

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Issues in Reasoning about Iffy Propositions: The Initial Representation of Conditionals

#### **Permalink**

<https://escholarship.org/uc/item/802121st>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

#### **ISSN**

1069-7977

#### **Author**

Shirouzu, Hajime

#### **Publication Date**

2005

Peer reviewed



finding that affirmation inferences, which the initial model yields, are more readily endorsed than denial inferences, which the initial model does not yield. This pattern has been corroborated in two meta-analytic studies (Schroyens et al., 2001a, 2001b). Initial models do not represent the categorical premise of the denial inferences. Hence, the premise eliminates the explicit model to leave only the implicit model from which nothing follows: one draws a blank. Individuals have to consider models of alternative possibilities before they can make a denial inference. The initial model assumption similarly explains the on-line inferences people tend to draw (and tend not to draw) during the comprehension of text (Lea, 1995). More direct evidence for the initial model assumption is provided by individual responses patterns in which individuals endorse the affirmation inference but reject the denial inferences. This pattern forms the explanatory basis for the main effect of the sort of inference (affirmation versus denial) that is corroborated by the abovementioned meta-analyses. More recent meta-analyses on the frequency of individual inference patterns, however, show that on average only 7% of the adults exhibit this so-called ‘conjunctive inference’ pattern where they endorse the affirmation inferences but reject the denial inferences. This is barely above chance level ( $1/4^2 = 6.25\%$ ; Schroyens & Schaeken, 2003b).

In the task of evaluating conditional inferences participants are given a conclusion to evaluate. It is common practice to call for participants to evaluate the determinate conclusion shown in Table 1. These conclusions are logically valid for MP and MT, but they are logically valid for AC and DA only in a bi-conditional interpretation of the major premise, i.e., it is taken to mean ‘if and only if A then C’. Table 1 also shows the inferences that individuals can make from their initial model of the premises. As readers will see, there is a conflict between the standard conclusions and conclusions derived from the initial models in the case of the two arguments based on denial. Individuals should be more likely to endorse the standard conclusion in these two cases than to reject ‘nothing follows’ if this response is presented for evaluation. In the case of the affirmation inferences, however, the initial models predict that individuals should both accept the standard conclusions and reject the ‘nothing follows’. It follows that there should be an interaction between the sort of inferences (affirmation or denial) and the sort of conclusion to be evaluated (standard versus ‘nothing follows’). The difference between the acceptance of standard conclusions and the rejection of nothing follows should be greater for denial inferences than for affirmation inferences. In order to test this prediction we carried out an experiment in which we manipulated the sort of conclusions that participants had to evaluate.

In the general discussion we will consider alternatives to the mental models theory, and will demonstrate how one such salient alternative model of human reasoning fails to make, or explain the central prediction of the present study. That is, to pre-empt the results of our study, the findings not only support the explanatory import of representational assumptions made in the mental models theory, but also allow us to counter alternative models of conditional reasoning.

## Experiment

### Method

**Design.** All participants evaluated 2 (Validity: Valid vs. Invalid) by 2 (Type: affirmation vs. denial) by 3 (conclusion: Standard/Opposite/Indeterminate) arguments. These 12 arguments were presented twice, in a blocked order.

**Materials and Procedure.** Each participant participated by running a custom-made computer program. Participants first read the instructions, then received two exercise problems (about ‘or else’), and subsequently received two blocks of 12 arguments. Within each block they were randomly given an MP/DA/AC/MT problem with its three conclusion types. With each of the four problem types participants were given the affirmation of the inferential clause, the denial of the inferential clause, or the nothing follows conclusions; thus producing the 12 (3x4) arguments. The affirmation of the inferential clause is the standard for the affirmation problems (MP/AC); it is the opposite of the standard denial inferences (MT/DA). The denial of the inferential clause is the standard denial inference; it is the opposite of the standard affirmation inference. A standard DA inference was presented as:

Given: If the letter is a B, then the number is a 2.

Given: The letter is NOT a B.

Walter Concludes: Hence, the number is NOT a 2.

The conclusion of the opposite DA read “Hence, the number is a 2”. Participants had to evaluate whether the derived conclusion was “logically correct or incorrect”, by clicking the mouse on a ‘correct/incorrect’ button [‘juist’/‘fout’ in Dutch].

**Participants.** Forty-four University of Leuven (Belgium) students participated (17-23 years of age,  $M = 18$ ). They received credit points or were paid for their participation.

### Results

Figure 1 presents the acceptance rates of the standard arguments and rejection rates of the null inferences. Unity minus the acceptance rates of the nothing follows inference corresponds, in principle, to the acceptance rates of the standard inferences. The results clearly show that though this equivalence is true in principle, in practice there is a clear discrepancy between the two arguments. The overall standard acceptance rates amount to .747, whereas the overall nothing follows rejection rates amount to .614 (Wilcoxon  $T = 48.0$ ,  $N = 30$ ,  $Z = 3.794$ ,  $p < .001$ ). Most importantly, the difference depends on the type of problem (Affirmation vs. Denial:  $d = -.028$  vs.  $d = .295$ ;  $T = 28.0$ ,  $N = 32$ ,  $Z = 4.413$ ,  $p < .001$ ). People are as likely to accept the standard affirmation arguments (MP/AC) as they are likely to reject the conclusion that nothing follows from the same premises (.784 vs. .812).<sup>1</sup> Phrased otherwise, the likelihood

<sup>1</sup> An anonymous reviewer suggested that under some conditions one might expect an increase of ‘correct’ responding to AC, that is, an increase of nothing follows acceptance rates. We agree. The

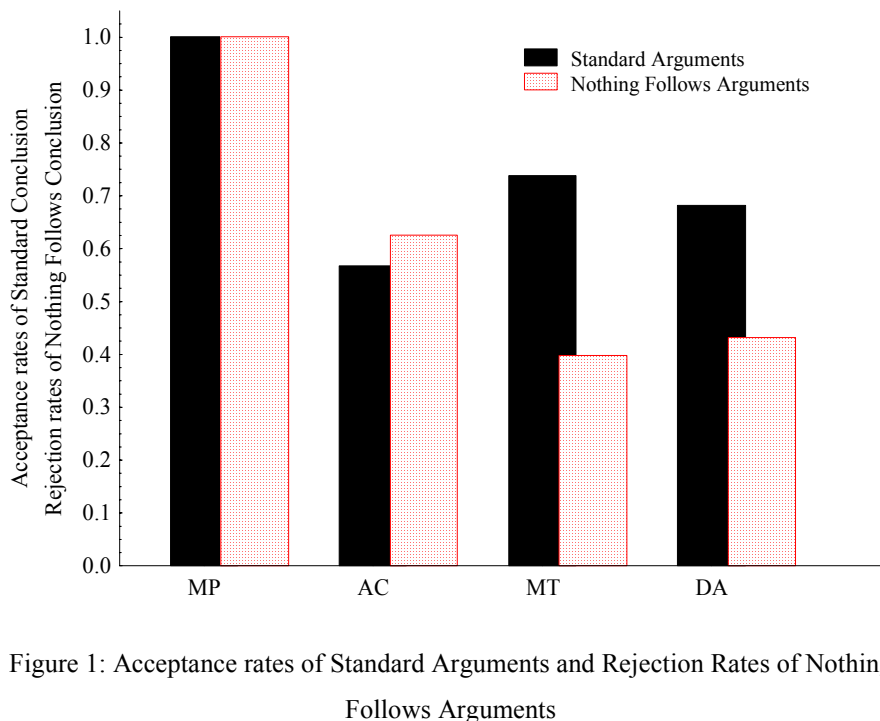


Figure 1: Acceptance rates of Standard Arguments and Rejection Rates of Nothing Follows Arguments

of accepting the standard affirmation inferences and accepting the nothing follows inference sums up nicely to unity. In case of denial (MT/DA), people are much more likely to accept the standard denial argument than they are likely to reject the nothing follows conclusion for the denial problems (.710 vs. .415;  $T = 11.0$ ,  $N = 27$ ,  $Z = 4.276$ ,  $p < .001$ ).

In summary, people are more likely to accept the initial-model conclusion (i.e., ‘nothing follows’) when this conclusion is given for evaluation. The fact that the conclusion type (standard vs. nothing follows) interacts with the type of problem shows that people are not simply more likely to accept the conclusion that is given for evaluation. Our argument hinges on the observed difference between accepting the standard inferences and rejecting the nothing follows inferences. One might argue that an equivalence between the standard acceptance rates and the nothing follows rejection rates would not be expected in the first place, because people might endorse the opposite of the standard conclusion. Those people who do not endorse the standard inferences, would accept ‘nothing follows’ or would accept the opposite conclusion. The results on the opposite-conclusion problems show that virtually nobody

conditions of the present experiment, however, are not conducive to observing such an effect. “Nothing follows” would need to be elaborated and explicated as the logically valid argument “[if A then C]; [C]; hence [A or not-A]”. That is, “nothing follows” would need to be interpreted as “nothing follows necessarily” (as indeed, [A] does follow possibly from AC). This requires that one explains to participants that a logical conclusion is a conclusion that follows necessarily, and not just possibly. (See, Schroyens et al., 2003, Experiment 2, for a demonstration and discussion of stressing the logical necessity instructions).

accepts this type of conclusions (acceptance rates on MP, AC, MT, DA are respectively: 0.00, 0.00, 0.02 and 0.01). That is, as assumed in mental models theory, responses fall into two classes. People infer the default standard conclusion, or infer that nothing conclusive can be inferred from the premises (see Schroyens, Handley, Evans, & Schaeken, 2002, for further discussion and experiments about opposite conclusion effects). Virtually nobody endorses the opposite conclusions. These are supported neither by the initial models, nor by a fully explicated representation of all possibilities (i.e., models) that would be consistent with the premises. This finding strengthens our argument and analysis of the difference between the acceptance rates of the standard conclusions and the rejection rates of the nothing follows inferences: Our finding fail to disrepute, that is, they support the initial-model assumption in mental models theory.

### General Discussion

Our findings provide clear support for the initial-model assumption in mental-models theory, and its import in explaining one of the most robust findings in the literature on conditional reasoning. People are generally much less certain about the standard denial (vs. affirmation) arguments. This is presumably due, in part, to the temporal layout of the human processing system which within the scope of its bounded rationality ensures that people initially represent as little information as possible explicitly and ab initio only represent what is explicitly narrated with the sentential information they are confronted with. The sentence ‘if it rains, then the streets get wet’, is about it raining and the streets getting wet. People would accordingly first consider this contingency.<sup>2</sup> The initial-model assumption implies that alternative possibilities need to be considered, if people are to infer anything specific from an additional bit of information that alludes to a possibility that is not represented in the initial representation (i.e., as is the case in a denial problem). We have shown that

<sup>2</sup> Note that it is theoretically false and hence misleading to state that mental-models theory considers the contingency between ‘it raining’ and ‘streets getting wet’ a simple conjunctive juxtaposition of the two events concerned (see, Evans et al., 2003; Grosset & Barrouillet, 2003). The abstract of Johnson-Laird and Byrne (2002) already proofs such claims wrong: “The 2 sorts of conditional have separate core meanings that refer to sets of possibilities. Knowledge, pragmatics, and semantics can modulate these meanings. *Modulation can add information about temporal and other relations between antecedent and consequent*” (p.646, italics added).

when we give people the initial-model conclusion for evaluation, the classic difference between the affirmation problems and the denial problems becomes even larger. This is clear evidence in favour of the initial-model assumption, and its import in explaining thinking and reasoning about conditional relation, as expressed in for instance conditional ‘if ...then ...’ utterances.

The initial-model assumption constitutes part of the model-based theory of interpretation, which is to be distinguished from a theory of meaning (Gilbert, 1991). Given that the meaning of a sentence can be captured by means of the possibilities it allows for, the core meaning of [if A then C] and its contra-positive [if not-C then not-A] is the same.<sup>3</sup> Both conditionals are falsified only by the joint contingency of [A] and [not-C]. There is abundant evidence that shows that a conditional is judged false when there are situations wherein the antecedent is satisfied, while the consequent is not (see, e.g., Evans, Ellis, & Newstead, 1996). However, this does not mean that [if A then C] and [if not-C then not-A] are interpreted in the same way, or that the logical equivalence corresponds to a psychological equivalence. A straightforward application of the initial-model principle shows that the initial interpretations/representations are fundamentally different. They represent the contingency between respectively [A] and [C], and [not-C] and [not-A]: quite something different.

The difference between meaning and interpretation is far from trivial. Indeed, it seems to be tempting to sacrifice faithfulness to the complexities of the phenomena for the sake of clarity of an (often polemic) exposition. Evans et al. (2003), for instance, argued against mental models theory in the process of proffering their own ‘theory’. As exposed by Schroyens and Schaeken (2004), they formed a misleading argument on the basis of the logical equivalence between the conditional and its contra-positive. This attests to a failure to distinguish logic from psychology and reflects a gross neglect of the intricate details of mental-models theory (i.e., the crucial initial-model assumption). Advances in the field are in part formed by critical analyses of extant models and theories. However, when a critique is based on an over-simplified misrepresentation of the theoretical advances in the field, we are setting a step backwards, not forwards.

The present set of findings also provides further evidence against the so-called conditional probability (CP) model of conditional reasoning (Oaksford et al, 2000). The CP model has already shown to be deficient for abstract reasoning about the type of standard inference evaluation problems

used in the present study (see, Schroyens & Schaeken, 2003a; also see Oaksford & Chater, 2003). Our findings obtained by providing a “nothing follows” inference for evaluation, demarcate further problematical findings for CP model, which as such is discredited and due for a serious maintenance job if it is to be reckoned with as an alternative theory of human reasoning. The CP model is very simplistic. The probability of the standard MP, AC, DA, MT arguments is supposed to equal the conditional probability of the conclusion, given the categorical premise:  $P(MP) = P(q|p)$ ;  $P(AC) = P(p|q)$ ;  $P(MT) = P(\text{not-}p|\text{not-}q)$ ;  $P(DA) = P(\text{not-}q|\text{not-}p)$ . Using basic probability calculus, i.e., Bayes Theorem, these probabilities can be reformulated into a set four equations with three parameters:  $P(p)$ ,  $P(q)$  and  $P(\text{not-}q|p)$ . We do not need to use these more complex formulae. Our critical demonstration is independent of such mathematical reformulations of the model. Oaksford et al. (2000) assume that the probability of the opposite conclusion endorsements is analogously equal to the conditional probability of the opposite conclusion, given the categorical premise:  $P(MP') = P(\text{not-}q|p)$ ;  $P(AC') = P(\text{not-}p|q)$ ;  $P(MT') = P(p|\text{not-}q)$ ;  $P(DA') = P(q|\text{not-}p)$ . That is, the opposite conclusion endorsements are the simple complements of the standard conclusion acceptance rates:  $P(MP) = 1 - P(MP')$ ;  $P(AC) = 1 - P(AC')$ ;  $P(MT) = 1 - P(MT')$ ;  $P(DA) = 1 - P(DA')$ .

Our results show that the opposite conclusions were rejected throughout, whereas the standard conclusions were not. That is, the probability of, for instance, AC (“q, therefore p”: .57) is not complementary to the endorsement rate of its opposite (AC’: “q, therefore not-p”: .02). This means that the opposite-conclusion effects counter the CP model, at least as it is presented by Oaksford et al. (2000). This is too simple though and would no advance our theoretical knowledge very much. A basic principle in the advancement of science and theories (as well as a basic ethical principle in inter-personal conduct), is the principle of charity: one gives the benefit of the doubt, and makes the best possible interpretation of a theory (or person; Davidson, 2001). As argued by Schroyens et al. (2002), the CP model needs to be extended by a decision-mechanism that maps a probabilistic evaluation of a conclusion onto a binary decision to accept or reject a conclusion that is only probabilistically truthful.

The CP model is not salvaged by extending it with a function that maps subjective probabilities onto decisions to accept/reject conclusions. Such a function expresses the idea that a decision to accept a conclusion is done above a particular threshold. The results on the “nothing follows” problems show that even the extended CP model is in trouble. First, consider though that one could make another simple argument against the CP model by pointing out that Oaksford et al. do not consider “nothing follows” inferences, and only talk about standard and opposite conclusions. This issue is remedied by taking up the in the literature commonly considered idea that the “nothing follows” conclusion reflects uncertainty about a determinate conclusion: both the standard and the opposite conclusions have a particular likelihood of being true. That is, “nothing follows”, vis-à-vis, e.g. AC, captures the maximally

<sup>3</sup> The core-meaning assumption in mental-models theory is part of its theory of meaning, whereas the initial-model assumption is part of the theory of interpretation. The theory of meaning reflects an idealisation wherein abstraction is made of the pragmatics of the context we live in, and the content we live with. Johnson-Laird and Byrne (2002) speak of the core-meaning of basic conditionals, which have “a neutral content that is as independent as possible from context and background knowledge, and which have an antecedent and consequent that are semantically independent apart from their occurrence in the same conditional” (p. 649). Basic conditionals are, thus, not ordinary conditionals, which are dependent from content and context. As such one cannot speak of “the core meaning of ordinary conditionals” (Evans et al., 2002).

uninformative (but logically valid) conclusion that “p or not-p” might be the case. Under these (charitable) assumptions, the extended CP model still remains deficient.

Let us assume a conclusion acceptance threshold at certainty level X. This implies that the distribution of subjective probabilities of the standard conclusions lies completely to the right of 1-X. Since the subjective probabilities of the standard and opposite conclusions are complementary this means that there is no point under the distribution of the subjective probabilities of the opposite conclusions that lies beyond the acceptance threshold. We

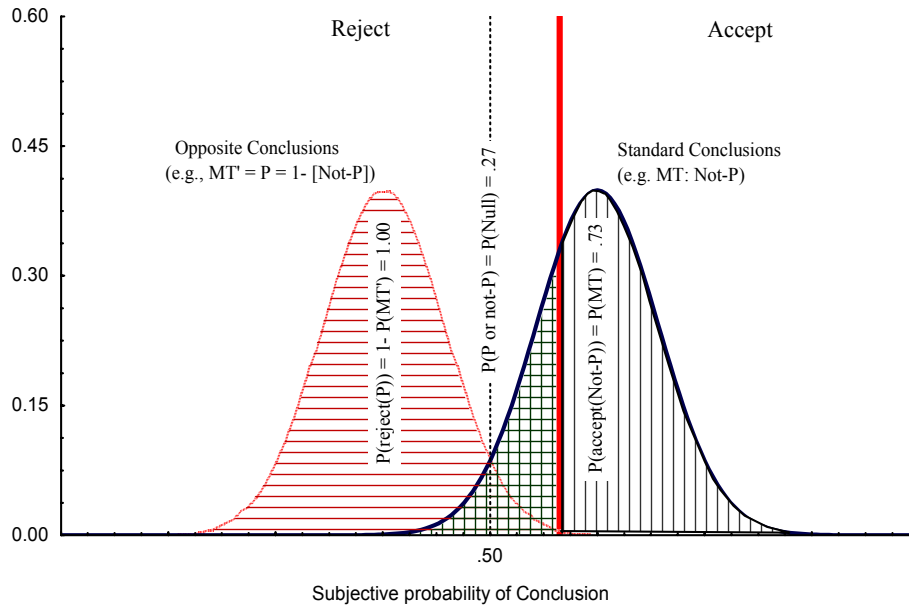


Figure 2: Example of two distribution functions corresponding to the probability of standard conclusions and their opposites, and a decision threshold that leads to accepting/rejecting these conclusions.

have to assume such a distribution (i.e. range limit) for all four subjective probability functions because, the results showed a uniform rejection of all opposite conclusions. Otherwise some opposite conclusions would be endorsed. Figure 2 illustrates two such complementary probability distributions, and provides an acceptance threshold such that the likelihood of accepting the standard conclusion amounts to .73 (i.e. the surface below the curve, to the right of the threshold is 73% of the total surface). This is the standard MT endorsement rate observed in the present study. As Figure 2 shows, the opposite MT endorsement rates would be zero. This is exactly what we observed. The probability of the standard conclusion (not-p) is not high enough to endorse the standard conclusion and neither is it high enough to endorse the opposite conclusion (p), whose probability ex hypothesis equals unity minus the probability of the standard conclusion.

The extended CP model, however, expects to observe a .27 (1-.73) acceptance rates of “nothing follows” (p or not-

p) on MT.<sup>4</sup> The acceptance rate of “nothing follows” in the extended CP model equals the acceptance rates of opposite conclusions in the simple CP model.) This is clearly countered by the data: the nothing follows endorsement rates actually amount to .71. No statistics are needed to see that this is quite a stretch away from a .27 endorsement rate. A similar problem for the CP model is observed on DA (1-.68 = .32. vs. .57 nothing follows endorsement rates). Only the results on the affirmation problems are in line with the extended CP model (MP: 1- 1 = 0 vs. 0; AC: 1-.57 = .43 vs. .38). This need not surprise us. Indeed, the discrepancy

between standard acceptance rates and the nothing follows rejection rates – or formally equivalent: the discrepancy between the standard rejection rates and the nothing follows acceptance rates, as formulated above – is exactly what we predicted on the basis of the initial-model assumption in mental models theory. The CP model does not subscribe to this principle, or any other representational principle that has the same import. As such it fails to account for our findings.

In summary, we extended the conclusion evaluation paradigm by changing the type of conclusion provided. The standard conclusions are the ones by which a determinate conclusion is inferred from the premises. A theoretical analysis of this task shows a confound as regards the status of the standard conclusions vis-à-vis the processes of reasoning by

constructing and manipulating mental models. The standard conclusions for affirmation problems can be drawn from the supposed initial representation of the premises, whereas the standard denial inferences cannot. This explains the robust difference between the acceptance rates of standard affirmation and denial inferences (see, Schroyens et al., 2001; Schroyens & Schaeken, 2003a, for a meta-analysis). By providing the initial model conclusion for the denial problems (“nothing follows”) we provided new, additional evidence for the initial model assumption. The difference between the affirmation and denial problems (.78 vs. .71) becomes even bigger when one considers the initial model conclusions (respectively the determinate affirmation inference for the affirmation problems and the indeterminate

<sup>4</sup> Changing the shape of the density function does not change expectations. One could increase the expected rate of null responses [p or not-p] by increasing the overlap between the distributions of [p] and [not-p]. This however would be in direct conflict with what Oaksford et al. (2000) proffered: The probabilities of the standard and opposite conclusions (though not necessarily their categorical acceptance rates) are complementary.

nothing follows inference for the denial problems: .78 vs. .59). We have illustrated how an alternative theory that does not support the initial-representation assumption, remains unsupported (i.e., is countered) by the present findings. First, the finding that opposite conclusion are not endorsed shows that the Conditional Probability model needs to be extended to capture the idea that conclusion are only accepted when their likelihood is beyond a decision threshold. Second, the CP must also be extended so that it allows to account for nothing follows conclusions, which it can be adopting the thesis that “nothing follows (with certainty)” reflects the idea that both the standard conclusion and its opposite are possible. The nothing follows rejection rates show that the extended CP remains problematical: it continues to underestimate nothing follows inferences on the denial problems. The CP model needs to follow mental models theory and subscribe to the initial-representation principle to account for the findings that we predicted on the basis of the general mechanisms of reasoning by constructing and manipulating mental models.

### Acknowledgments

The present research was funded in part with the support of the Flanders (Belgium) Fund for Scientific Research (G.0320.05) and the Canadian Natural Sciences and Engineering Research Council (NSERC 297517). I also like to extent my grateful acknowledgments to Walter Schaeken, Kristien Dieussaert, Niki Verscheuren, Wim De Neys and all other members of the R.A.T. research group at the University of Leuven, Belgium.

### References

- Davidson, D. (2001). *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- 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, 1086-1114.
- Evans, J. St. B. T., Handley, S. J., & Over, D. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29(2), 321-335.
- Gilbert, D. (1991). How mental systems believe. *American Psychologist*, 46(2), 107-119.
- Grosset, N., & Barrouillet, P. (2003). On the nature of mental models of conditional: The case of If, If then, and Only if. *Thinking & Reasoning*, 9(4), 289-306.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109(4), 646-678.
- Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, 99, 418-439.
- Lea, B. E. (1995). On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1469-1482.
- Oaksford, M., & Chater, N. (2003). Computational levels and conditional inference: Reply to Schroyens and Schaeken (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 150-156.
- Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(4), 883-899.
- Schroyens, W., Handley, S., Evans, J. St. B. T., & Schaeken, W. (2002). Evaluating conditional arguments: Opposites do not attract. Technical Report (No. 292). Leuven: University of Leuven, Laboratory of Experimental Psychology, <http://www.psy.kuleuven.ac.be/reason/report292.pdf>.
- Schroyens, W., & Schaeken, W. (2003a). A critique of Oaksford, Chater and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 140-149.
- Schroyens, W., & Schaeken, W. (2003b). *Essays in reasoning by model: Individual differences in/and models of conditional reasoning*. Internal Technical Report (Report No. 302). Leuven: Laboratory of Experimental psychology, University of Leuven, <http://www.psy.kuleuven.ac.be/reason/report302.pdf>.
- Schroyens, W., & Schaeken, W. (2004). Guilt by association: On iffy propositions and the proper treatment of mental-models theory. *Current Psychology Letters: Behaviour, Brain & Cognition*, 12(1), <http://cpl.revues.org/document404.html>.
- Schroyens, W., Schaeken, W., & d'Ydewalle, G. (2001a). A meta-analytic review of conditional reasoning by model and/or rule: Mental models theory revised. Internal Technical Report (No. 278). Leuven: University of Leuven, Laboratory of Experimental Psychology, <http://www.psy.kuleuven.ac.be/~walters/report278.pdf>.
- Schroyens, W., Schaeken, W., & d'Ydewalle, G. (2001b). The processing of negations in conditional reasoning: A meta-analytic study in mental model and/or mental logic theory. *Thinking and Reasoning*, 7(2), 121-172.