-called paradoxes of the material conditional, are valid if and only if the conditional is the material conditional (Evans & Over, 2004, pp. 19–21). These are invalid inferences when, as a result of semantic modulation, 'if p then q' means 'p raises the probability of q,' or 'p causes q,' as obviously neither of these statements follows from 'not-p' alone or from q alone. These inferences are also invalid if P(if p then q) is the conditional probability (Adams, 1998; Edgington, 1995). Johnson-Laird and Byrne will need new account of which inferences are valid, and which are invalid, for non-basic conditionals.

There is additionally the problem of accounting for the probability of non-basic conditionals. People make probability judgments about these conditionals by using heuristics, causal models, or other non-extensional procedures. But the mental model theory of Johnson-Laird and his collaborators so far only applies to extensional probability judgments (Johnson-Laird et al., 1999, p. 63).

If the extensional material conditional is not the ordinary conditional of natural language, then what is? We will consider two possibilities, the Adams conditional and the Stalnaker conditional. Both of these conditionals were founded on the Ramsey test and have influence some psychologists (see Evans & Over, 2005, for a review and Oaksford, 2006, for comment). Rips and Marcus (1977) used the conditional of Stalnaker (1968) as the basis of their psychological theory. But before considering the Stalnaker conditional, we return to Adams conditional, which we introduced in the discussion of Experiment 1.

5.3. The adams conditional

An Adams conditional, of the form 'if p then q,' is evaluated by means of the Ramsey test and its probability, P(if p then q), is exactly the conditional probability P(q|p). Edgington (1995, 2003) argued that this conditional can represent both the indicative and counterfactual conditionals of natural language. There is a truth value gap for Adams conditionals with false antecedents: the conditionals have a subjective probability (or believability) but no objective truth value in not-p cases. People do use the ordinary words 'true' and 'false' in 'not-p' cases, but supporters of the Adams conditional can point out that these words are ambiguous in natural language. For example, the words 'true' and 'false' do not mean the same applied to scientific statements and to expressions of moral principles or other subjective matters. Applied to a conditional in 'not-p' cases, these words can only express, for supporters of the Adams conditional, high or low confidence in the conditional. There is evidence that basic conditionals are Adams conditionals, but whether non-basic conditionals are Adams conditionals is a much more difficult question to answer (Evans & Over, 2004, Ch. 7 & Ch. 9).

An Adams conditional is not equivalent to an explicit statement that p *raises* the probability of q, as in (3) above, nor that p *causes* q, as in (4) above. A conditional probability P(q|p) can be high when p does not raise the probability of q and when p does not cause q. For example, P(q|p) can be high simply because P(q) is high. Does this mean that supporters of the view that these conditionals are Adams conditionals cannot account for the weak negative effect of P(q|¬p) in the current studies? Not necessarily, for they can argue that the use of a conditional *pragmatically* suggests, in certain ordinary contexts, that p raises the probability of q or that p causes q. Consider for example:

(6) If you take extra vitamin C, then your cold will be gone in three days.

In most contexts, asserting (6) would be misleading, and very bad advice, if extra vitamin C was not a causal factor raising the probability that the cold will be gone in 3 days. The argument would be that there is often a *pragmatic implicature* when a conditional like (6) is asserted: that not taking extra vitamin C will make it probable that the cold will last longer than 3 days (Evans & Over, 2004, pp. 158–159). This inverse implicature could cause a second conditional, 'if not-p then not-q,' to be added to the original representation. A Ramsey test on this added conditional would be believable to the extent that P(¬q|¬p) is high. A positive effect of this probability is the same as a negative effect of its complement P(q|¬p). Hence, our secondary finding could be accounted for by pragmatic implicature. It is worth noting that pragmatic implicatures, unlike semantic implications, are open to individual interpretation and can be 'cancelled' in some contexts (Grice, 1989). Future studies could investigate contexts in which possible implicatures would normally be cancelled. For example, a doctor who asserted (6) could do so in a context in which she had already expressed some doubt about the claim that extra vitamin C shortens the length of colds. She might then be taken by her patient as making an 'even if' assertion. That is, even if extra vitamin C is taken, the cold will still be gone in 3 days.

5.3.1. The Stalnaker conditional

There is also the possibility, as presupposed originally by Rips and Marcus (1977), that the conditionals we studied are not Adams conditionals but rather Stalnaker conditionals (Stalnaker, 1968, 1975). Stalnaker (1968) held that, when in the Ramsey test we make the smallest possible change to our belief state to suppose p consistently, we are thinking about the closest possibility in which p holds. The notion of the closeness of possibilities has inspired work in cognitive psychology on how people understand conditionals in general and counterfactuals in particular (Rips & Marcus, 1977; Evans & Over, 2004). It has also led to research in judgment and decision making and social psychology on how people conceive and respond to the closeness of possibilities and counterfactuals (Byrne, 2005;

Kahneman & Miller, 1986; Kahneman & Varey, 1990; Mandel et al., 2005; Roese, 1997, 2004, 2005; Teigen, 1998, 2005).

In Stalnaker's analysis, indicative and counterfactual conditionals are non-truth functional conditionals that are always either true or false—there are no truth value gaps. A Stalnaker conditional is true in TT possibilities and false in TF possibilities, but unlike an Adams conditional, a Stalnaker conditional *does* have a truth value in FT and FF possibilities. It is true in FT, or true in FF, if and only if TT is closer than TF as a possibility to FT, or to FF. Stalnaker (1968, 1975) takes his approach to both indicative and counterfactual conditionals, with only a pragmatic difference between them (Evans & Over, 2004).

To illustrate, assume that (1)—'if the cost of petrol increases then traffic congestion will improve'—is a Stalnaker conditional, and that the cost of petrol will not increase (the world is in a FT or FF state). Now consider the closest possible state of the world in which the cost of petrol increases. Perhaps there are many people in Government who want to put an extra tax on petrol to increase its cost, and they almost have enough votes to make this tax change. That would make the tax change the closest possibility in which the cost of petrol goes up, and (1) will be true if traffic congestion improves in that closest possibility. If it does not improve in that closest possibility, then (1) is false. A material conditional, of the form 'not-p or q,' is truth functional and is of course always true in FT and FF possibilities. But a Stalnaker conditional is not truth functional: it is only *sometimes* true in FT and FF possibilities.

Stalnaker (1968) argued that the probability of a conditional, P(if p then q), in his analysis is the conditional probability, P(q|p), given by the Ramsey test. However, the probability of a Stalnaker conditional is not in general identical with the conditional probability (as famously proved by Lewis, 1976; see also Bennett, 2003, for a review of work on Lewis's proof and Evans & Over, 2004, Ch. 2, for an account of it in a psychological context). Nevertheless, it can be argued that Stalnaker's theory was 'almost right' (Lewis, 1976), and that the probability of a Stalnaker conditional is often too close, in many ordinary contexts, to the conditional probability for a difference to be easily found in experiments (Evans & Over, 2004, pp. 28–30).

It may be that people use the Ramsey test, in bounded thought, as a quick and generally effective way to get close enough to the probability of a Stalnaker conditional. The points we made above about pragmatic implicature for Adams conditionals can also be made for Stalnaker conditionals. For example, the use of (6) as a Stalnaker conditional could have the same pragmatic implicatures that (6) would have as an Adams conditional. Moreover, supporters of Stalnaker can also explain the low ratings given to LL conditionals in our data. The LL case, in which both p and q are improbable, represents a distant possibility, and people may use a 'closeness' or 'proximity' heuristic (Roese, 2004; Teigen, 1998, 2005) as a quick way to reject such a conditional as highly improbable. (This heuristic takes very close possibilities to be highly probable and very distant ones to be highly improbable.) Since the probability of a Stalnaker conditional will often be very close to the conditional probability, supporters of Stalnaker, such as Rips and Marcus, could use this fact to explain why the conditional probability hypothesis is so strongly supported by our results.

5.3.2. The covariation and causal proposals

Causal indicative conditionals, and related counterfactuals, may be neither Adams conditionals nor Stalnaker conditionals, but a stronger conditional, as implied by both the covariation and causal proposals. Suppose a conditional, ‘if p then q,’ means that p raises the probability of q, as the covariation proposal states, or that p causes q, as the causal proposal states. As we have already pointed out, an Adams conditional is weaker than the statement that p raises the probability of q, or that p causes q. A Stalnaker conditional is also weaker than either of these statements. As a Stalnaker conditional, (1) would be true if a petrol cost increase was followed, in a mere coincidence, by traffic congestion improvement. The covariation and causal proposals might also be able to explain our confirmation of the conditional probability hypothesis and our other results.

Consider again the findings in Experiment 2 and Experiment 3 of a high correlation between the judged probability of the conditional and causal strength judgments. This correlation could be taken as evidence for the causal proposal: that a conditional like (1) is equivalent in meaning to an explicit causal statement like (4). In that case, the negative effect of $P(q|\neg p)$ in Experiment 2 would not be pragmatic but rather semantic. The delta-p rule is too simple on its own to measure causal strength, but $P(q|p) - P(q|\neg p)$ must normally be positive before one would say that p is a ‘cause’ of q (Cheng, 1997; Pearl, 2000). As observed earlier, however, we found only a weak negative effect of $P(q|\neg p)$ in participants’ judgments of causal strength.

The weakness of both negative effects of $P(q|\neg p)$ could be the result of a well known bias studied in the literature on contingency judgment. It is well established that people tend to focus too much on the a and b cells, rather than the c and d cells, of a 2×2 contingency table when tested on the delta-p rule (Anderson, 1990; Kao & Wasserman, 1993; Shanks, 1995; Stanovich & West, 1998). This tendency would be equivalent, in our results, to placing too much emphasis on $P(q|p)$ rather than $P(q|\neg p)$, in spite of understanding the conditional as a causal assertion. The weakness of the negative effects of $P(q|\neg p)$ in our results could simply be the result of this bias, which may arise because focusing more on $P(q|p)$ than on $P(q|\neg p)$ is often an effective heuristic for judging correlation or causal strength in ordinary affairs (Anderson, 1990; Cheng, 1997; Over & Green, 2001; Shanks, 1995; Stanovich & West, 1998).

Previous discussions in the literature of the delta-p rule have almost always been directly related to explicit statements of correlation or causation, such as (3) and (4) above. There is not extended work in linguistics, philosophical logic, or psychology that tries to explain the meaning of some type of conditional in terms of the delta-p rule. For example, as far as we are aware, no normative logic for a conditional has ever been formulated on the basis that its formal semantics is given by the delta-p rule. Owing to the bias that could explain the weak negative effect of $P(q|\neg p)$, more work will have to be done theoretically and experimentally before it can be definitely determined whether or not the delta-p rule is basic to the semantics of causal conditionals and counterfactuals.

6. Conclusions

In this paper, we have provided the first experimental psychological evidence on the subjective probability of causal indicative conditionals, and closely linked counterfactuals, commonly used in natural language. We have shown that these conditionals cannot be material, truth functional conditionals. Our results strongly disconfirm the material conditional hypothesis, which is implied by Johnson-Laird and Byrne (1991) and still found in aspects of Johnson-Laird and Byrne (2002). We have also shown that evidence for a conjunctive probability response, predicted by Johnson-Laird and Byrne (1991, 2002), is very weak. We have found that people's confidence in causal conditionals and related counterfactuals is predominantly determined by the conditional probability, P(q|p), as implied by the conditional probability hypothesis, which is implied in turn by the Ramsey test. We have studied, for the first time, possible relations between probability judgments about causal and counterfactual conditionals, causal strength judgments, and the delta-p rule. We have found a weak effect of the delta-p rule and pointed out how this could be connected to a bias in contingency judgment. An extended mental models framework may be able to explain our results, but it is an open question whether these are best explained by the supposition that natural language conditionals are Adams conditionals, Stalnaker conditionals, or stronger conditionals. The only way to make progress is to integrate the study of conditionals with the psychology of causal judgment (Sloman, 2005) and, more generally, with work on probability in judgment and decision making (Evans & Over, 2004). Only this integration can tell us how the Ramsey test is implemented in detail, by the use of deductive and inductive inferences, heuristics, and causal models for probability judgment. In this paper, we have taken an important step in that direction by demonstrating how very closely probability judgments about causal and counterfactual conditionals are related to conditional probability judgments.

7. Uncited references

Over (1993), Over (2004), Roberts (1998), Stanovich and West (2003).

Acknowledgments

We are grateful to Stefan Kaufmann, Mike Oaksford, and two anonymous referees for their comments on an earlier version of this paper.

References

Adams, E. (1975). *The logic of conditionals: an application of probability to deductive logic*. Dordrecht: Reidel.
Adams, E. (1998). *A primer of probability logic*. Stanford: CLSI publications.
Allan, L. G. (1980). A note on the measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychometric Society, 15*, 147–149.
Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.
Byrne, R. M. J. (2005). *The rational imagination: how people create alternatives to reality*. Cambridge, MA: The MIT Press.
Cheng, P. (1997). From covariation to causation: a causal power theory. *Psychological Review, 104*, 367–405.
Dudman, V. H. (1984). Conditional interpretations of if-sentences. *The Australian Journal of Linguistics, 4*, 143–204.
Edgington, D. (1995). On conditionals. *Mind, 104*, 235–329.
Edgington, D. (2003). What if? Questions about conditionals. *Mind and Language, 18*, 380–401.
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*, 321–355.
Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford: Oxford University Press.
Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality, and conditionals: a critique of Johnson-Laird & Byrne (2002). *Psychological Review, 112*, 1040–1052.
Girotto, V., & Johnson-Laird, P. N. (2004). The probability of conditionals. *Psychologia, 47*, 207–225.
Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hadjichristidis, C., Stevenson, R. J., Over, D. E., Sloman, S. A., Evans, J. St. B. T., & Feeney, A. (2001). On the evaluation of If p then q conditionals. In Proceedings of the 23rd annual meeting of the cognitive science society. Edinburgh.

Hair, J. F., Jr., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

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

Kahneman, D., & Miller, D. (1986). Norm theory: comparing reality to its alternatives. *Psychological Review, 93*, 136–153.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristic and biases* (pp. 201–210). Cambridge: Cambridge University Press.

Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: the loser who almost own. *Journal of Personality and Social Psychology, 59*, 1101–1110.

Kaufmann, S. (2004). Conditioning against the grain: abduction and indicative conditionals. *Journal of Philosophical Logic, 33*, 583–606.

Kaufmann, S. (2005). Conditional predictions: a probabilistic account. *Linguistics and Philosophy, 28*, 181–231.

Kao, S. F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1363–1386.

Keren, G., & Teigen, K. H. (2004). Yet another look at the heuristics and biases approach. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 89–109). Oxford: Blackwell.

Lagnado, D. A., & Sloman, S. A. (2004). Inside and outside probability judgment. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 157–176). Oxford: Blackwell.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review, 95*, 581–589.

Liu, I.-M., Lo, K.-C., & Wu, J.-T. (1996). A probabilistic interpretation of "If-Then". *The Quarterly Journal of Experimental Psychology, 49A*, 828–844.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 149–157.

Macchi, L., Osherson, D., & Krantz, D. H. (1999). A note on superadditive probability judgment. *Psychological Review, 106*, 210–214.

Mandel, D. R., Catellani, P., & Hilton, D. J. (Eds.). (2005). *The psychology of counterfactual thinking*. London: Routledge.

Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*, 172–175.

Oaksford, M. (2005). Probability logic and the MP-MT asymmetry in conditional inference. Manuscript submitted for publication.

Oaksford, M. (2006). Making connections in conditional inference. *American Journal of Psychology*, 170–175.

Oaksford, M., & Chater, N. (1998). *Rationality in an Uncertain World*. Hove, UK: Psychology Press.

Oaksford, M., & Chater, N. (2003). Conditional probability and the cognitive science of conditional reasoning. *Mind and Language, 18*, 359–379.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking and Reasoning, 5*, 193–243.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 883–889.

Oaksford, M., & Moussakowski, M. (2004). Negations and natural sampling in data selection: ecological vs. heuristic explanations of matching bias. *Memory & Cognition, 32*, 570–581.

Oaksford, M., & Wakefield, M. (2003). Data selection and natural sampling: probabilities do matter. *Memory & Cognition, 31*, 143–153.

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*, 688–693.

Oberauer, K., Geiger, S., Fischer, K., & Weidenfeld, A. (in press). Two meanings of 'if"? Individual differences in the interpretation of conditionals. *The Quarterly Journal of Experimental Psychology*.

Ohm, E. & Thompson, V. A. (in press). Conditional probability and pragmatic conditionals: Truth, reasoning, and behavioural effectiveness. *Thinking and reasoning*.

Over, D. E. (1993). Deduction and degrees of belief. *Behavioral and Brain Sciences, 26*, 361–362.

Over, D. E. (2004). The psychology of conditionals. In K. I. Manktelow & M. C. Chung (Eds.), *Psychology of reasoning: theoretical and historical perspectives*. Hove: Psychology Press.

Over, D. E., & Evans, J. St. B. T. (2003). The probability of conditionals: the psychological evidence. *Mind and Language, 18*, 340–358.

Over, D. E., & Green, D. W. (2001). Contingency, causation and adaptive inference. *Psychological Review, 108*, 682–684.

Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge: Cambridge University Press.

Politzer, G., & Bourmand, G. (2002). Deductive reasoning from uncertain conditionals. *British Journal of Psychology, 93*, 345–381.

Ramsey, F. P. (1990). General propositions and causality (original publication, 1931). In D. H. Mellor (Ed.), *Philosophical papers* (pp. 145–163). Cambridge: Cambridge University Press.

Rips, L. J., & Marcus, S. L. (1977). Suppositions and the analysis of conditional sentences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 185–219). New York: Wiley.

Roberts, M. J. (1998). Inspection times and the selection task: are they relevant? *The Quarterly Journal of Experimental Psychology, 51A*, 781–810.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133–148.

Roese, N. J. (2004). Twisted pair: counterfactual thinking and hindsight bias. In N. Harvey & D. J. Koehler (Eds.), *Blackwell handbook on judgment and decision making*. Oxford: Blackwell.

Roese, N. J. (2005). *If only: how to turn regret into opportunity*. New York: Broadway Books.

Sanford, D. H. (1989). *If p,then q: conditionals and the foundation of reasoning*. London: Routledge.

Schroyens, W., & Schaeken, W. (2004). Guilt by association: on iffy propositions and the proper treatment of mental-models theory. *Current Psychology Letters, 12*, 1. Available at http://cpl.revues.org/document411.html.

Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge: Cambridge University Press.

Sloman, S. A. (2005). *Causal models: how people think about the world and its alternatives*. New York: Oxford University Press.

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

Stalnaker, R. (1975). Indicative conditionals. *Philosophia, 5*, 269–286.

Stanovich, K. E., & West, R. F. (1998). Who uses base rates and P(D/¬H)? An analysis of individual differences. *Memory & Cognition, 28*, 161–179.

Stanovich, K. E., & West, R. F. (2003). Evolutionary versus instrumental goals: how evolutionary psychology misconceives human rationality. In D. E. Over (Ed.), *Evolution and the psychology of thinking* (pp. 171–230). Hove, UK: Psychology Press.

Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *The Quarterly Journal of Experimental Psychology, 48A*, 613–643.

Stevenson, R. J., & Over, D. E. (2001). Reasoning form uncertain premises: effects of expertise and conversational context. *Thinking and Reasoning, 7*, 367–390.

Swartz, N. (1999). 'The' modal fallacy. Available at http://www.sfu.ca/philosophy/swartz/modal_fallacy.htm.

Teigen, K. H. (1998). When the unreal is more likely than the real: post hoc probability judgments and counterfactual closeness. *Thinking and Reasoning, 4*, 147–177.

Teigen, K. H. (2005). The proximity heuristic in judgments of accident probabilities. *British Journal of Psychology, 96*, 423–440.

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

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.