Interpretations of Conditional and Causal Statements

Alice McEleney (alice.mceleney@unn.ac.uk)
Psychology Division, Northumbria University
Northumberland Building, Newcastle upon Tyne, NE1 8ST, UK

Abstract

Two experiments compared people's interpretations of indicative conditionals (e.g., if the shape is blue, then it's a square), subjunctive conditionals (e.g., if the shape were blue, then it would be a square), causation statements (e.g., the shape being blue causes it to be a square) and prevention statements (e.g., the shape being blue prevents it from being a square). In the first experiment, participants rated the extent to which the statements were true of arrays of coloured shapes. In the second experiment, participants constructed their own arrays of coloured shapes to show the statements to be true or false. The results suggest that people tend to interpret all four statements in the same way and they often interpret them extensionally rather than probabilistically. The implications of the findings for theories of conditional and causal thinking are discussed.

Keywords: conditional; causal; thinking; reasoning

Introduction

People often think about conditional or 'if…then…' relations, such as 'if you eat a lot of fat, then you have high cholesterol' (an indicative conditional) or 'if you were to eat a lot of fat, then you would have high cholesterol' (a subjunctive conditional). Conditionals often express causal relations, and so they are closely linked to causal statements such as 'eating a lot of fat causes high cholesterol' (a causation statement) and 'not eating a lot of fat prevents high cholesterol' (a prevention statement). Conditional and causal thinking are essential to reasoning and decision making, yet despite centuries of philosophical debate and decades of psychological research, a very fundamental question remains controversial: what do ordinary people take conditional and causal statements to mean? This paper attempts to shed light on how people interpret indicative conditionals, subjunctive conditionals, causation statements and prevention statements.

The widespread unproblematic use of conditional and causal constructions in everyday conversation shows that ordinary people share an implicit understanding of their meanings. The interpretation of particular conditional and causal assertions depends very much on their content and context (e.g., Cummins, 1995; Thompson, 1994). However, people readily understand statements with neutral contents and contexts, such as *If A then B* or *A causes B*, which suggests that conditional and causal statements have 'basic' meanings which allow people to understand them even when their content and context are impoverished (Goldvarg & Johnson-Laird, 2001; Johnson-Laird & Byrne, 2002). What are the basic meanings of conditional and causal assertions? Among psychologists there are currently two main competing views of how people interpret conditional and causal statements, the extensional view and the probabilistic view.

According to the extensional view, conditional and causal statements refer to sets of possibilities, which are represented as mental models (e.g., Goldvarg & Johnson-Laird, 2001; Girotto & Johnson-Laird, 2004; Johnson-Laird & Byrne, 2002). On this account, a basic indicative or subjunctive conditional *if A then B* or causation statement *A causes B* refers to the following set of possibilities:

$$a \quad b$$
$$\neg a \neg b$$
$$\neg a \quad b$$

where each line represents a different possibility, a denotes that A is satisfied, b denotes that B is satisfied and $\neg$ denotes negation (e.g., $\neg a$ denotes that A is not satisfied). This implies that *if A then B* and *A causes B* both mean that A is sufficient (but not necessary) for B and B is necessary (but not sufficient) for A. Similarly, a basic prevention statement *A prevents B* would refer to the following set of possibilities:

$$a \neg b$$
$$\neg a \quad b$$
$$\neg a \neg b$$

in which A is sufficient (but not necessary) to prevent B and $\neg B$ is necessary (but not sufficient) for A. People do not necessarily represent all of the above possibilities explicitly when they think about conditional or causal statements. For example, instead of the fully explicit 'conditional' representation shown above, they may construct a 'biconditional' representation, which includes just the first two possibilities explicitly, or a 'conjunction' representation, which includes just the first (e.g., Barouillet, Grosset & Lecas, 2000; Johnson-Laird & Byrne, 2002).

In contrast, theorists who take a probabilistic view argue that conditional and causal statements refer to probabilistic relations, rather than sets of possibilities (e.g., Cheng, 1997; Evans, Handley & Over, 2003; Oberauer & Wilhelm. 2003). For example, Evans and collaborators argue that a conditional *if A then B* is interpreted as meaning that the conditional probability of B given A is high. On this account, A is neither sufficient nor necessary for B and B is neither necessary nor sufficient for A. Instead, B is probable given A. Similarly, Cheng (1997) argues that *A causes B* means that the probability of B given A is noticeably higher than the probability of B given $\neg A$, and *A prevents B* means that the probability of B given A is noticeably lower than the probability of B given $\neg A$.

One aim of the experiments reported in this paper is to contribute to the debate on whether basic conditional and causal statements are typically interpreted extensionally or probabilistically. A second aim is to determine whether indicative and subjunctive conditionals and causation and prevention statements are interpreted in the same way or in different ways. According to the extensional account developed by Johnson-Laird and collaborators, *if A then B*, *A causes B* and *A prevents not-B* all refer to the same set of possibilities, as described above. In contrast, probabilistic accounts suggest that conditional and causal statements have different meanings (Over, Hadjichristidis, Evans, Handley & Sloman, 2005). Moreover, recent research on the meanings of 'cause' and 'prevent' suggests that *A causes B* is not directly equivalent to *A prevents not-B* (Walsh & Sloman, 2005). To my knowledge, no previously published studies have directly compared people's interpretations of conditional and causal statements.

A major limitation of previous research on the meaning of conditionals is that the vast majority of studies have used some version of one of only three tasks. One of those is the truth table task, which is based on the notion in propositional logic that the truth or falsity of a conditional *if p then q* depends on the truth or falsity of its component propositions, *p* and *q*, which can be shown in a truth table. As shown in Table 1, the truth table for the material conditional shows that *if p then q* is false if and only if *p* is satisfied and *q* is not. In truth table tasks, participants' interpretations of conditionals are assessed by requiring them to classify truth table cases as true, false or irrelevant.

Table 1: Truth table for the material conditional *if p then q*

Case	Truth value
TT (pq)	T
TF (p¬q)	F
FT (¬pq)	T
FF (¬p¬q)	T

Note: T = true, F = false

The two other commonly used tasks are Wason's (1966) selection task, in which participants have to select the cases they need to examine in order to test whether a conditional is true; and inference tasks, in which participants are given a conditional and a second premise and asked to generate or evaluate a conclusion. Recently, some researchers have also used probabilistic truth table tasks, in which participants judge the probability of conditionals given frequency information about the truth table cases (see Evans & Over, 2004, for reviews of all of these methods).

All of these methods have their drawbacks. The selection task and inference task are both very indirect ways of assessing people's mental representations of conditionals, as shown by researchers' disagreement about how to interpret people's responses (e.g., Braine & O'Brien, 1998; Evans & Over, 1996). The truth table approach offers a more direct

measure of the interpretation of conditionals, but it makes some contentious assumptions. Asking participants to judge whether each of the four logically possible cases either confirms or falsifies a conditional assumes that people interpret conditionals as referring to sets of possible cases and may bias them towards making this kind of interpretation. Probabilistic truth table tasks avoid biasing people towards extensional interpretations, but they may instead bias them towards a probabilistic interpretation.

The experiments reported in this paper avoid these problems by using the methods reported by Evans, Ellis and Newstead (1996). In the first experiment, participants were presented with statements about the colour and shape of symbols (e.g., if the shape is blue, then it's a square) and asked to rate the extent to which the statements were true of accompanying arrays of coloured shapes containing varying ratios of TT (e.g., blue square) to TF (e.g., blue circle) cases. In the second experiment, participants constructed their own arrays of coloured shapes to show the statements to be true or false. The use of these methods was intended to avoid biasing participants towards any particular kind of interpretation of the statements.

Evans et al (1996) report that participants did not consider indicative conditionals to be falsified when a small number of TF cases was included in an array and they often included TF cases when constructing arrays to show the statements to be true. These findings are frequently referred to as evidence for probabilistic interpretations of conditionals (e.g., Evans et al, 2003; Evans & Over, 2004). However, as only the mean responses were reported, it is unclear how individual participants were typically interpreting the statements. This paper reports both the mean responses and patterns of individual responses to these tasks.

Experiment 1

Method

Participants
32 students and members of the general public (17 female, 15 male) aged 18 to 47 (mean age 20) were paid approximately £5 for their participation. None of the participants had any training in logic.

Procedure
Participants were tested on computers. They were given written instructions stating that they were going to be presented with statements, each accompanied by an array of coloured shapes, and that their task would be to indicate how true they felt each statement to be with respect to the accompanying array. They were instructed that they were to rate this on a scale of +5 to −5, where +5 is 'absolutely true', -5 is 'absolutely false' and 0 is 'can't tell'. 11 adjacent keys on the top row of the keyboard were labeled from +5 to −5 and participants had to press one of those keys to record their response.

Participants worked through the trials at their own pace. For each trial, the computer displayed the statement at the

top of the screen with a 6 by 6 array of 36 coloured shapes underneath. At the bottom of the screen was a scale from +5 to –5, with +5 marked 'absolutely true', -5 marked 'absolutely false' and 0 marked 'can't tell'. Each participant completed the trials in a different random order.

Materials and design

There were four types of statement: indicative conditional (e.g., if the shape is blue, then it's a square), subjunctive conditional (e.g., if the shape were blue, then it would be a square), causation statement (e.g., the shape being blue causes it to be square) and prevention statement (e.g., the shape being blue prevents it from being a square)[1]. Four shapes (square, circle, triangle and diamond) and four colours (blue, yellow, black and white) were used in the statements. The colours and shapes were randomly assigned to each trial.

Each of the statements was evaluated with respect to five different arrays of coloured shapes. All of the arrays contained 9 FF cases and 9 FT cases, with the frequencies of TT and TF cases manipulated as shown in Table 2.

Table 2: Distribution of truth table cases in each array in Experiment 1

Array	TT	TF	FT	FF
18:0	18	0	9	9
17:1	17	1	9	9
9:9	9	9	9	9
1:17	1	17	9	9
0:18	0	18	9	9

In some cases, more than one symbol could be used to represent a truth case (e.g., for the statement 'if the shape is blue, then it's a square', TF cases would include blue circles, triangles and diamonds). In such cases, the symbols used were selected randomly from the possible symbols. Examples of the four cases for the four different statement types are as follows: for 'if the shape is blue, then it's a square', 'if the shape were blue, then it would be a square' and 'the shape being blue causes it to be a square', TT = blue square, TF = blue circle, FT= yellow square, FF = yellow circle; for 'the shape being blue prevents it from being a square', TT = blue circle, TF = blue square, FT = yellow circle, FF = yellow square. The positions of the

---

[1] In addition, the polarity of the antecedents and consequents of the statements was manipulated so that there were four forms of each statement type: affirmative-affirmative (e.g., if the shape is blue, then it's a square), affirmative-negative (e.g., if the shape is blue, then it's not a square), negative-affirmative (e.g., if the shape is not blue, then it's a square) and negative-negative (e.g., if the shape is not blue, then it's not a square). This resulted in 16 statements in total. Due to space limitations, the results reported here are for affirmative-affirmative assertions only (for further details, see McEleney, 2005).

different symbols in the array were randomly determined for each trial.

Results and Discussion

Are the statements interpreted in the same way or in different ways?

Figure 1 shows participants' mean ratings of the truth of the statements, given the different ratios of TT to TF cases. A four (statement: indicative vs subjunctive vs causation vs prevention) by five (TT:TF ratio: 18:0 vs 17:1 vs 9:9 vs 1:17 vs 0:18) ANOVA on participants' ratings showed that there was a main effect of ratio, $F, 1, 31 = 1230.76$, $p < 0.01$, but no main effect of statement, $F, 1, 31 = 0.25$, $p = 0.62$, and no interaction between statement and ratio, $F, 1, 31 = 0.13$, $p , 0.72$. The findings suggest that people interpret all four statements in the same way.
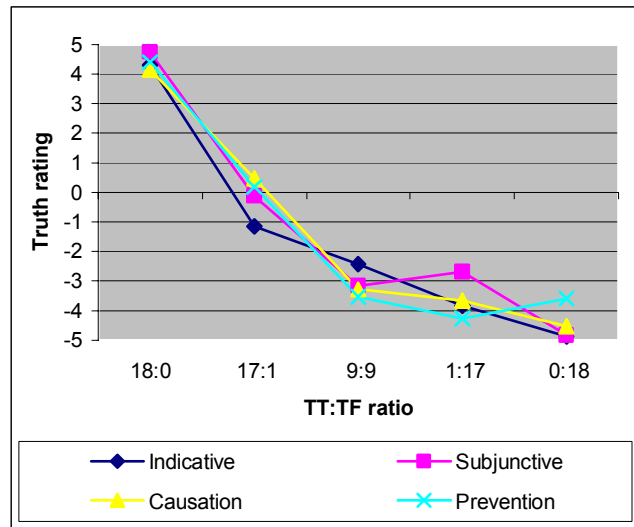

Figure 1: Mean ratings of statements for different TT:TF ratios in Experiment 1

Are the statements interpreted extensionally or probabilistically?

Post-hoc comparisons showed that ratings for the 18:0 ratio were higher than for any of the other ratios; ratings for the 17:1 ratio were higher than the 9:9, 1:17 and 0:18 ratios; and ratings for the 9:9 ratio were higher than the 0:18 ratio (all $ps < 0.05$). This general pattern of a decrease in mean truth ratings as the ratio of TT:TF cases decreases is consistent with the findings of Evans et al. (1996) with indicative conditionals and seems to support the idea that conditional and causal statements are interpreted in a probabilistic way.

However, examination of individual participants' responses suggests that only about half of the participants interpreted the statements probabilistically, whereas the other half interpreted them extensionally. As shown in Figure 2, the overwhelming majority of participants rated all four statements as 'absolutely true' of arrays containing 18 TT cases and no TF cases. For the 18:0 array, the modal

response for all four statements was +5 (absolutely true), which was used for 86% of responses (87% of indicative, 91% of subjunctive, 84% of causation, 81% of prevention). As shown in Figure 3, approximately half the participants also rated the statements as true (although not necessarily 'absolutely true') of arrays containing 17 TT cases and one TF case, whereas the other half considered the statements to be falsified by the presence of just one TF case. For the 17:1 ratio, the modal response for all four statements was -5 (absolutely false), which was used for 39% of responses (50% of indicative, 41% of subjunctive, 37% of causation and 28% of prevention). The second most common response was +5 (absolutely true), which was used for 23% of all responses (12% of indicative, 16% of subjunctive, 31% of causation and 31% of prevention) and the third most common response was +4, which was used for 16% (6% indicative, 9% of subjunctive, 28% of causation and 22% of prevention).
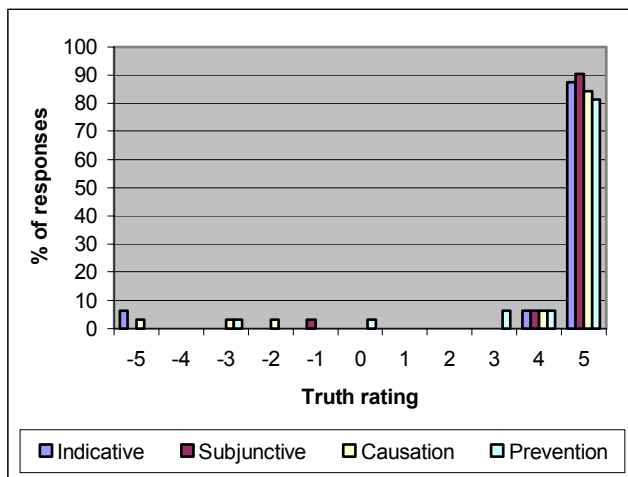

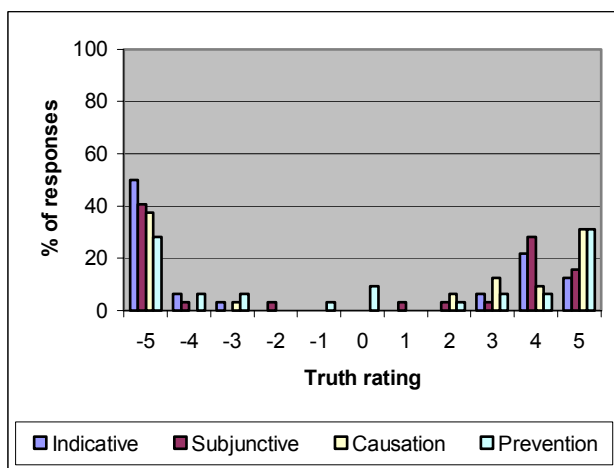Figure 2: Percentages of each response for the 18:0 arrays in Experiment 1

Figure 3: Percentages of each response for the 17:1 arrays in Experiment 1

Experiment 2

Method

Participants
24 students and members of the general public (19 female, 5 male) aged 18-47 (mean age 23) were paid approximately £5 for their participation. None of the participants had any training in logic.

Procedure
Participants were given written instructions stating that they were going to be presented with statements about the colour and shape of symbols, each accompanied by an empty 6 by 6 grid and that their task would be to fill the grid with coloured shapes to show the statement to be either true or false with respect to the appearance of the grid. The materials were presented in a booklet with one trial on each page. For each trial, the statement was displayed at the top of the page with the empty grid below it. Participants were provided with a blue pen and a red pen, which they used to draw coloured shapes in the grid.

Materials and design
The same statements were used as in Experiment 1[2]. For each of the statements, there was one trial in which it had to be verified and one in which it had to be falsified. Two shapes (square, circle) and two colours (blue, red) were used. The colours and shapes were randomly assigned to each trial. Each participant completed the trials in a different random order.

Results and Discussion

Are the statements interpreted in the same way or in different ways?
Figures 4 and 5 show the mean numbers of TT, TF, FT and FF cases participants included to verify and falsify the statements, respectively. Four (statement: indicative vs subjunctive vs causation vs prevention) by two (task: verify vs falsify) ANOVAs revealed no effects of statement on the numbers of TT, TF, FT or FF cases (TT: $F_{1, 23} = 3.03$, $p = 0.09$; TF: $F_{1, 23} = 0.72$, $p = 0.54$; FT: $F_{1, 23} = 0.65$, $p = 0.58$; FF: $F_{1, 23} = 0.99$, $p = 0.76$), whereas there were effects of task on all four measures (TT: $F_{1, 23} = 57.31$, $p < 0.01$; TF: $F_{1, 23} = 102.86$, $p < 0.01$; FT: $F_{1, 23} = 4.87$, $p = 0.04$; FF: $F_{1, 23} = 13.74$, $p < 0.01$). There were no interactions between statement and task for TT, TF or FT cases (TT: $F_{1, 23} = 3.24$, $p = 0.08$; TF: $F_{1, 23} = 0.7$, $p = 0.55$; FT: $F_{1, 23} = 1.85$, $p = 0.14$), but there was an interaction for FF cases, $F_{1, 23} = 8.7$, $p < 0.01$. Post-hoc tests showed that there were significantly more FF cases for verification than falsification for causation and prevention

[2] As in Experiment 1, there were four versions of each statement, based on varying the polarity of the antecedent and consequent. Due to space limitations, only the results for affirmative-affirmative statements are reported here (see McEleney, 2005, for further details).

statements (both ps < 0.01), whereas this difference was not significant for indicative or subjunctive conditionals. Overall, the findings suggest that people tend to interpret all four statements in the same way.
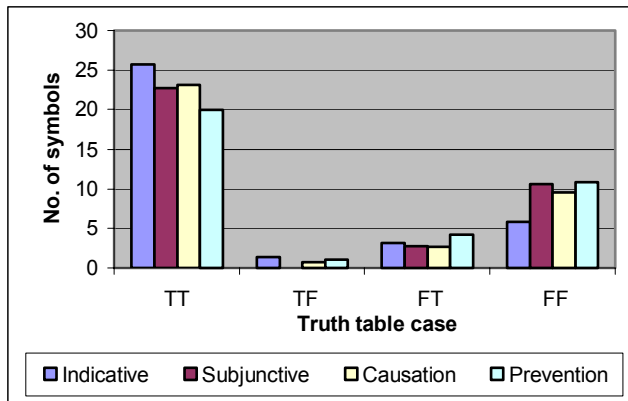

Figure 4: Mean numbers of symbols included in the verification task of Experiment 2
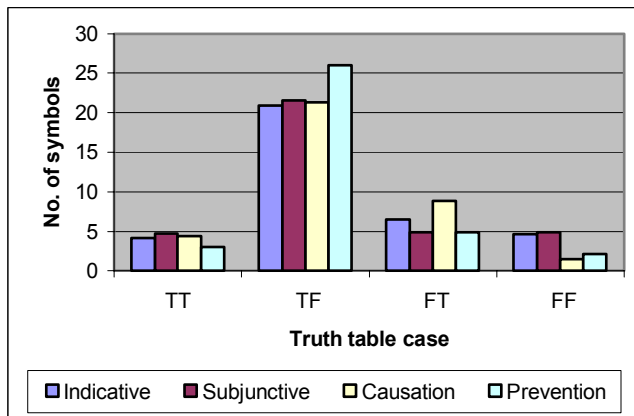

Figure 5: Mean numbers of symbols included in the falsification task in Experiment 2

Are the statements interpreted extensionally or probabilistically?

As shown in figure 4, the mean numbers of TF cases included in the verification task were 1.33 for indicative conditionals, 0 for subjunctive conditionals, 0.75 for causation statements and 1 for prevention statements. These findings are consistent with those of Evans et al. (1996) for indicative conditionals and suggest that participants may be interpreting the indicative conditionals, causation statements and prevention statements in a probabilistic way.

However, as in Experiment 1, examination of participants' individual responses suggests a different interpretation. Only one participant included any TF cases to verify an indicative conditional, only one included any to verify a causation statement and only one included any to verify a prevention statement. The numbers of TF cases included in these trials were very high (32, 18 and 24

respectively), which suggests that they were not included in order to set the probability of the consequent given the antecedent to be high, but less than 1 (pace Evans et al., 1996; Evans & Over, 2004). It appears more plausible that in these three cases, participants misunderstood the task or misread the statements and included these TF cases by mistake. The results suggest that the kind of situation people expect to hold when a conditional or causal statement is true is one in which TF cases do not occur.

Participants' patterns of symbol selections in the verification task were classified into categories based on whether they included TT cases only (consistent with a 'conjunction' interpretation), TT and TF cases only (consistent with a 'biconditional' interpretations), TT, FF and FT cases only (consistent with a 'conditional' interpretation) and other patterns of responses, as shown in Table 3.

Table 3: Percentages of participants constructing arrays consistent with 'conjunction' (TT), 'biconditional' (TT & FF), 'conditional' (TT, FF & FT) and other interpretations in Experiment 2

	TT	TT & FF	TT, FF & FT	Other
Indicative	37	29	29	4
Subjunctive	33	29	37	0
Causation	33	25	33	8
Prevention	21	29	42	8
Mean	**32**	**29**	**36**	**5**

The results show that for all four statements, the overwhelming majority of participants constructed arrays that were consistent with either a 'conjunction', 'biconditional' or 'conditional' interpretation, with very few participants constructing arrays containing other combinations of truth table cases.

Discussion

The findings of the two experiments reported here suggest that people tend to interpret basic indicative conditionals, subjunctive conditionals, causation statements and prevention statements in the same way. Participants in the first experiment made very similar ratings of the extent to which the different kinds of statements were true of arrays of coloured shapes. Participants in the second experiment constructed very similar arrays of coloured shapes to verify or falsify the different kinds of statement.

The findings also suggest that people often interpret all four statements in an extensional rather than probabilistic way (pace Evans et al., 1996; 2003; Evans & Over, 2004). In the first experiment, as would be expected, the overwhelming majority of participants rated all four statements as 'absolutely true' of arrays of coloured shapes containing 18 TT cases and no TF cases. Approximately half the participants also rated the statements as true (although not necessarily 'absolutely true') of arrays

containing 17 TT cases and one TF case, suggesting a probabilistic interpretation in which the antecedent increases the likelihood of the consequent, rather than determining its occurrence. However, the other half of the participants considered the statements to be falsified by the presence of just one TF case, suggesting an extensional interpretation in which TF cases are impossible rather than just improbable. The finding of individual differences in participants' interpretations is consistent with evidence reported by Evans et al. (2003) and Oberauer and Wilhelm (2003) for indicative conditionals. However, the results reported here are not consistent with their argument that probabilistic interpretations are more common than extensional ones.

The second experiment provided stronger evidence for extensional interpretations. When constructing their own arrays to show the statements to be true, participants extremely rarely included any TF cases. Instead, the vast majority of participants constructed arrays consistent with one of three extensional interpretations: conjunction (TT cases only), biconditional (TT and FF cases only) or conditional (TT, FF and FT cases only).

Of course, the experiments reported here do not examine all aspects of people's interpretations of conditional and causal statements. For example, although indicative and subjunctive conditionals may be interpreted as consistent with the same sets of possibilities, people's initial mental models of indicative and subjunctive conditionals may be different, with the result that people make different inferences from them (e.g., Thompson & Byrne, 2002). Similarly, a causation statement may refer to the same set of possibilities as a conditional, but the causation statement may also convey that the cause precedes the effect in time, whereas the conditional may not specify the temporal order of the antecedent and consequent (Goldvarg & Johnson-Laird, 2001). Another possibility is that the meaning of causation includes understanding of the mechanisms by which the cause produces its effect (e.g., Schultz, 1982; White, 1989).

The experiments reported here only examined 'basic' conditional and causal statements, that is, those with '…neutral content that is as independent as possible from context and background knowledge...' (Johnson-Laird & Byrne, 2002, p. 648). Although it is important to understand how people interpret such statements, further research is needed to determine whether the findings also apply to more realistic statements used in context (see McEleney, 2005).

Acknowledgments

References

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

Braine, M. D. S. & O'Brien, D. P. (Eds.) (1998). *Mental Logic*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367-405.

Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory and Cognition, 23 (5),*646-658.

Evans, J. St. B. T., Ellis, C. E. & Newstead, S. E. (1996). On the mental representation of conditional sentences. *Quarterly Journal of Experimental Psychology, 49A, 4,* 1086-1114.

Evans, J. St. B. T., Handley, S. J. & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology, 29, 2,* 321-335.

Evans, J. St. B. T., & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review, 103,* 356-363.

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

Girotto, V. & Johnson-Laird, P. N. (2004). The probability of conditionals. *Psychologia, 47, 4,* 207-255.

Goldvarg, E. & Johnson-Laird, P. N. (2001). Naive causality: a mental model theory of causal meaning and reasoning. *Cognitive Science, 25, 4,* 565-610.

Johnson-Laird, P. N. & Byrne, R. M. J. (2002). Conditionals: A Theory of meaning, pragmatics and inference. *Psychological Review, 109, 4,* 646-678.

McEleney, A. (2005). Mental representations of conditional and causal statements. *Manuscript in preparation.*

Oberauer, K. & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology, 29, 4,* 680-693.

Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J. & Sloman, S. A. (2005). The probability of ordinary indicative conditionals. *Manuscript submitted for publication.*

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47,* 1-51.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition, 22 (6),* 742-758.

Thompson, V. A. & Byrne, R. M. J. (2002). Reasoning counterfactually: Making inferences about things that didn't happen. *Journal of Experimental Psychology, 28, 6,* 1154-1170.

Walsh, C. R. & Sloman, S. A. (2005). The meaning of cause and prevent: the role of causal mechanism. *This volume.*

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

White, P. E. (1989). A theory of causal processing. *British Journal of Psychology, 80,* 431-454.