What Underlies the Assessment of Conditionals?

Uri Hasson (uhasson@uchicago.edu)
Departments of Psychology and Neurology,
Chicago, IL 60615, USA.

Clare Walsh (clare.walsh@plymouth.ac.uk)
Department of Psychology,
Plymouth, PL48AA, UK.

Philip N. Johnson-Laird (phil@princeton.edu)
Department of Psychology
Princeton, NJ 08544, USA.

Abstract

According to some theorists, the strength of a conditional statement is determined solely by the extent to which the conditional's antecedent is sufficient for its consequent. On this view, the strength of conditionals is independent of the degree to which the antecedent is necessary for the consequent. Data from four experiments reported here demonstrate that when people evaluate conditionals, they are also sensitive to whether the antecedent is necessary for the consequent. Conditionals are perceived as stronger the more necessary the antecedent is for the consequent. We discuss the implications of these results for theories of the meaning of conditional statements.

Introduction

A longstanding position in logic and in the philosophy of language is that the meaning of a conditional statement of the form *if p then q* is related to the likelihood that the conditional's consequent (q) is true given that the conditional's antecedent (p) is true (e.g., Adams, 1975). Conditionals are more assertable, or more likely to be true, when the likelihood of the consequent given the antecedent is high.

Recent empirical studies have evaluated the psychological reality of this position (e.g., Evans, Over, & Handley, 2003; Oberauer & Wilhelm, 2003). They have shown that when people evaluate conditionals, they are very sensitive to the conditional probability of the consequent given the antecedent, P(q | p); i.e., they rate conditionals as more likely to be true when this probability is higher.

The conditional probability of the consequent given the antecedent ("conditional probability" henceforth) designates the relative frequency of just those instances in which the antecedent of the conditional is true (*pq*/ (*pq* + *p*¬*q*), *where* ¬ *refers to not*); it is independent of the frequency of those cases in which the antecedent of the conditional is false *(*¬*pq* and ¬*p*¬*q*). Consequently, experimental findings that have shown a correlation between people's ratings for the likelihood of conditionals and conditional probability have been interpreted as indicating that when people assess conditionals they consider only those two states of affairs in which the antecedent holds. This view is well summarized by Hadjichristidis et al. (2001), "as far as the basic evaluation of a conditional is concerned, our results suggest that people construct an imaginary world in which the antecedent holds and then consider the likelihood that the consequent holds in the same world."

On this view, when people assess conditionals of the form *if p then q*, they consider whether the antecedent is sufficient for the consequent, which is reflected by P(q | p), but they do not consider whether it is *necessary* for the consequent.

By definition, considering necessity means considering states of affairs in which the antecedent is false. The degree of necessity might be influenced, e.g., by the absolute frequency of the ¬*pq* cases, or by the relative frequency of the *pq* and ¬*pq* cases. This latter measures, pq /(pq + ¬pq), is captured by P(p | q). In any case, a basic assumption of the conditional probability hypothesis is that the degree to which the antecedent is *necessary* for the consequent does not, and should not affect the strength of the conditional, because the degree of necessity is *independent* of conditional probability. In other words, any theory consistent with the idea that people also evaluate necessity would no longer be *just* a theory of conditional probability, because it would have abandoned the basic assumption that all that matters are those states of affairs in which the antecedent is true.

In this paper, we examine the hypothesis that when people evaluate whether a conditional is likely, or reasonable, they do consider whether the antecedent is necessary for the consequent. They consider whether the consequent can hold in the absence of the antecedent, and when such cases are frequent, or easily accessible, people perceive conditionals as being weaker.

As a starting point, we report a re-analysis of findings that have been interpreted to support the conditional probability hypothesis, and we show that they also reveal an independent effect of necessity (Section I). We report a replication of those findings (Section II). We then examine the role of necessity and sufficiency in the assessment of causal conditionals (Section III), and finally discuss the implications of the results for theories of the conditional (Section IV).

I. The Likelihood of Conditionals

This section presents a re-analysis of data reported by Evans, Handley, and Over (2003). These authors conducted three experiments in which participants judged the likelihood of conditionals on the basis of information

presented to them. For example, participants were told that a pack of card contains: 1 yellow circle, 4 yellow diamonds, 16 red circles, and 16 red diamonds. Given this information, the participants judged how likely it was that a certain claim was true for a card drawn at random, (e.g., "if the card is yellow then it has a circle printed on it"). The ratings were made on a scale of 1 (*very unlikely*) to 5 (*very likely*). All three experiments showed a strong correlation between participants' judgments of likelihood and the conditional probability of the consequent given the antecedent; here, P(circle printed | card is yellow) = 0.2.

One unexpected finding, however, was that the experiments revealed an independent tendency for conditionals to be rated as more likely when the frequency of cases in which both the antecedent and consequent were true increased (i.e., *pq* cases; 1/37 in the example above). Although this effect is independent of conditional probability, the authors suggested it might reflect the tendency of some people to employ shallow processing in which they evaluate a conditional as a conjunction: these people consider just the probability of the *pq* cases, but do not consider the $p\neg q$ cases. In any case, apart from this effect, which seemed to have occurred only for certain individuals, the results were interpreted as strongly supporting the conditional probability hypothesis.

While all three experiments were well suited to demonstrate that variations in conditional probability affect participants' ratings, only Experiment 2 offered the opportunity to refute the conditional probability hypothesis, because this experiment made it possible to test whether the participants considered the necessity of the antecedent for the consequent. Within each level of conditional probability, the experiment varied the distribution of the false-antecedent cases (i.e., the frequency of $\neg pq$ and $\neg p\neg q$ cases) in 4 ways. An example is presented in Table 1 (the experiment contained 8 such quartets, which varied in their conditional probabilities).

Table 1: An example of four different distributions from Evans et al. (2003, Experiment 2) with the resulting conditional probability and the probability of the conjunction.

	Distribution				Measures	
	pq	p¬q	¬pq	¬p¬q	P(q/p)	P(pq)
A	1	1	1	4	1:2	1:7
B	1	1	4	1	1:2	1:7
C	1	1	1	1	1:2	1:4
D	1	1	4	4	1:2	1:10

On the conditional probability hypothesis, the conditional *if p then q* should be judged as equally likely under all four distributions, because they yield identical conditional probabilities. On a modified version of this approach (Evans, Handley, & Over, 2003), conjunctive probability may have an independent effect, and so the conditional should be judged as most likely under condition *C*, and as least likely under *D*, with conditions *A* and *B* somewhere between these two extremes. A crucial prediction is that conditions *A* and *B* should not differ systematically in the likelihood judgments they elicit, because in all the quartets they were identical on the measures of P(q | p), P(pq) and even, P(p). However, they do differ in another way: the antecedent is more necessary for the consequent in condition *A* than in condition *B*. Hence, if individuals are affected by necessity, then they should judge the conditional as more likely to be true in condition *A* than in *B*. Necessity may correspond to the likelihood that the consequent holds when the antecedent is false, P(p | q), or to the potency of the antecedent, e.g., ΔP; P(q | p) – P(q | ¬p). In the example above, conditions *A* and *B* differ on both.

Although Evans et al. did not analyze their data for this effect, the data they report (p.328, table 2: conditionals) is rather surprising. The mean ratings for the four types of conditions in Table 1 were as follows: condition *C:* 3.14; condition *A:* 3.13; Condition *B:* 2.80, Condition *D:* 1.04. Statistical re-analysis of these data confirmed that there was a highly significant difference between the ratings in conditions *A* and *B*; $F(1, 47) = 8.4$, $p = .006$.

The interpretation of the re-analysis is clear: when participants evaluate the likelihood of conditionals, they are sensitive to the distribution of cases in which the antecedent is false, and they rate conditionals as more likely when the antecedent is more necessary for the consequent. Hence the difference between the ratings given for conditions *A* and *B*. These conditions differed in the necessity of the antecedent for the consequent; the antecedent was more necessary in condition *A* than in condition *B*. However, these conditions did not differ in conditional probability, P(q | p), conjunctive probability, P(pq), or in the probability that the antecedent is true, P(p). Before we draw any conclusions, it was essential to obtain an independent replication of the results.

II. The Likelihood of Conditionals: The Role of Necessity

This study was designed to test the predictions and analysis presented above. Participants evaluated the likelihood of conditionals on the basis of information given to them about the distribution of four sorts of cards in a stack of cards. The design controlled for conditional probability and conjunctive probability, and manipulated the necessity of the antecedent for the consequent, P(p | q) by varying the frequency of the ¬p cases.

Method

Participants. Thirty nine Princeton University undergraduates participated in the study for course credit.

Materials. Participants were provided with descriptions of packs of cards. In each pack the cards could be printed in one of two shapes, and in one of two colors. The colors used were blue, red, yellow and green, and the shapes used were square, circle, diamond and triangle. Shapes and colors were randomly assigned twice to the different conditions.

Design and Procedure. The design was based on manipulating three factors: conditional probability (0.25, 0.5, 0.75), conjunctive probability (0.2, 0.13) and the

distribution of false-antecedent (¬p) cases: in one distribution, the majority of the ¬*p* cases were ¬*pq* cases, and in the other, the majority of ¬*p* cases were ¬*p*¬*q* cases. These manipulations yielded the 12 experimental conditions presented in Table 2.

Table 2: The 12 different frequency distributions in Experiment 1.

Distribution				Measures	
pq	p¬q	¬pq	¬p¬q	P(q/p)	P(pq)
1	3	1	0	0.25	0.2
1	3	0	1	0.25	0.2
1	3	3	1	0.25	0.13
1	3	1	3	0.25	0.13
1	1	2	1	0.5	0.2
1	1	1	2	0.5	0.2
1	1	5	1	0.5	0.13
1	1	1	5	0.5	0.13
3	1	8	3	0.75	0.2
3	1	3	8	0.75	0.2
3	1	15	5	0.75	0.13
3	1	5	15	0.75	0.13

The trials were presented on a computer screen and their order was randomized for each participant. Within each trial, the descriptions of the four sorts of cards were presented one below the other, and the order in which the four types of cards were displayed on the screen in each trial was separately randomized for each trial. The study also included 8 filler problems, which were not analyzed, and which were meant to encourage a full use of the rating scale: in four of these problems conditional probability was 1, and in four it was 0. Each participant completed the twenty problems. The participants were told to read each scenario carefully before rating how likely the conditional was to be true given the pack of cards provided. The scale ranged from 1 (*very unlikely*) through 5 (*very likely*). The study was self-paced, and participants pressed the spacebar to advance to the next trial after each rating.

Results and Discussion

As predicted, there was a reliable effect of the distribution of ¬*p* cases on the ratings of the conditional: ratings were lower when the majority of ¬*p* cases were ¬*pq* cases (M = 2.87, SE = 0.05) than when the majority were ¬*p*¬*q* cases (M = 2.95, SE = 0.05), $F(1, 38) = 5.15$, $p < .03$. There was a main effect of conditional probability, $F(2, 76) = 144$, $p < .001$; ratings increased with higher conditional probability (Means = 2.07, 2.96, 3.71, respectively). Finally, there was an unpredicted interaction between conditional probability and conjunctive probability, $F(2, 76) = 5.26$, $p = .007$. When conditional probability was 0.25 or 0.5, higher conjunctive probability was associated with higher strength, but when conditional probability was 0.75, there was an unexpected reversal of that pattern. No other effects or interactions approached significance.

An analysis of individual responses showed that 22 of the 39 participants gave responses that corresponded exactly to the levels of conditional probability. These participants gave uniform ratings of 2 to all conditionals with a low conditional probability, 3 to all conditionals with a medium conditional probability, and 4 to all conditionals with a high conditional probability. The remaining effects and interactions were therefore due to the remaining 17 participants. When the mean ratings of just those 17 participants were analyzed, the analysis revealed the main effects and interactions reported above. This sample responded with a mean rating of 2.7 when the majority of ¬*p* cases were of the ¬*pq* sort, and with a mean rating of 2.89 when the majority of cases were of the ¬*p*¬*q* sort.

The results of this study corroborate those of the re-analysis presented earlier. They show that when people evaluate conditionals on the basis of provided frequency-data, they are sensitive to the degree to which a conditional's antecedent is necessary for its consequent. The results also indicate that this task elicits different sorts of responses from different participants (see also Evans et al., 2003, Exp. 3).

The method employed here, and in previous studies, is based on asking participants to evaluate a conditional on the basis of data that correspond to a 2 x 2 contingency table. This method is ill suited for examining what information people spontaneously consider when evaluating a conditional, because it presents participants with a full partition of events, and people do not normally encounter such distributions when understanding conditionals. In the next study, we examined whether people spontaneously consider the necessity of the antecedent when they evaluate conditionals in the absence of such frequency data.

III. The Role of Necessity in Assessing Causal Conditionals

Conditionals are commonly used to express causal relations, and any theory that aims to account for people's evaluation of conditionals should also aim to explain how people evaluate causal conditionals, e.g., *If a person is obese then this person suffers from backache*.

A causal conditional should be more sensible the more likely the cause is to bring about the effect, i.e., the greater its sufficiency: P(effect | cause). Both the conditional probability hypothesis, and the one developed here predict that this conditional probability should correlate positively with participants' evaluations of conditionals. However, as we have seen in the previous sections, people are also sensitive to the necessity of the antecedent for the occurrence of the consequent. The present hypothesis accordingly predicts that when alternative causes are easily accessible, a conditional should be rated as weaker.

When the number of alternative causes for an effect increases, the necessity of the antecedent to bring about the consequent goes down. Hence, given the consequent, it is

less certain that the antecedent occurred, i.e., P(cause | effect) is lower. Conditionals should therefore be rated as less sensible when P(cause | effect) decreases.

We designed an experiment to test this prediction. The participants evaluated causal conditionals that were equated for sufficiency, P(effect | cause), but that differed on necessity, P(cause | effect). We predicted that the accessibility of alternative causes would affect the evaluation of causal conditionals.

Norming the Materials

To develop the materials for our experiment, we first constructed 80 causal conditionals of four sorts depending on whether sufficiency was high or low and on whether necessity was high or low. In a norming study, ten Princeton undergraduates rated the probability of the effect given the cause for these 80 conditionals (i.e., sufficiency), and ten other participants rated the probability of the cause given the effect (i.e., necessity). Each trial consisted of the presentation of one clause of the conditional as a fact and the other clause as a putative conclusion, e.g.:

You know for a fact that a person is obese.

Conclusion: this person suffers from backache.

The participants rated (on a scale of 1 - 9) how strongly the conclusion followed from the fact. The ratings enabled us to select ten examples of four sorts of conditionals that varied independently in the degree of necessity and sufficiency of the cause. The ratings for the four sorts of conditionals, and an example of each sort are presented in Table 3.

Table 3: Examples of materials chosen from the norming study, with mean ratings of each of the four conditions

Sufficiency P(effect| cause)	Necessity P(cause |effect)	Conditional statements
High (M=7.5)	High (M=7.8)	If a metal is heated, then that metal expands.
High (M=7.5)	Low (M=3.8)	If a person jumps into the ocean, then that person will get wet.
Low (M=5.4)	High (M=7.8)	If a person exercises regularly, then that person can run a marathon.
Low (M=5.4)	Low (M=3.8)	If a person goes without food for a long time, then that person will develop heartburn.

In order to demonstrate an effect of necessity on the evaluation of conditionals we could merely have contrasted two of the conditions in Table 3 (i.e., either the first two or latter two), as both pairs were equated on sufficiency and differed in necessity. However, we used a complete two by two design because we wanted to test a subtler prediction. The prediction was that in cases where a cause is highly sufficient for an effect (i.e., rows 1 and 2 in Table 3), people should be highly sensitive to the necessity of the cause. In such cases, the conditional asserts a relation in which the sufficiency of the cause is undeniable, and so such conditionals are likely to be understood as making a statement about the necessity of the cause. In contrast, when the cause is not sufficient for the effect (i.e., rows 3 and 4 in Table 3), the conditional asserts a causal relation that is highly debatable. In this case, the participants are likely to be less sensitive to the cause's necessity.

To summarize, our hypothesis predicts that both sufficiency and necessity should affect the evaluation of causal conditionals. But, the two factors are likely to interact, so that people are more sensitive to the necessity of a cause when its sufficiency is high than when its sufficiency is low.

Study I

The participants rated how reasonable the four sorts of conditional were. We did not ask the participants to rate the likelihood of the conditionals, because a reference to likelihood could bias participants to focus on extensional probability calculations, and inflate the importance of conditional probability. Ratings of reasonableness have been used previously (cf. Rips, 2002) and have been shown to be quite sensitive.

Method

Participants. Twenty Princeton University undergraduates participated in the study for course credit.

Materials, Design and Procedure. Each participant rated the 40 conditionals chosen from the norming study (10 conditionals of each type in Table 3) in a different random order. The materials were presented on a computer running the EPRIME software. The participants were told that they would be presented with statements making various claims, and their task was to judge how reasonable each claim was on a scale ranging from 1 (*not at all reasonable*) to 9 (*very reasonable*). They made their judgments at their own pace, and they pressed the spacebar when they were ready for the next conditional.

Results and Discussion

The mean ratings for the four sorts of conditionals (+ *SE*) are presented in Fig. 1. As the figure shows, conditionals with high sufficiency were rated as more reasonable than conditionals with low sufficiency, $F(1, 22) = 377$, $p < .001$. Likewise, conditionals with high necessity were rated as slightly more reasonable than those with low necessity, $F(1, 22) = 5.2$, $p < .05$. Furthermore, these two variables interacted significantly, $F(1, 22) = 5.4$, $p < .05$. With high sufficiency, conditionals with high necessity were rated as more reasonable than those with low necessity, $t(22) = 4.3$, $p < .001$; whereas, with low sufficiency, whether conditionals had high or low necessity had no reliable effect on ratings, $t(22) = 0.3$, *ns*.

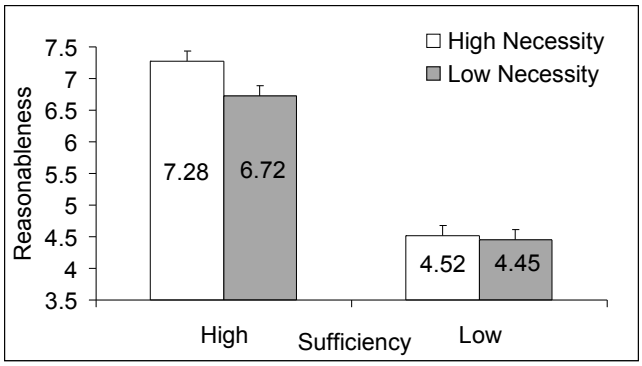

Figure 1: Mean Ratings of Reasonableness in Study 1 as a Function of the Sufficiency and Necessity of the Cause.

The results show that when participants could think of cases where the effect holds in the absence of the cause, they rated the conditional as weaker. The effect was found for conditionals in which the antecedent was judged as highly sufficient for the effect. Hence the two left bars in Fig. 1 differ reliably. These bars show the reasonableness ratings for two sorts of conditionals that are equated for sufficiency of the cause, but not for its necessity. The interaction between the degree of sufficiency and necessity shows that people attribute a greater role for necessity when the cause is highly sufficient for the effect.

Study II

In this study, one group of participants rated how reasonable the conditionals were, in a replication of Study 1. A second group of participants rated how likely these conditional were to be true. Instructions that refer to likelihood should emphasize assessment of conditional probability, because the question "how likely is *if p then q*" could be interpreted as "if p, how likely is q?" Such instructions could therefore prompt responses that reflect more sensitivity to the sufficiency of the cause.

Method

Thirty-eight Princeton University undergraduates participated in the study for course credit. The materials were those used in Study 1. Nineteen participants rated the reasonableness of the four sorts of conditional on a scale of 1 - 9, as in Study 1. The other nineteen participants rated the likelihood that the four sorts of conditional were true on a scale of 1 (*not at all likely to be true*) to 9 (*very likely to be true*). The procedure was otherwise the same as in Study 1.

Results and Discussion

Figures 2 and 3 present the two groups' mean ratings for the four sorts of conditional (+*SE*). The ratings of reasonableness in Fig. 2 replicated the results of Study 1: conditionals with high sufficiency were rated as more reasonable than those with low sufficiency, $F(1, 18) = 278$, $p < .001$. Conditionals with high necessity were rated as more reasonable than those with low necessity, $F(1, 18) = 4.63$, $p < .05$. Once again, the two variables interacted, $F(1, 18) = 4.83$, $p < .05$: the effect of necessity occurred only for conditionals with high sufficiency.

The ratings of the likelihood of the conditionals in Fig. 3 showed only a reliable effect of sufficiency, $F(1, 18) = 146$, $p < .001$. Hence, in the experiment as a whole, there was a reliable three-way interaction between the rating task (reasonableness or likelihood), the level of sufficiency, and the level of necessity, $F(1, 36) = 5.4$, $p < .05$. An analysis of the judgment latencies showed that people were faster to estimate reasonableness than likelihood (M = 5.7 vs. 7.2 sec); $t(36) = 3.9$, $p < .001$.

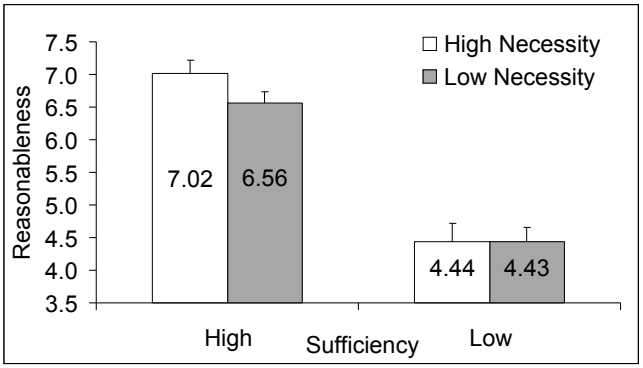

Figure 2: Mean Ratings of Reasonableness in Study 2 as a Function of the Sufficiency and Necessity of the Cause.

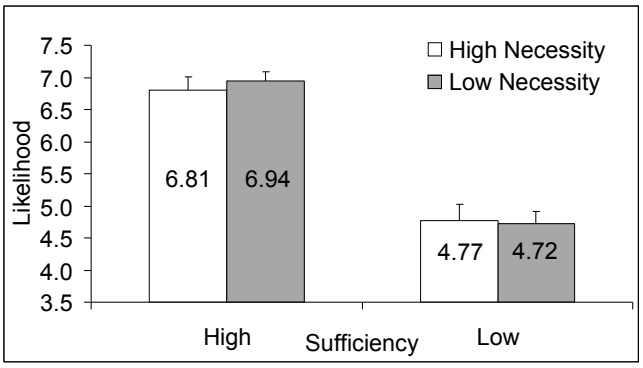

Figure 3: Mean Ratings of Likelihood in Study 2 as a Function of the Sufficiency and Necessity of the Cause.

IV. Discussion

The interpretation of conditional statements is an ongoing topic of inquiry in philosophy and psychology. One view is that when people understand a conditional statement, they consider only those two possibilities in which the conditional's antecedent holds true: in one possibility, both the antecedent and consequent hold true; in the other possibility, the antecedent holds true but the consequent is false. The relative plausibility of these two states of affairs determines the sensibility of the conditional statement: a conditional is more sensible when the probability of the consequent given the antecedent is high, and less sensible when that probability is low.

On this view, when people comprehend conditionals, they do not consider states of affairs in which the antecedent is false. As summarized by Over and Evans (2003), "A conditional causes hypothetical thinking about the *p* [true antecedent] possibility, without any representation of the *not-p* [false antecedent] possibility."

The relative probability of the consequent given that antecedent is no doubt important for the assessment of conditionals. However, the data presented here demonstrate that when people evaluate conditionals, they are sensitive to the necessity of the antecedent as well as its sufficiency. That is, when people evaluate conditionals they do consider states of affairs in which the antecedent is false.

The re-analysis of data from Evans, Over and Handley (2003, Experiment 2) demonstrated that when people evaluate conditional statements when provided with frequency data, they are sensitive to the necessity of the antecedent. That study offered a contrast between two experimental conditions that were equated on the sufficiency of the antecedent, but different in its necessity. The data showed that participants rated conditionals as stronger when the antecedent was more necessary for the consequent. A follow-up study (Section II) replicated these results.

In our studies of causal conditionals (Section III), we departed from the method employed in previous studies because we did not present participants with the frequency distribution of the contingencies of events referred to in the antecedent and consequent of the conditionals. Instead, we used a norming study to select conditionals that independently varied on the necessity and sufficiency of the cause. We could thus examine which sort of information is spontaneously considered when evaluating a conditional.

Our studies revealed that when people assess the reasonableness of causal conditionals they may be sensitive to the necessity of the cause. These findings corroborate results reported by Newstead, Ellis, Evans, and Dennis (1997, Experiment 1). They reported that when participants assessed causal conditionals, they rated cases of ¬cause and effect as either inconsistent with the conditional (48% of the time) or irrelevant to the conditional (52% of the time), but they never rated such cases as consistent with the conditional. Hence, when individuals can readily think of cases in which the effect occurs in the absence of the cause, they should judge conditionals as less reasonable than when they cannot think of such cases. Other studies have shown that when people assess the strength of a causal relationship, they are sensitive to those two cells of the frequency table in which the cause is absent, and, they consider cases of ¬cause and effect as constituting evidence *against* a causal relationship (see e.g., Wasserman, Elek, Chatlosh, & Baker, 1993; Mandel & Lehman, 1998).

We found that the impact of the antecedent's necessity on the assessment of conditionals depends on the sufficiency of the cause, as well as the task required of participants. The necessity of the cause had a strong effect when its sufficiency was high, but not when its sufficiency was low (see Figures 1 and 2). When sufficiency is high, a conditional is asserting a causal relationship that is undeniable, and so the assertion might be understood as being about the necessity of the cause. However, when sufficiency is low, the conditional is unreasonable, and so participants are less sensitive to the cause's necessity. We also found that the task instructions affect the relative impact of necessity: ratings of reasonableness were affected by the necessity of the cause, but ratings of likelihood were not.

In summary, we have shown that when people assess conditionals, they are sensitive to the availability of alternative causes. Further research may very well demonstrate the that the assessment of conditionals is sensitive to additional factors that affect the strength of contingencies, e.g., ΔP, P(effect), P(cause), etc' (cf., White, 2004, for the independent contributions of such factors on judgments of causal contingencies). Our results clearly demonstrate that people's assessment of conditionals cannot be explained by appealing to conditional probability alone.

References

Adams, E. W. (1975). *The Logic of Conditionals*. Dordrecht: Reidel.

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

Hadjichristidis, C., Stevenson, R. J., Over, D. E., Sloman, S. A., & Evans, J. S. B. T. (2001). *On the evaluation of If p then q conditionals.* Paper presented at the Twenty-Third Annual Conference of the Cognitive Science Society, Edinburgh, Scotland.

Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General, 127*, 269-285.

Newstead, S. E., Ellis, M. C., Evans, J. S. B. T., & Dennis, I. (1997). Conditional reasoning with realistic material. *Thinking & Reasoning, 3*, 49-76.

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

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

Rips, L. J, (2002). Circular reasoning. *Cognitive Science, 26*, 767-795.

Wasserman, F.A., Elek, S.M., Chatlosh, D.L., & Baker, A.G. (1993). Rating causal relations: The role of probability in Judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 174-188.

White, P. A. (2004). Causal judgment from contingency information: a systematic test of the pCI rule. *Memory and Cognition, 32*, 353-368.