A Meta-Analysis of Conditional Reasoning

Marco Ragni (ragni@cs.uni-freiburg.de)
Cognitive Computation Lab, Technical Faculty, University of Freiburg, Germany

Hannah Dames (dames@cs.uni-freiburg.de)
Cognitive Computation Lab, Technical Faculty, University of Freiburg, Germany

Phil Johnson-Laird (phil@princeton.edu)
Princeton University, Princeton NJ 08540, USA
New York University, New York, NY 10003, USA

Abstract

Conditional premises are assertions with "if", e.g., *If I have measles, then I have fever.* They provide a connection between different propositions and can express causal relations. Conditional inferences often comprise conditional and categorical assertions, e.g., such as modus tollens: *If I have measles, then I have fever*; *I don't have fever*; So, *I don't have measles*. Most research has concerned four sorts of conditional inference, examining them separately. Only a few studies have focused on the patterns over the four sorts of inference (e.g., Oberauer, 2006). Our meta-analysis was of 39 experiments (with 2378 participants) that reported these patterns. It showed that a version of the mental model theory best fits the results when participants produced their own conclusions or evaluated a given conclusion, whereas the suppositional theory provided the best fit when participants chose a conclusion from a list of options.

Keywords: Conditional reasoning; Information; Mental models; Suppositions; Probabilities

Introduction

Conditionals allow humans to describe hypotheses, causal dependencies, diagnoses, and other relations between pieces of information. They tend to be expressed in assertions of the sort, *If A then B*, where *A* and *B* are sensible clauses in natural language, which may be simple or compound, i.e., contain sentential connectives of their own. Classical studies of reasoning use inferences consisting of a conditional and an additional categorical premise, as in:

If he has measles, then he has a fever.	(A conditional)
He has measles.	(A categorical)
What, if anything, follows?	

Almost all reasoners infer: he has a fever (see, e.g., Oberauer, 2006). This sort of inference is the first of four sorts (called modus ponens) as shown below with their conventional names and abbreviations. These four sorts of inferences share a conditional premise, but have different categorical premises and so yield different conclusions. We use '∴' to preface conclusions. For the given conditional *If A then B* and we have the respective categorical premise and conclusion:

A.	∴ B.	(Modus Ponens: MP);
B.	∴ A.	(Affirmation of Consequent: AC);
Not A.	∴ Not B.	(Denial of Antecedent: DA);
Not B.	∴ Not A.	(Modus Tollens: MT).

In classical logic, MP and MT are valid, i.e., given that their premises are true, their conclusions are also true. DA and AC are valid only if the conditional has a biconditional interpretation, equivalent to: *If and only if A then B*. The biconditional inference pattern often occurs in studies (e.g., Oaksford & Chater, 2007, p. 140). While most studies report the response frequencies of the four sorts of inference, they do not give any information about the inference patterns of each participant, such as the number of participants who drew only MP and MT inferences. A few studies, however, do report the frequencies of these inference patterns over the four sorts of premises (e.g., Oberauer, 2006, Barrouillet, Gauffroy, & Lecas, 2008; Evans & Over, 2004). In what follows, we also show that the separate overall frequencies of each of the four inferences yields a misleading picture of the process of reasoning.

Psychologists have proposed five main sorts of theory of conditional reasoning: theories based on formal logic, on mental models, on suppositions, on dual-processes with suppositions, and on probabilities. In what follows, we briefly review them.

Theories based on formal logic (e.g., Rips, 1994) postulate that the mind contains a formal rule for MP but no rule for MT. Thus, its inference depends on the three steps: i) make a supposition of the conditional's *if*-clause, *A*; ii) the rule for MP yields *B*; iii) its conjunction with the categorical premise *not B* is a self-contradiction. As a consequence, one can deny the supposition to yield the conclusion: *not A*. Readers should note that formal rule theories are not included in the meta-analysis, because their processes have never been formulated as multinomial processing trees (see below).

The theory of mental models. The theory of mental models (e.g., Johnson-Laird & Byrne, 2002) postulates two systems of reasoning: intuitive and deliberative. The first system is the intuitive process in which reasoners rely on mental models that represent only what is true. Hence, for the conditional *If A then B* a reasoner forms the following, mental models:

A B
. . .

The first model represents the the possibility in which *A*, and thus *B,* both hold. The second model – the ellipsis – stands for the possibility in which *A* is not possible. MP follows at once from these models given the premise *A*. MT, however, does not. It calls for the second process, which is deliberative and in which mental models, including the ellipsis, are fleshed out into fully explicit ones:

A	B
¬ A	¬ B
¬ A	B

This process yields the possibilities in the order above (see, e.g., Barrouillet, Grosset, & Lecas, 2000). The categorical premise, *not B*, now yields the conclusion, *not A*. Hence, MP should be easier than MT. The model theory explains the discrepancy between human reasoning and logically correct inferences as a result of reliance on intuitive mental models. A further relevant prediction is that MT is easier with a biconditional, which has only two fully explicit models, than with a conditional, which has three (Johnson-Laird, Byrne & Schaeken, 1992).

The directional model theory. A variant of the mental model theory – the directional model theory – introduces the assumption that inferences are easier from the *if*-clause to the *then*-clause of a conditional than in the opposite direction (Evans, 1993; Oberauer, 2006). Hence, it follows that an MP-inference is easier than an AC-inference. If the inferences are based on biconditionals, a DA-inference is easier than an MT-inference.

The suppositional theory. The suppositional theory (Evans & Over, 2004) also assumes that two cognitive systems underlie conditional reasoning: a heuristic, automatic, and fast system (1), and an analytical, controlled, and slow system (2). In later versions, the theory assumes that conditionals have a probabilistic interpretation in which there is high conditional probability of the *then*-clause given the *if*-clause. System 1 takes background knowledge, context, and the content of the premises into account. System 2, however, can focus on the information given in the premises and principles of deductive reasoning. Oberauer (2006) formulated two versions of the theory in order to fit data. In the *sequential* version, system 1 operates first and then system 2 generates a conclusion on the basis of this outcome. In the *exclusive* version, only one of the two systems operates on a given problem, i.e., they are mutually exclusive.

The dual process theory of suppositions. There is a family of dual-process theories (see, e.g., Evans, 2008 for a review). However, one prominent version is similar to the suppositional theory (Verschueren, Schaeken, & d'Ydewalle, 2005) because it has the same system 1. But, in this version, system 2 makes inferences using mental models in the same way as the model theory does, instead of the proof-based system in the suppositional theory. The two systems are assumed to be mutually exclusive.

The probabilistic theory. The probabilistic theory shares a general assumption of the suppositional theory, that is, that conditionals are interpreted in terms of subjective conditional probabilities (Oaksford, Chater, & Larkin, 2000). Conditionals have a high conditional probability of the *then*-clause given the *if*-clause. The process for drawing inferences, however, differs from the suppositional theory. Reasoners accept a conclusion based on its subjective conditional probability given the minor premise. This theory was not included in this meta-analysis, because its parameters for MP and MT inferences have only the 'exceptions' parameter (1 – P(*then*-clause | *if*-clause)) in common, which is close to zero. De facto. the theory treats the four inferences as independent, and Oaksford et al. (2000) do not report the frequencies of the patterns of inference (cf. Singmann et al., 2016).

Table 1 summarizes the predictions of the three main sorts of theory. But, as we will see, our meta-analysis was able to examine four theories.

Table 1: Three predictions that discriminate about theories based on logic, suppositions, and mental models.

	Logical	**Suppositional**	**Mental models**
The meaning of *If A then C*:			
1. implies the possibilities: *A C*, *¬A ¬C*, *¬A C*	-	-	+
2. implies that only cases of *A* are relevant to verification	-	+	+
3. implies that MT with a biconditional is easier than with a conditional	-	-	+

Note: + indicates that a theory makes the prediction, and - indicates that it does not.

Prior to the work of Oberauer (2006), theories tended to consider individual sorts of inference, whereas he formalized versions of theories with multinomial processing trees – henceforth, we refer to them as 'trees' – for all 16 possible patterns of responses to the four sorts of inference (MP, AC, DA, and MT). Every reasoner is bound to yield of $2^4 = 16$ possible patterns of responses for the four sorts of inference. These patterns give a more accurate understanding of the cognitive processes underlying conditional reasoning than analyzing the four sorts of inferences separately. As we will see later, the four sorts of inferences are not drawn independently from each other.

Oberauer's trees included all the cognitive processes leading from inputs to the 16 leaves that represented the responses. He added a single fixed guessing component to each of the trees and evaluated the goodness of fit using G-tests. In the following, we use the formulations of Oberauer's (2006) trees for the original model theory, the directional model theory, the suppositional theory (sequential and exclusive), and the dual-process theory (Verschueren et al., 2005).

The main goals of our analyses were (1) to analyze the four sorts of inference in three types of experimental task: the production of conclusions, the choice of conclusions from options, and the evaluations of single given conclusions; and (2) to carry out a new sort of meta-analysis that includes assessments of the reliability of the data, of the inter-dependence of conclusions over the four sorts of inference, and the goodness of fit of the different theories. Finally, the paper discusses the implications of its results for the various theories.

Three Types of Reasoning Task

Studies of conditional reasoning have used three main tasks (for an overview see Schroyens & Schaeken, in preparation). In the *production* task, the participants are given the premises and asked to state what, if anything, follows from them, i.e.., what must be true given that the premises are true. In the *option* task, they are asked to choose such a conclusion from a set of multiple options, which usually include one for "nothing follows". In the *evaluation* task, they are presented with the premises and a single putative conclusion, and they evaluate whether or not it follows from the premises. These three tasks are likely to call on different mental processes, e.g., reasoners can work backwards from a given conclusion in the evaluation task, but they have to formulate or guess a conclusion to carry out the production task (see Schroyens & Schaeken, in preparation). These authors were the first to show that the different sorts of task affect the conclusions that individuals draw (e.g., Schroyens et al., 2001; Schroyens & Schaeken, in preparation). They formulated the following predictions about differences among the three sorts of task: more conjunctive conclusions should occur in the production task, and fewer selections of fallacious conclusions for AC and DA should occur in the option task. We therefore follow Schroyens and Schaeken and conducted separate analyses of performance for the three types of task.

Meta-analysis

The meta-analysis included the data collected and prepared by Schroyens and Schaeken (in preparation). They carried out their own meta-analysis, which included a detailed report of the patterns of inference for the three sorts of task. Their results were from adult participants and high school students in their final year. Furthermore, the studies used abstract conditionals, and other logically equivalent formulations, such as: *all A are B*, *B if A*, *A unless not B*, and *B only if A*, and the biconditional: *if, and only if, A then B*. In addition to these data, the meta-analysis included results reported in Oberauer (2006). We searched the literature in April 2018 on Google Scholar and PubMed. But, none of the other papers that we found reported the frequencies of the 16 different patterns for the four sorts of inference.. Yet, these patterns were essential for our meta-analysis. Thus, in the end, our work relies on results of one study by Klaus Oberauer and 14 studies that Walter Schaeken kindly provided to us (e.g., Barrouillet, Grosset, & Lecas, 2000; Byrne & Tasso, 1999; Evans, Clibbens, & Rood, 1995; Schroyens, Schaeken, & Handley, 2003; and more). In sum, the meta-analysis included data from 39 experiments (from 15 studies) that tested a total of 2378 participants.

MP as the most basic inference pattern

MP is the fundamental inference in conditional reasoning. It is commonplace in everyday life, and most experimental participants make it, though a few failures do occur (see, e.g., Oberauer, 2006). In our view, individuals who do not make MP in an experiment have failed to reason, and so we have excluded their data from our analyses. It therefore focused on the eight patterns of response that include MP.

The Dependency of the Inference Patterns

Some theories of conditional reasoning assume that inferences of the four sorts of inference are independent of one another (Evans & Lynch, 1973). Other theories do so de facto in that they consider only the frequencies of each of the four sorts of inference, not the frequencies of their patterns (e.g., Oaksford et al., 2000). But, are the four sorts of inference independent of one another?

The question is an empirical one, and to examine it we used an algorithm based on Shannon's measure of information, which we used to show that the selections of potential evidence to test a conditional hypothesis are dependent on one another (see Ragni, Kola, & Johnson-Laird, 2018). The intuition motivating the algorithm is simple. Suppose that the inferences in an experiment are more redundant – less informative – than inferences based only on the individual probabilities of each of the four inferences in the experiment. It follows that something is constraining the inferences over and above their independent frequencies. Hence, the selections are dependent. Consequently, if the inferences are dependent, theories implying their independence are wrong. We therefore tested whether the patterns of inference in the experimental data were significantly more redundant (using Shannon's measure) than those of 10,000 simulations of each experiment based on independent selections.

To analyze the potential dependence of the conditional inferences we examined the data for each experiment in our sample following four main steps:

1. Compute N, the number of participants, and the probabilities of the eight inference patterns (each including MP) in the set of the participants' inferences.

2. Compute Shannon's entropy H for the experiment.

3. Carry out 10,000 simulated experiments based on the probabilities of making each inference, assigning a pattern to N hypothetical participants.

4. Return the number of simulated experiments that were more informative than the actual experiment and the number with the same or lower information values.

Table 2 shows the relative frequencies of the main patterns of inference in our sample of 39 experiments.

Table 2: The relative percentages of five patterns of inference in 39 experiments categorized according to the task (evaluation, option or production of a conclusion). Three of the eight patterns occurred less often than 5% and are not included in the table.

Response pattern	Evaluation	Option	Production	Overall
All inferences	39.0	43.1	50.3	42.3
MP, MT	17.7	21.3	12.4	18.3
MP	14.8	8.4	5.4	10.9
MP, AC	8.4	5.8	11.0	7.8
MP, AC, MT	4.0	8.2	8.5	6.3
Number of Experiments	8	22	9	39
Number of Participants	1103	921	354	2378

Note. The different response patterns indicate whether the MP, DA, AC, MT inferences were accepted, selected, generated or not. We do not present patterns that occurred less than 5% in each task.

Table 3 presents the information value of the 39 experiments investigating the three sorts of task, the mean value of each of their 10,000 simulations, and the results (of Wilcoxon's test comparing the two values) indicating a reliable dependence over the four sorts of inference.

Table 3: The mean information value (in bits, with a theoretical maximum of 3 bits) of 39 experiments using three tasks, their mean information value, and that of sets of 10,000 simulations of each experiment.

Sort of task	Evaluation	Option	Production
Mean information of experiments	1.69	2.03	1.93
Mean information of their simulations	2.05	2.29	2.12
Wilcoxon's W and p-value	$W = 3$, $p < .04$	$W = 3$, $p < .04$	$W = 6$, $p = .055$

In sum, these results demonstrate that the four sorts of conditional inference in the experiments depend on each other.

An evaluation of theories of conditional reasoning

We evaluated five theories using multinomial processing trees (based on the formulated trees in see Oberauer, 2006; see Figure 1 for the tree of the mental model theory).

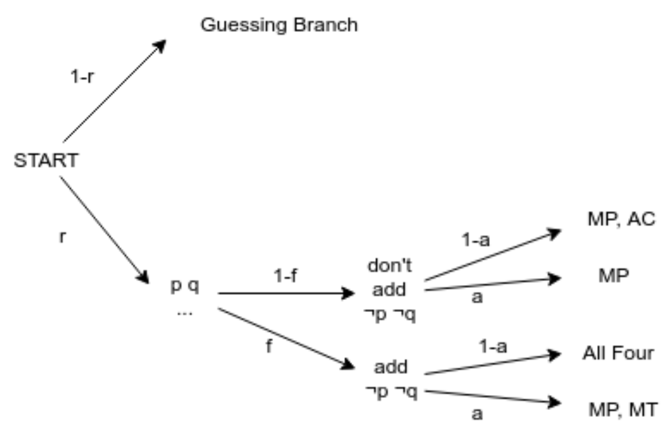

Figure 1: The multinomial processing tree for the mental model theory. The parameter r stands for the reasoning part and $1\text{-}r$ for the guessing part (for an explanation, see Oberauer, 2006). Parameter f controls whether or not models are fleshed out to include a model of ¬ p and ¬ q.

Within each tree for a theory, the probability of a particular cognitive state is estimated from the observed frequencies of inferences (Riefer & Batchelder, 1988). We used the maximum-likelihood method from the R-package for multinomial processing trees (MPTinR, Singmann & Kellen, 2012) to fit each theory's tree to the frequencies of the four patterns of inference, separately for the three different types of tasks. To compare the models, we calculated the Bayesian information criterion (BIC), which indicates how much information is lost when a model represents the process that generates the data. This criterion takes into account both a tree's parsimony and its goodness of fit with the data. Thus, the BIC rewards a good fit and punishes a higher number of free parameters. A lower BIC indicates a better theory, because it has fewer parameters or fits the data better, or both. Table 4 presents the BIC for the different trees we fitted for the three different types of conditional tasks (evaluation, option, and production).

The results in Table 4 show that the model theory is the best in accounting for the conclusions that reasoners draw for themselves (the production task). Its directional version is best for the evaluation of given conclusions (the evaluation task). The best theory for the task of selecting a conclusion from a multiple choice (the option task) is the exclusive version of the suppositional theory, which postulates that either system 1 or else system 2 is engaged in the process of reasoning.

Table 4: The fit of the different trees based on the theories of reasoning for the different sorts of task.

The theory	Evaluation (BIC)	Option (BIC)	Production (BIC)
Suppositional-exclusive	42	**43**	45
Dual process suppositions	48	48	42
Directional mental model	**37**	60	42
Mental model	46	62	**38**
Suppositional	63	54	60

Note. 'Evaluation' refers to the evaluation of a given conclusion; 'Option' refers to the choice of a conclusion as an option in a multiple-choice format; and 'Production' refers to the production of a conclusion from a set of premises. A lower Bayesian information criterion (BIC) indicates that a theory has fewer parameters or fits the data better, or both. Best fits are shown in bold.

General Discussion

Following previous research (e.g., Oberauer, 2006; Schaeken, 2001), we carried out a meta-analysis to determine which theory of conditional reasoning and type of tasks gave the best account of individuals' patterns of inference in the four basic sorts of conditional reasoning inferences (modus ponens: MP, affirmation of the consequent: AC, denial of the antecedent: DA, and modus tollens MT). Following Schaeken and Schroyens (in preparation), we separated the studies into those that called for the evaluation of a given conclusion (evaluation task), its selection from a set of options (selection task), and its production from the premises alone (production task) and considered their differences in the analyses.

Our results showed that the most frequent pattern of inferences in all three tasks was to make all four inferences – a pattern that is valid only if the conditionals are interpreted as biconditionals (see Table 2). The next most frequent pattern was to make MP and MT inferences – the two inferences that are valid given a classical conditional interpretation. These results do not discriminate among the various theories of conditional reasoning, though some theories, such as the model theory, predict that AC can occur in the absence of a biconditional interpretation.

An analysis of the amount of information, using Shannon's measure, showed that the patterns of actual inferences were more redundant than those of 10,000 simulations of each experiment (see Table 3). This result corroborated our elimination of any theory in which each selection is in principle independent of the others, e.g., the probabilistic theory of conditional reasoning (Oaksford et al., 2000).

Finally, we fit the multinomial processing trees for each of the remaining five theories to the results of 39 experiments using the three different tasks (Table 4). The Bayesian information criterion, which credits a fewer number of parameters and goodness of fit, showed that the model theory gave the best account of the production of conclusions. Furthermore, its variant that reflects the direction of an inference (the directional model theory) – from *if*-clause to *then*-clause, or vice versa – gave the best account of the evaluation of a given conclusion. We speculate that this result may reflect the order of clauses, i.e., *A C* versus *C A,* in some of the putative conclusions that the participants had to evaluate. In the suppositional theory, which proposes that participants rely either on system 1 or else on system 2 (its exclusive variant), gave the best account of the selection of an option from a set of multiple conclusions.

But why do participants seem to differ in their inference patterns given different types of tasks? There is strong evidence that the response modality of conditional tasks (e.g., scaled or dichotomous response format) affects the way participants process the presented information (e.g., Markovits, Forgues, & Brunet, 2010). These results are consistent with the idea that scaled responses promote a probabilistic mode of processing. Yet, the current studies concentrated on data from non-rating tasks. This constraint was necessary in order to compare the three sorts of task. Future studies should extend the present findings by considering different types of response modality.

By far the most important task for future studies is to formulate tasks in which participants make conditional interpretations. As we mentioned, the most frequent pattern in the present studies was for a biconditional interpretation and so, for now, it is not possible to determine how well the various theories would fit tasks in which the main interpretation is for a conditional and not a biconditional interpretation.

The difference among the theories' fit to the data over the three types of task shows that theories should account for performance in different tasks. Different tasks yield different patterns of inference. One factor, for instance, could be that the need to formulate a conclusion discourages guesswork in comparison with the "option" task in which participants choose a conclusion from a multiple set of possible responses. Future studies should therefore separate different tasks in their analyses of theories.

Acknowledgements

This paper was supported by DFG grants RA 1934/3-1, RA 1934/2-1 and RA 1934/4-1 to MR.

References

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

Byrne, R. M. J., & Tasso, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory & Cognition*, *27*, 726–740.

Evans, J. S. B., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, *64*(3), 391-397.

Evans, J. S. B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition*, 48, 1–20.

Evans J. St.B.T., Clibbens, J. & Rood, B. (1995). Bias in conditional inference: Implications for mental models and mental logic. *The Quarterly Journal of Experimental Psychology Section a*, *48*, 644–670.

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

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

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

Markovits, H., Forgues, H. L., & Brunet, M. L. (2010). Conditional reasoning, frequency of counterexamples, and the effect of response modality. *Memory & Cognition*, *38*(4), 485-492.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 883.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology*, *53*, 238–283.

Ragni, M., Kola, I., & Johnson-Laird, P. N. (2018). On selecting evidence to test hypotheses: A theory of selection tasks. *Psychological Bulletin*, *144*(8), 779.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. Psychological Review, 95(3), 318-339.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA, US: MIT Press.

Schaeken, W., & Schroyens, W. (2000). The effect of explicit negatives and of different contrast classes on conditional syllogisms. *British Journal of Psychology*, *91*, 533–550.

Schaeken, W., Schroyens, W., & Dieussaert, K. (2001). Conditional assertions, tense, and explicit negatives. *European Journal of Cognitive Psychology*, *13*, 433–450.

Schroyens, W. & Schaeken, W. (in preparation). Individual differences and/in models of conditional reasoning by model.

Schroyens, W., Schaeken, W., & d'Ydewalle, G. (2001). *A meta-analytic review of conditional reasoning by model and/or rule: Mental models theory revised.* (Report No. 278). Leuven: University of Leuven, Laboratory of Experimental Psychology.

Schroyens, W., Schaeken, W., Fias, W., & d'Ydewalle, G. (2000). Heuristic and analytic processes in propositional reasoning with negatives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1713.

Schroyens, W., Schaeken, W., & Handley, S. (2003). In search of counter-examples: Deductive rationality in human reasoning. *The Quarterly Journal of Experimental Psychology Section a*, *56*, 1129–1145.

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, *45*(2), 560-575.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning*, *11*(3), 239-278.

Weidenfeld, A., Oberauer, K., & Hörnig, R. (2005). Causal and noncausal conditionals: An integrated model of interpretation and reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *58*(8), 1479-1513.