Inferential Conditionals and Evidentiality*

Karolina Krzyżanowska
Faculty of Philosophy, University of Groningen
k.h.krzyzanowska@rug.nl

Sylvia Wenmackers
Faculty of Philosophy, University of Groningen
s.wenmackers@rug.nl

Igor Douven
Faculty of Philosophy, University of Groningen
i.e.j.douven@rug.nl

Abstract

Many conditionals seem to convey the existence of a link between their antecedent and consequent. We draw on a recently proposed typology of conditionals to argue for an old philosophical idea according to which the link is inferential in nature. We show that the proposal has explanatory force by presenting empirical results on the evidential meaning of certain English and Dutch modal expressions.

1. Introduction. Among the numerous theories of conditionals that have been proposed so far, none seems able to account for all the empirical data concerning how people use and interpret such sentences.[1] At least prima facie, a theory of conditionals appears materially inadequate if it validates, in whichever precise sense, sentences like these:

(1) a. If badgers are cute, then 2012 was a leap year.

*The final version appeared in *Journal of Logic, Language and Information*, Volume 22, Issue 3, 2013, pp. 315-334.

[1]For a survey of the main accounts of conditionals as well as the problems they face see, for instance, Edgington (1995), Bennett (2003), or Douven (2011).

b. If weasels are people's best friends, then dogs have four legs.

It is easy to understand why we are reluctant to accept conditionals like (1a) and (1b); the antecedents of those conditionals have nothing to do with their consequents. And it seems that using a conditional construction is meant to convey, possibly among other things, the existence of some sort of link between the content of the if-clause and the content of the main clause.

What kind of link might this be? According to a suggestion that has been repeatedly made in the history of philosophy, the link is inferential in nature. That this idea has been dismissed as often as it has been floated may be due to the fact that it was always understood, implicitly or explicitly, that the inferential connection had to be of the same type—namely, deductive—for all conditionals. Recently, a typology of conditionals has been proposed that takes seriously the aforementioned suggestion and argues that it is correct for at least a large class of conditionals, aptly termed "inferential conditionals" in the linguistics literature, while also pointing out that the type of inferential connection may be different for different types of conditionals.

The primary aim of this paper is to provide empirical support for the psychological reality of this typology. We present two empirical studies which show that the typology helps to explain certain systematic differences in people's ratings of the assertability of conditionals, depending on subtle differences in the phrasing of those conditionals. The differences in phrasing all involve expressions that can plausibly be regarded to serve as evidential markers, among other possible roles that they can play, and more specifically be regarded as signalling the presence of particular types of inference. A subsidiary aim of this paper is to provide some evidence that the said expressions can indeed function as evidential markers.

As a preliminary disclaimer, we note that we remain non-committal on the question of whether in using a conditional we *assert* the existence of an inferential link between the conditional's antecedent and its consequent, which would suggest that the existence of such a link is part of the truth conditions of the conditional, or whether we only *implicate* the existence of the link, which might be compatible with conditionals having, for instance, the truth conditions of their material counterparts or even with their having no truth conditions at all.

2. Inferential conditionals. The most general distinction to be made when it comes to classifying conditionals is the distinction between indicative and subjunctive conditionals, of which the paradigmatic examples are, respectively:

(2) a. If Oswald did not kill Kennedy, someone else did.

b. If Oswald had not killed Kennedy, someone else would have.

In this paper, we will only be concerned with indicative conditionals and refer to these simply as "conditionals."[2] For many theorists, this is only the beginning of a typology, though there is little unanimity as to what the typology should further look like. In linguistics, even if not so much in the philosophical or psychological literature on conditionals, it has become common practice to classify conditionals as *inferential* and *content* conditionals.[3] The class of content conditionals is not particularly well defined. Its members are sometimes loosely said to describe relations between states of affairs or events as they happen in reality.[4] However, those will not concern us here. We limit our attention to inferential conditionals. These are conditionals that can be regarded as expressing a reasoning process, having the conditional's antecedent as a premise and its consequent as the conclusion, such as:

(3) a. If she has not had much sleep recently, she will perform poorly on her exams.

b. If he owns a Bentley, he must be rich.

[2]As many authors have pointed out, the distinction between indicative and subjunctive conditionals is not as clear-cut as one might wish. However, the conditionals that were used in the materials of our experiments to be reported in this paper were all uncontroversial cases of indicative conditionals.

[3]See, among others, Dancygier (1998, 2003), Dancygier and Sweetser (2005), Declerck and Reed (2001), Haegeman (2003), and Verbrugge (2007).

[4]This description is too broad to allow for a demarcation of content conditionals from other types of conditional sentences. Even though sentences such as "If she never answers his e-mails, he will get very disappointed with her" or "If you take ice out of the deep freeze, it melts" have been described as typical examples of content conditionals (Verbrugge 2007, p. 4), it would seem that those may well be characterized in terms of inferential relations between their antecedents and consequents, and hence labelled as "inferential."

Inferential conditionals constitute a common type among the conditionals we encounter in natural language.

The idea that a conditional can be considered as somehow embodying a kind of "condensed argument" (Woods 2003, p. 15) is not altogether new to philosophy; it can be traced back at least to Chrysippus, a stoic logician from the third century BC. Chrysippus is believed to have held the view that a conditional is true if it corresponds to a valid argument (see Sanford 1989). Obviously, if one insists on understanding validity in terms of classical deductive inference, it is easy to find counterexamples to the aforementioned idea. Yet deduction is not the only type of inference people rely on in their reasoning, and theories of inferential conditionals should not neglect this fact.

Although linguists have proposed various finer-grained typologies of inferential conditionals (see, e.g., Declerck and Reed 2001), most of these stem from grammatical distinctions. We are interested in a differently based typology recently presented by Douven and Verbrugge (2010), who acknowledge the variety of inferential relations that may exist between a conditional's antecedent and its consequent.

The first distinction these authors make is between *certain* and *uncertain* inferences, where *certain* inferences guarantee the truth of the conclusion given the truth of the premises while *uncertain* inferences only tend to make the truth of the conclusion likely given the truth of the premises.

In Douven and Verbrugge's typology, the certain inferences coincide with the deductively valid ones. The uncertain inferences are, following standard philosophical practice, further divided into *abductive* and *inductive* ones, where the former are inferences based on explanatory considerations and the latter are inferences based on statistical information. More exactly, in an abductive inference we infer a conclusion from a set of premises because the conclusion provides the best explanation for those premises. For example, we may infer that Sally failed her exam from the premises that Sally had an exam this morning and that she was just seen crying and apparently deeply unhappy. That she failed the exam is the best explanation for her apparent unhappiness. Inductive inferences rely on information about frequencies (which may be more or less precisely specified). For instance, we infer that Antonio likes pasta from the premise that he is Italian because we know that by far the most Italians like pasta. It is largely uncontested that people

engage in abductive and inductive inferences on a routine basis. However, it is still a matter of some controversy how to best characterize the notions of abductive and inductive validity.[5] Douven and Verbrugge do not commit to any specific proposals in this regard, and we will not do so here either.

Douven and Verbrugge's typology of inferential conditionals mirrors the aforesaid typology of inference. That is to say, they distinguish between certain (or deductive) and uncertain inferential conditionals, and then divide the latter class further into abductive and inductive inferential conditionals.[6] More specifically, they propose the following:

Definition 2.1 *A sentence "If p, then q" is*

- *a deductive inferential (DI, for short) conditional if and only if q is a deductive consequence of p;*
- *an inductive inferential (II) conditional if and only if q is an inductive consequence of p;*
- *an abductive inferential (AI) conditional if and only if q is an abductive consequence of p.*

Douven and Verbrugge point out that often the inference may rely on the antecedent p together with background assumptions that are salient in the context in which the conditional is asserted or evaluated. They call such conditionals *contextual* DI, AI, or II conditionals, depending on the type of inference involved.[7]

Definition 2.2 *Where* $\mathbf{K} = \{p_1, \ldots, p_n\}$ *is the set of salient background premises, "If p, then q" is*

- *a contextual DI conditional if and only if q is a deductive consequence of* $\{p\} \cup \mathbf{K}$*;*

[5] See for some proposals, Cialdea Mayer and Pirri (1993, 1995), Kyburg and Teng (2001), and Gabbay and Woods (2005).

[6] Note that this typology is not necessarily exhaustive. Following Douven and Verbrugge, we remain non-committal as to whether conditionals expressing, for instance, causal or analogical inferences should be analyzed as separate types or as subclasses of, say, inductive inferential conditionals.

[7] As Douven and Verbrugge (2010, p. 304) note, in contextual AI conditionals, the consequent need not always be the best explanation of the antecedent. It may also be that the consequent is, *in light of the antecedent*, the best explanation of one of the background assumptions.

- *a contextual II conditional if and only if q is an inductive consequence of* $\{p\} \cup \mathbf{K}$*;*
- *a contextual AI conditional if and only if q is an abductive consequence of* $\{p\} \cup \mathbf{K}$*.*

Douven and Verbrugge do not claim that their typology of inferential conditionals is *correct* and the ones that so far have been propounded by other theorists are *incorrect*. What they do claim is that their typology is exceedingly simple and that it is non-ad hoc in that it relies on a time-tested distinction between types of inference. More importantly still, they show in their 2010 paper that the typology has considerable explanatory force by recruiting it in service of testing a thesis, first proposed by Adams (1965) and championed by many since, according to which the acceptability of a conditional is measured by the probability of its consequent conditional on its antecedent. (This thesis is now commonly referred to as "Adams' Thesis.")

In their experiment, Douven and Verbrugge divided the participants into two groups, asking one group to judge the acceptability of ten DI, ten AI, and ten II conditionals and the other group to judge the corresponding conditional probabilities. For all sentences taken together, they were able to disprove Adams' Thesis both in its strict form and in some of its looser forms. That is to say, where Ac stands for the degree of acceptability of a sentence and Pr for the probability operator, they demonstrated that neither is it generally true that $\mathrm{Ac}(\text{If } p, q) = \Pr(q \,|\, p)$, nor that $\mathrm{Ac}(\text{If } p, q) \approx \Pr(q \,|\, p)$, nor that $\mathrm{Ac}(\text{If } p, q)$ is high (middling / low) whenever $\Pr(q \,|\, p)$ is high (middling / low). Splitting out the results for the three types of conditionals showed that Adams' Thesis in its strict form holds only for DI conditionals and that for AI conditionals the most that can be said is that acceptability and conditional probability are highly correlated, for II conditionals not even that much was found to be true.

The typology of inferential conditionals proposed by Douven and Verbrugge explains the systematic differences in the acceptability judgements of different types of conditionals.[8] We take this to be evidence for the sig-

[8]The deeper explanation of these results might be in terms of acceptability conditions, which might be different for the different types of conditionals, or in terms of truth conditions, which might also be different for the different types. See Krzyżanowska (2012)

nificance and cognitive plausibility of the typology.[9] In the following, we aim to provide further support for this typology by relating it to the use of certain linguistic expressions that in the literature have been identified as evidential markers or that can reasonably be assumed to act as such markers. Specifically, we want to show that the typology is able to explain certain patterns in how people evaluate conditionals depending on whether or not these conditionals contain a particular evidential marker. Inserting this or that evidential marker should make the conditional sound more or less natural, depending on the type of inference the conditional expresses. In Sections 4 and 5, we report experimental results that relate the typology of conditionals at issue to various English and Dutch evidential markers. First, we briefly present the broader context of evidentiality and motivate the choice of markers we use in our experiments.

3. Evidential markers. Not all that we believe or assert rests on an equally solid footing. Some things we believe because we saw them with our own eyes. Other things we believe because we heard them from others, or we read them in the newspaper or on the Internet. And again other things we believe on the basis of inferences we made. The source of a belief typically will, and arguably also should, have an effect on the firmness with which we hold the belief. Things we believe because we saw them happening may be particularly firmly held. Second-hand beliefs may also be firmly held, but doubts about the reliability of the source from which we obtained the information giving rise to the belief may have a tempering effect on the firmness with which we hold the belief. Ditto for beliefs based on inference if the inference was non-deductive.

It can be useful for a hearer to know what the source is of the belief a speaker expresses. Even if we take a speaker to be completely sincere,

and Krzyżanowska et al. (2013) for an exploration of the idea that the different types of inferential conditionals have different truth conditions.

[9]Note that by taking its explanatory force as evidence for the typology we are relying on abduction. While neither for the purposes of Douven and Verbrugge's (2010) paper nor for the current use we are making of their proposal is it necessary to make any assumptions about the confirmation-theoretic status of abduction, for independent reasons we do believe that abduction is in much better normative standing than is generally believed. See Douven (2013).

we tend to accept with greater confidence the proposition asserted by the speaker if the source of the speaker's belief in the proposition was their eyesight than if they inferred it from things they read on the Internet (e.g., because we have more confidence in the reliability of the speaker's eyes than in their inferential capacities or in the quality of the information available on the Internet). It is thus no surprise that we often communicate information about the evidential grounds for the contents of our assertions.

In fact, there exist languages, equipped with so called "grammatical evidentiality," in which doing so is mandatory (Aikhenvald 2004, p. 6). In these languages, evidentials are typically expressed by means of morphosyntactic items such as affixes, particles, clitics or special forms of verbs. By contrast, speakers of languages that do not encode evidentiality grammatically, having only lexical means at their disposal, may omit the evidential signal entirely.[10] The evidentiality systems of different languages vary with respect to the number and types of information sources they discriminate. The distinction which is most commonly marked is that between direct (perceptual) access to the evidence and indirect access, where the latter can often be further divided into inferential and reportative access. Some evidentiality systems are more elaborate still and allow to distinguish between different modalities of perception, or between different types of witness reports, or different types of inference (Willett 1988; Aikhenvald 2004).

We are mostly interested in the strategies used by speakers allowing to determine the type of inference that underlies their assertion (if the assertion is based on an act of inference). It is customary in the literature on evidentials to group inferences into those that are based on observations and those that are based on reasoning.[11] The first class is typically said to include inferences from premises that the speaker has direct evidence for, whereas inferences based on reasoning are supposed to draw on general knowledge, common sense, or conjectures. Linguists working on evidentiality have, to

[10] According to Aikhenvald (2004), evidentiality is a grammatical category, and hence lexical items used to mark the source of information, which are available in all languages, are not evidentials in this strict, narrow sense. She argues that what can be found in English and many other European languages are mere *evidential strategies*. However, not all linguists agree on such a restrictive view. For a discussion of Aikhenvald's position see, e.g., Diewald and Smirnova (2010, p. 3-6).

[11] See, e.g., Willett (1988), Faller (2002), and Matthewson et al. (2007).

our knowledge, made no attempt to relate this distinction to the types of inference commonly distinguished in philosophy and logic,[12] and the definitions found in the literature on evidentiality are too vague to permit any definite conclusions on this point. Nevertheless, examples used to illustrate this linguistic typology at least somewhat suggest that the former is more closely related to abductive reasoning while the latter is more closely related to inductive reasoning.[13]

Given that we aim to do experimental work on inferential conditionals, and given that we can only recruit for our experiments native speakers of either English or Dutch, we are interested in strategies that speakers of these two languages, both of which are devoid of any grammatical evidentiality, can use to signal the evidential grounds for their assertions. It is always possible to convey information about one's evidential grounds in direct ways, as when we say that we *saw* that John crossed the street; or that it *seems* to us that Harriet is worried; or by the use of such words as "presumably," "apparently," "allegedly," or "they say," and so on. However, we hardly ever are explicit about the exact kind of inference that led us to the conclusion we are communicating. For instance, we do not normally say that we inductively inferred that the weather tomorrow will be nice or that we deduced that the beer is in the fridge. But in a more indirect way, we may sometimes indicate the type of inference underlying whatever it is that we are claiming to be the case.

In the linguistics literature (Aksu-Koç 1988; Matthewson et al. 2007; Haßler 2010), evidentiality has frequently been associated with modality. Even though modality and evidentiality are argued to be distinct categories (de Haan 1999; Aikhenvald 2004), they do seem to overlap. In particular, the category of epistemic modality and the category of inferential evidentiality seem to be closely related (van der Auwera and Plungian 1998; Faller 2002).

The idea that epistemic modals may function as evidential markers indicating the presence of an inference has also been put forward in the debate concerning the meaning of English epistemic "must." A number of authors

[12] Kwon (2012), who identified the Korean evidential *-napo-* as signalling the presence of an inductive inference, seems to be an exception.

[13] See, for instance, Matthewson et al. (2007, p. 205 ff) for an analysis of an evidential system of St'át'imcets with two inferential markers; or Smirnova (2012, p. 12 ff) for a discussion of a clearly abductive inferential evidential in Bulgarian.

have argued that insertion of this modal auxiliary verb makes an assertion weaker,[14] and that "It must be that p" does not entail p. They noted that, for instance, when an English speaker tries to call her friend's land line but does not get an answer, she may infer that her friend is out and express the resulting belief by saying:

(4) She must be out.

Were she to see her friend walking on the street, her assertion of (4) would seem odd or even inappropriate. On the other hand, there are contexts in which the content of an assertion is entailed by premises assumed in the context yet "must" does fit in. To give an example, if one knows that Mary has put a bottle of wine either in the fridge or in the cupboard, and one has checked that it is not in the cupboard, it would be perfectly all right to say:

(5) The bottle of wine must be in the fridge.

As noticed by von Fintel and Gillies (2007, 2010), what the uses of "must" in (4) and (5) have in common is that both signal the presence of an inference.[15]

While we are not aware of any relevant discussion in the literature of "should," this auxiliary, too, often seems to play the role of an inferential marker. For instance, when we are wondering about the translation of a phrase in Latin and we know that Susan studied classical languages for a number of years, we might say:

(6) Susan should be able to translate this phrase.

"Should" here seems to signal an uncertain inference: An assertion of (6) would seem odd if we knew that (say) the phrase is from a text which Susan recently published in English translation.

In the following, we assume that "must" and "should" can both serve as inferential markers. We will then be interested in the question of whether

[14] See, e.g., Karttunen (1972, p. 12), Groenendijk and Stokhof (1975, p. 69), Veltman (1985, p. 161 ff), and Kratzer (1991, p. 645).

[15] Dietz (2008, p. 246) also notes that in "It must be raining," the auxiliary indicates that the speaker only has (what he calls) "inferential evidence," and no direct observational evidence, that it is raining. See in the same vein Anderson (1986); Papafragou (1998); van der Auwera and Plungian (1998); Nuyts and Vonk (1999); Salmon (2011), and Mortelmans (2012).

the use of "must" and "should" gives us any indication as to what *kind* of inference (if any at all) led the speaker to feel warranted in making the assertion she did on the basis of the evidence she had.

Our hypothesis is that "must" marks either abductive or deductive inference, while "should" is rather a marker of inductive inference. Consider, for instance, that the inference underlying the assertion of (4) in our earlier example is most plausibly thought of as being abductive, that is, as an inference to the best explanation: that the friend is out is the best explanation for the evidence that the speaker has, to wit, that her friend does not answer the phone. In the example of Susan, it rather seems to be some form of inductive reasoning that warrants the assertion of (6): the people we met in our lives who had studied classical languages for a number of years were typically able to translate Latin phrases; given that Susan studied classical languages for a number of years, we expect her to be able to translate the designated phrase. That "must" may equally serve as a marker of a deductive inferential connection between evidence and assertion is suggested by (5).

We include in our study the epistemic adverb "probably," which we hypothesize to mark uncertain inference generically. While strictly speaking something that is certain could be said to be probable, neo-Gricean pragmatists have argued that saying of something one is certain of that it is *probably* the case generates the misleading (scalar) implicature that one is *not* certain of it (see, e.g., Levinson (1983, p. 134)). So, one would expect "probably" to go well with uncertain inferences, but not with certain ones.

We in fact want to broaden the scope of our investigations at least slightly by comparing "must" and "should" in their putative roles as inferential markers to what, according to our best judgement, are the closest counterparts of these markers in Dutch, the native language of two of the present authors, to wit, "moet" and "zou moeten."[16] In the Dutch study, we also look at "waarschijnlijk," which is the Dutch translation of "probably." The first study to be reported below concerns the English markers, the second one their Dutch counterparts.

[16]In Dutch, "should" is expressed by means of the verbal complex consisting of a counterfactual auxiliary "zou" and the infinitive "moeten" ("must"). See Huitink (2008) for a discussion of modal concord in Dutch.

It is to be noted that, ideally, these markers together yield something like an acid test for classifying conditionals. It cannot be generally read off from an inferential conditional to which type it belongs: a conditional that qualifies as a contextual DI conditional relative to one set of background premises may qualify as an AI or II conditional relative to another such set and similarly with the broader distinction between certain and uncertain inferential conditionals. The markers mentioned above may provide means of identifying the type to which an inferential conditional belongs, in that a speaker's use of a given marker in a conditional or her willingness to assent to a re-assertion of the conditional but now with a particular marker inserted, may show what kind of inference the speaker takes the conditional to embody.

Before we turn to the experiments, we want to clarify an aspect of our approach that might otherwise raise methodological worries. On the one hand, we are interested in a typology of conditionals that groups conditionals according to the type of inference that they embody. We aim to show that this typology helps explain how the assertability of conditionals can be influenced by inserting in them particular lexical expressions. On the other hand, we are interested in whether precisely those lexical expressions have the linguistic properties that they have been said to have by us and other authors, mainly based on intuitive judgements of a handful of examples (such as (4)–(6) above). The worry might now be that this involves us in circular reasoning, given that, as it would seem, we are assuming the truth of the one hypothesis in testing the second, and assuming the truth of the second in testing the first.

As has been argued by Glymour (1980) and as has been accepted by many philosophers since, there is nothing per se objectionable to using one hypothesis as an auxiliary in testing a second while also using the second as an auxiliary in testing the first. In fact, in his book Glymour gives many examples from the history of science that are generally considered to constitute good science in which this kind of mutual scaffolding occurs. As Glymour convincingly argues, what matters in this kind of situation is that the mutual scaffolding construction does not by itself *guarantee* success for the hypotheses involved and leaves open the possibility of failure; the test should be (what Glymour calls) *non-trivial*. Using a quantitative version of Glymour's theory of confirmation (as developed in Douven and Meijs

(2006)), we can even strengthen the non-triviality requirement by demanding that the scaffolding construction does not by itself make it more likely that we will get positive results for the hypotheses at issue.

The non-triviality requirement is clearly met in the present case, even in the more demanding probabilistic sense. Assuming that our and others' intuitions about which lexical markers go with which types of inferences are correct will do nothing to ensure, or even to make more likely, that there will be any pattern to be discovered in the data of our experiments that aligns in any significant way with how we are proposing to carve up the class of inferential conditionals. Conversely, assuming the typology marks theoretically importantly different classes of inferential conditionals does not make it any likelier that the lexical markers we consider will have any effect on the assertability of our stimuli, and, if they do have an effect, that effect might be completely out of line with our predictions (e.g., "should" might turn out to go better with abductive conditionals, "must" better with inductive conditionals, and "probably" better with deductive conditionals).

4. Experiment 1: English "should," "must," and "probably." Both of our experiments concern the typology of inferential conditionals discussed above. We look at a number of conditionals of the various types and consider whether they are perceived as more naturally assertable depending on whether or not "must," "should," or "probably" are inserted.

Before describing the experiment, we should be clear about the operational criteria that we will assume in determining whether an expression can count as an inferential marker. That an expression is a marker of a particular type of inference does not necessarily have to mean that inserting it in the consequent of a given conditional embodying that type of inference raises the conditional's degree of assertability. Even if the insertion leaves the degree of assertability more or less as it is, it serves as a marker for the type of inference if it does have an outspoken effect on the assertability of conditionals embodying other types of inference. For instance, an expression might have no effect on the degree of assertability of, say, II conditionals, and it would still qualify as an inferential marker of inductive inference if at the same time it lowers *substantively* the degree of assertability of the other types of inferential conditionals.

4.1. Method

Participants

Participants were recruited via the CrowdFlower interface,[17] which directed them to the Qualtrics platform[18] on which the experiment was run. The participants received a small amount of money for their participation. All participants were from Australia, Canada, the United Kingdom, and the United States. Of the 138 participants who started the survey, 136 completed it. We removed responses from participants who indicated a mother tongue other than English as well as from participants who failed at least one of two comprehension questions. The data are based on the remaining 68 participants. The average age of these participants was 35 (±11); the gender balance was 59 % females, 41 % males. Of these participants, 84 % had a college education or higher, 15 % only had a high school education, and the remaining 1 % had a lower level of education. The average time spent on the survey was 14 minutes (±18). On a scale from 1 (very easy) to 7 (very difficult), the survey was judged as 2.88 (±1.21) by the participants.

Design

The type of conditional (DI / AI / II) as well as the markers were manipulated within subjects.

Materials and Procedure

All materials were in English, the participants' mother tongue. Fifteen items were presented on screen. The participants were presented five items involving a DI conditional, five items involving an II conditional, and five items involving an AI conditional. Each item consisted of a context and four versions of the same conditional, where the first version had no marker and the second, third, and fourth versions contained "should," "must," and "probably," respectively. The participants were asked to rate on a seven-point scale the assertability of each version of the conditional in light of the given context. The order of the items was randomized per individual.

The following is an example of an item involving an AI conditional:

[17] `http://www.crowdflower.com`.

[18] `http://www.qualtrics.com`.

CONTEXT: Nelly lives on the sixth floor of an apartment building. The elevator has been broken since earlier this morning. A good friend of Nelly's who lives on the third floor of the same building hears someone rushing down the stairs. She knows that Nelly tends to avoid exercise as much as possible. How assertable are the following conditionals given this context?

CONDITIONAL: If that's Nelly rushing down the stairs, then she is in a hurry.

Highly unassertable 1 2 3 4 5 6 7 Highly assertable

CONDITIONAL: If that's Nelly rushing down the stairs, then she should be in a hurry.

Highly unassertable 1 2 3 4 5 6 7 Highly assertable

CONDITIONAL: If that's Nelly rushing down the stairs, then she must be in a hurry.

Highly unassertable 1 2 3 4 5 6 7 Highly assertable

CONDITIONAL: If that's Nelly rushing down the stairs, then she probably is in a hurry.

Highly unassertable 1 2 3 4 5 6 7 Highly assertable

See the Appendix for the rest of the materials.

Results and Discussion

We conducted for each of the three types of conditionals a separate one-way repeated measures ANOVA with type of marker (no marker / should / must / probably) as independent variable and degree of assertability as dependent variable.

For the three types of conditionals, DI, AI, and II, Mauchly's test indicated a violation of the assumption of sphericity ($\chi^2(5) = 95.17$, $p < .0001$; $\chi^2(5) = 97.08$, $p < .0001$; $\chi^2(5) = 80.01$, $p < .0001$; respectively). Because of this, the Huynh–Feldt correction was used to determine degrees

of freedom ($\epsilon = .852$ for the DI conditionals; $\epsilon = .839$ for the AI conditionals; and $\epsilon = .865$ for the II conditionals). The outcomes showed that assertability rates for all types of conditionals are significantly affected by type of marker: $F(2.56, 866.70) = 15.491$, $p < .0001$, for the DI conditionals; $F(2.51, 853.34) = 164.221$, $p < .0001$, for the AI conditionals; and $F(2.60, 880.03) = 221.169$, $p < .0001$, for the II conditionals.

The mean assertability ratings for the three types of conditionals with and without the various markers are given in Table 1. Inspecting the means for the DI conditionals shows that inserting any of the markers negatively impacts assertability. Post hoc tests using Bonferroni's adjustment revealed that the means for "should," "must," and "probably" are all significantly lower than the mean for no marker ($p < .0001$ for "should" and "must"; $p = .016$ for "probably"). Also, the mean for "should" is significantly lower than the means for "must" ($p = .017$) and "probably" ($p = .001$). As for the AI conditionals, the mean for "probably" is highest, that for "must" is second highest, followed by the mean for no marker. The lowest mean is for "should." Post hoc tests using Bonferroni's adjustment showed that the mean for "should" is indeed significantly lower than the others (all ps $< .0001$). The difference between the means for "must" and no marker is not significant. The mean for "probably" is significantly higher than the other means (all ps $< .0001$). Finally, for the II conditionals, the mean for "probably" is again highest but is now followed by that for "should." The mean for "must" is lowest. Post hoc tests using Bonferroni's adjustment showed that the mean for "probably" is significantly higher than the other means, while the mean for "must" is significantly lower than the other means (all ps $< .0001$). The mean for "should" is significantly higher than those for no marker and for "must" (both ps $< .0001$).

In order to make the impact that the insertion of the markers has on the different types of conditionals easier to see, Figure 1 plots the differences in the mean *relative* assertability of the conditionals. By the relative assertability of a conditional, we mean the degree of assertability of a conditional with a marker minus the degree of assertability of the conditional without marker. The graph clearly shows that inserting "should" has a positive impact on the assertability of II conditionals but a negative impact on the assertability of both DI and AI conditionals. By contrast, the insertion of "must" has a somewhat positive impact on the assertability of AI conditionals and a

Table 1: Mean assertability (averaged over the five items per type of inference) and standard deviations (SD) for the three types of inferential conditionals.

	DI		AI		II	
	Mean	SD	Mean	SD	Mean	SD
no marker	6.52	0.87	5.01	1.32	4.71	1.59
should	6.11	1.13	4.13	1.56	5.38	1.32
must	6.29	1.07	5.14	1.51	4.06	1.65
probably	6.32	1.01	6.01	1.02	6.15	0.96

somewhat, respectively strong, negative impact on the assertability of DI and II conditionals. "Probably" does very well with AI and II conditionals but less so with DI conditionals.

Given how we previously operationalised the notion of an inferential marker, the above findings support the hypothesis that "should" serves as an inductive inferential marker, that "must" serves as an abductive inferential marker, and that "probably" is a marker of uncertain inference. On the other hand, we found no evidence for the claim that "must" serves as a deductive inferential marker, which was suggested by considering (5).

5. Experiment 2: Dutch "zou moeten," "moet," and "waarschijnlijk." We wanted to investigate whether repeating the first experiment in Dutch will yield similar results. We hypothesized that the Dutch expressions "zou moeten" and "moet" come closest to being equivalents, *qua* inferential markers, of "should" and "must," respectively. And given that "waarschijnlijk" is a straightforward translation of "probably," it may be expected to have a similar role as a marker of uncertainty.

5.1. Method

Participants

We recruited Dutch and Flemish participants via CrowdFlower ($N = 25$) and via social media ($N = 19$), which directed them to the Qualtrics plat-

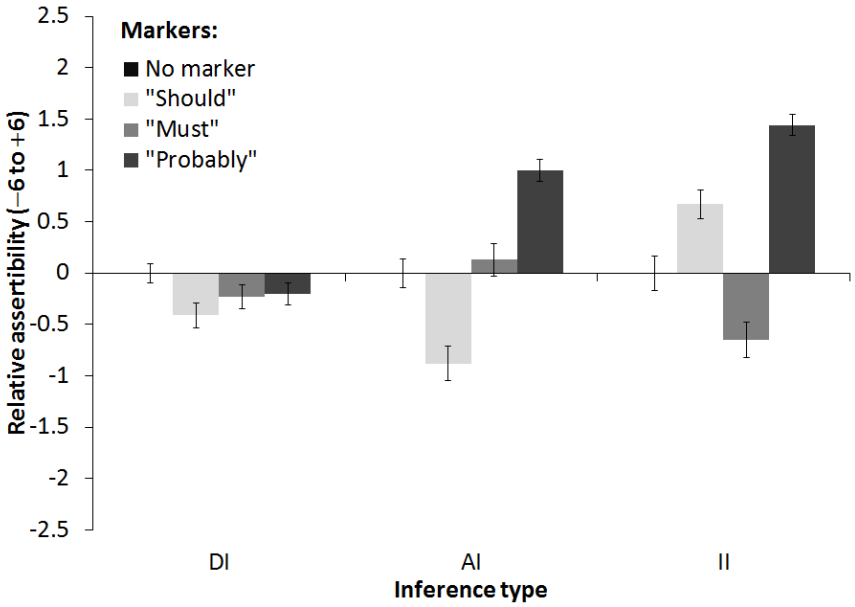

Figure 1: Effect of the various English markers on relative assertability for the different types of conditionals. Error bars represent 95 % confidence intervals.

form on which the survey was run. The former participants were paid a small amount of money in return for their cooperation. We excluded from the analysis participants who did not complete the survey as well as participants who answered incorrectly at least one of the two comprehension questions. This left us with 15 participants. The average age of these participants was 35 (±11), with 67 % of them being females. Of these participants, 60 % had a college education or higher, 27 % had only high school, and the remaining 13 % had a lower level of education. The average time spent on the survey was 19 (±13) minutes. On a scale from 1 (very easy) to 7 (very difficult), the survey was judged as 2.81 (±1.21) by the participants.

Design

The type of conditional (DI / AI / II) as well as the markers were manipu-

lated within subjects.

Materials and Procedure

We used the same materials as in Experiment 1 translated into Dutch, the participants' mother tongue. Here too, the first version of each conditional that was presented had no marker and the second, third, and fourth versions contained "zou moeten," "moet," and "waarschijnlijk," respectively. Participants were again asked to rate the assertability of each version of the conditional in light of the given context. The order of the items was randomized per individual.

Results and Discussion

The analysis was identical to the analysis of Experiment 1. We again conducted for each of the three types of conditionals a separate one-way repeated measures ANOVA with type of marker (no marker / zou moeten / moet / waarschijnlijk) as independent variable and degree of assertability as dependent variable.

For the DI and AI conditionals, Mauchly's test indicated a violation of the assumption of sphericity ($\chi^2(5) = 38.76$, $p < .0001$, and $\chi^2(5) = 14.01$, $p = .016$, respectively). The Huynh–Feldt correction was used to determine degrees of freedom ($\epsilon = .782$ for the DI conditionals and $\epsilon = .913$ for the AI conditionals). The outcomes showed that assertability rates for both types of conditionals are significantly affected by type of marker: $F(2.35, 173.69) = 11.179$, $p < .0001$, for the DI conditionals, and $F(2.74, 202.69) = 59.380$, $p < .0001$ for the AI conditionals. For the II conditionals, the outcomes also showed that assertability rates are significantly affected by type of marker, $F(3, 222) = 47.536$, $p < .0001$. (Mauchly's test did not reach statistical significance in this case, so no sphericity corrections were applied.)

Table 2 gives the mean assertability ratings for the three types of conditionals with and without the various markers. Also for the Dutch DI conditionals, inserting any of the markers has a negative impact on assertability. Post hoc tests using Bonferroni's adjustment showed that the means for "zou moeten," "moet," and "waarschijnlijk" are significantly lower than the mean for no marker ($p = .002$ for "zou moeten"; $p = .001$ for "moet"; $p < .0001$ for "waarschijnlijk"). The differences between the means for the other markers did not reach significance. Comparable to what we found for the English AI conditionals, for the Dutch AI conditionals the mean for "waarschijnlijk" is highest, followed by that for "moet," which is followed by

Table 2: Mean assertability (averaged over the five items per type of inference) and standard deviations (SD) for the three types of inferential conditionals.

	DI		AI		II	
	Mean	SD	Mean	SD	Mean	SD
no marker	6.05	1.22	4.71	1.63	4.04	1.68
zou moeten	5.19	1.63	3.65	1.52	4.53	1.66
moet	5.40	1.34	5.07	1.40	3.69	1.72
waarschijnlijk	4.80	1.73	6.16	0.92	6.08	1.10

the mean for no marker. The lowest mean is for "zou moeten." Here, too, post hoc tests using Bonferroni's adjustment showed that the mean for "zou moeten" is significantly lower than the other means (all ps $< .0001$); the mean for "waarschijnlijk" is significantly higher than the other means (all ps $< .0001$); and the mean for "moet" is higher than that for no marker, but not significantly so. And for the II conditionals, the mean of "waarschijnlijk" is again highest, followed by that for "zou moeten". The mean for "moet" is lowest. Post hoc test using Bonferroni's adjustment revealed that the mean for "waarschijnlijk" is significantly higher than the other means, while the mean for "moet" is significantly lower than the means for "zou moeten" and "waarschijnlijk" (both ps $< .0001$) but not significantly lower than the mean for no marker. By contrast to the result for "should" in the context of II conditionals, the mean for "zou moeten" is not significantly higher than the mean for no marker.

Again, it is easiest to show the impact of the various markers on the assertability of the different types of conditionals by plotting the differences in the mean relative assertability of the conditionals (see Figure 2). The patterns are qualitatively almost identical to the ones shown in Figure 1 (there is a small difference between the pattern for "must" and that for "moet"). Similarly to "should," "zou moeten" has a positive impact on the assertability of II conditionals but a negative impact on the assertability of DI and AI conditionals. And similarly to "must," "moet" has a somewhat positive impact on the assertability of AI conditionals and a negative impact

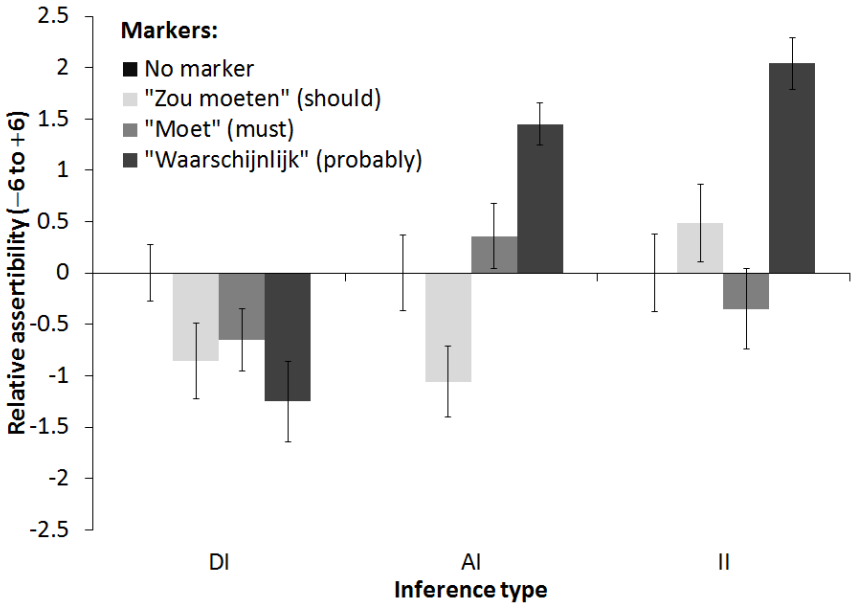

Figure 2: Effect of the various Dutch markers on relative assertability for the different types of conditionals. Error bars represent 95 % confidence intervals.

on the assertability of DI and II conditionals. Finally, "waarschijnlijk" does, like its English translation, very well with AI and II conditionals but less so—even much less so—with DI conditionals.

These results support the hypothesis that "zou moeten" serves as an inductive inferential marker, "moet" serves as an abductive inferential marker, and "waarschijnlijk" serves as a generic marker of uncertain inference. Here too, there is no evidence that "moet" also serves as a deductive inferential marker.

6. General discussion. In Douven and Verbrugge (2010), it was shown that the typology of inferential conditionals proposed in that paper helps to explain certain systematic differences in how people's acceptability judge-

ments of conditionals relate to their corresponding conditional degrees of belief. That was the first piece of evidence in favour of our hypothesis that the typology cuts at the joints. We have added to this another piece of evidence by showing that the typology helps to explain systematic differences in how people's assertability judgements vary depending on whether a marker and, if so, which marker, is inserted in a conditional.

As was predicted, both English "probably" and its Dutch equivalent, "waarschijnlijk", have a tendency to increase the assertability of uncertain (that is, II and AI) conditionals when added to the consequent of such a conditional and an opposite tendency to decrease the assertability of DI conditionals. It is worth noting that, as Figures 1 and 2 show, the effect of "probably" on the assertability ratings of II conditionals in both English and in Dutch is stronger than its effect on the assertability ratings of AI conditionals. We speculate that this is because the defeasibility of inductive inferences tends to be more conspicuous to people than that of abductive inferences. Stating that something happens *most* of the time, or that it has a 95 % chance of happening, conveys the information that it does not happen all the time or that it is not certain to happen. Hence, the very premises of an inductive argument direct a hearer's attention to the possibility of an exception, whereas the conclusion of an abductive argument, which is supposed to be the *best* explanation of (one of) its premises, might be easily thought to be the *only* explanation given that alternative explanations are often hard to conceive.

Our studies further support the hypothesis that the English modal verb "should" functions as an evidential marker of inductive inference. This may not seem a surprising result considering that, for instance, the Cambridge Dictionary Online lists one of the uses of "should" as "[showing] when something is likely or expected,"[19] and people's expectations regarding future events often result from inductive reasoning. Given that inductive inference is overtly defeasible, it might be perceived as providing relatively weak grounds for an assertion. But a comparison with "probably" suggests that "should" does more than just weaken an assertion given that, unlike "probably," it does not go well with abductive conditionals. Similar remarks apply to the Dutch expression "zou moeten."

[19] `http://dictionary.cambridge.org/dictionary/british/should_2`.

We saw that both "must" and "moet" have a positive effect on the assertability of AI conditionals, though for neither marker does this effect reach statistical significance. However, the combined fact that "must" and "moet" have a negative effect on the assertability of II conditionals, a weaker negative effect on the assertability of DI conditionals, and yet no negative effect (and in fact an, albeit insignificant, positive effect) on the assertability of AI conditionals constitutes a pattern which warrants the conclusion that "must" and "moet" are evidential markers indicating abductive inference.

Acknowledgments. We are greatly indebted to two anonymous referees and to the editors, Rineke Verbrugge and Jakub Szymanik, for valuable comments. Previous versions of this paper were presented at the Logic & Cognition Workshop at ESSLLI 2012 and at the London Reasoning Workshop 2013. We thank the audiences on those occasions for stimulating questions and discussions.

Appendix. This appendix presents the contexts and conditionals that, together with the example given in Section 4, were used in Experiment 1. Experiment 2 used the Dutch translation of these materials. We state here only the conditionals without marker. From these and the explanation in Section 4, the versions of the conditionals with the markers can be straightforwardly inferred.

DI Items:

Context: All students in class 6C have at least a B for their math test paper.
Conditional: If Ben is in class 6C, then he has at least a B for his math test paper.

Context: Last year, all people older than 65 have been vaccinated for the flu.
Conditional: If Mrs Harris is 70 years old, then she has been vaccinated for the flu.

Context: All Indian elephants have small ears.
Conditional: If Babou is an Indian elephant, then it has small ears.

Context: All white cats possess a gene that predisposes them to develop blindness late in their lives.
Conditional: If Paul's kitten is white, then it possesses a gene that predisposes it to develop blindness late in its life.

Context: Two friends are wondering whether Cynthia passed the exam. They know that it was an absolute requirement for the exam to hand in a thesis before the end of the semester.
Conditional: If Cynthia did not hand in her thesis before the end of the semester, then she failed.

AI Items:

Context: You know that Tom and Hank recently had a flaming row which, you think, ended their friendship for good. Now a friend tells you that she thinks she just saw Tom and Hank jogging together.
Conditional: If Tom and Hank are jogging together again, then they are friends again.

Context: Someone tells you that a nearby village, located in a valley below a dammed reservoir, has been flooded. You doubt that this is true. On the other hand, the dam has been in a rather bad state for some time.
Conditional: If the village has been flooded, then the dam has broken.

Context: Judy is waiting for the train. She is looking for her iPod to listen to some music while she waits. It is not in her coat. Yet she is sure that she took it this morning. Perhaps it is in her bag. Then she sees that the bag has been cut open. At that moment there is an announcement that pickpockets are active in the train station.
Conditional: If Judy's iPod is not in her bag, then someone has stolen it.

Context: Pete had to play the finals of a tennis tournament earlier today. Two friends of his, who do not yet know the result of the match, are walking to Pete's house. Pete is not really a party-person, but from a distance, it seems to them that there is a party going on in Pete's garden.
Conditional: If Pete is partying, then he has won the match.

II Items:

Context: 99 % miners develop silicosis, a disease caused by inhaling fine dust for a prolonged period of one's life.
Conditional: If Rudolph has worked in the mines for all his life, then he has developed silicosis.

Context: According to the local bus company, none of their buses has been more than 5 min late in the past 10 years. You are presently waiting for a bus of this company.
Conditional: If our bus is not exactly on time, then it will be at most a few minutes late.

Context: A pharmaceutical company unexpectedly got into financial trouble. They had to cut many jobs and decided to fire most employees above 50. Mark is an employee of this company.
Conditional: If Mark is above 50, then he is among the employees who will be fired.

Context: Bernard is a bit of an irregular student: sometimes he works hard, but he can also be lazy. So far he had excellent grades for most courses for which he had worked hard.
Conditional: If Bernard works hard for the linguistics course, then he will get an excellent grade for it.

Context: The hepatitis A virus may be transmitted by contact with an infected person, so people working in health care are at a higher risk of getting ill from the virus. The vaccine against hepatitis A is 95 % effective. Adam has recently started volunteering at a hospital.
Conditional: If Adam has been vaccinated against hepatitis A, then he will not get ill from the virus.

References

Adams, E. W. (1965). The logic of conditionals. *Inquiry*, 8:166–197.

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford University Press, Oxford.

Aksu-Koç, A. (1988). *The Acquisition of Aspect and Modality: The Case of Past Reference in Turkish*. Cambridge University Press, Cambridge.

Anderson, L. B. (1986). Evidentials, paths of change and mental maps: Typologically regular asymmetries. In Chafe, W. L. and Nichols, J., editors, *Evidentiality: The Linguistic Coding of Epistemology*, pages 273–312. Ablex, Norwood, NJ.

van der Auwera, J. and Plungian, V. (1998). Modality's semantic map. *Linguistic Typology*, 2:79–124.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press, Oxford.

Cialdea Mayer, M. and Pirri, F. (1993). First order abduction via tableau and sequent calculi. *Logic Journal of the IGPL*, 1:99–117.

Cialdea Mayer, M. and Pirri, F. (1995). Propositional abduction in modal logic. *Logic Journal of the IGPL*, 3:907–919.

Dancygier, B. (1998). *Conditionals and Predictions: Time, Knowledge and Causation in Conditional Constructions*. Cambridge University Press, Cambridge.

Dancygier, B. (2003). Classyfying conditionals: Form and function. *English Language and Linguistics*, 7:309–323.

Dancygier, B. and Sweetser, E. (2005). *Mental Spaces in Grammar: Conditional Constructions*. Cambridge University Press, Cambridge.

Declerck, R. and Reed, S. (2001). *Conditionals: A Comprehensive Empirical Analysis*. Mouton de Gruyter, Berlin/New York.

Dietz, R. (2008). Epistemic modals and correct disagreement. In García-Carpintero, M. and Kölbel, M., editors, *Relative Truth*, pages 23–262. Oxford University Press, Oxford.

Diewald, G. and Smirnova, E. (2010). Introduction. Evidentiality in European languages: the lexical-grammatical distinction. In Diewald, G. and Smirnova, E., editors, *Linguistic Realization of Evidentiality in European Languages*, pages 1–14. Mouton de Gruyter, Berlin/New York.

Douven, I. (2011). Indicative conditionals. In Horsten, L. and Pettigrew, R., editors, *A Companion to Philosophical Logic*, pages 383–405. Continuum Press, London.

Douven, I. (2013). Inference to the best explanation, Dutch books, and inaccuracy minimisation. *Philosophical Quarterly*, 69:428–444.

Douven, I. and Meijs, W. (2006). Bootstrap confirmation made quantitative. *Synthese*, 149:97–132.

Douven, I. and Verbrugge, S. (2010). The Adams family. *Cognition*, 117:302–318.

Edgington, D. (1995). On conditionals. *Mind*, 104:235–329.

Faller, M. (2002). *Semantics and Pragmatics of Evidentials in Cuzco Quechua*. PhD thesis, Stanford University, Palo Alto, CA.

von Fintel, K. and Gillies, A. (2007). An opinionated guide to epistemic modality. *Oxford Studies in Epistemology*, 2:32–63.

von Fintel, K. and Gillies, A. S. (2010). *Must* ... stay ... strong! *Natural Language Semantics*, 18:351–383.

Gabbay, D. M. and Woods, J. (2005). *The Reach of Abduction*. Elsevier, Amsterdam.

Glymour, C. (1980). *Theory and Evidence*. Princeton University Press, Princeton.

Groenendijk, J. A. and Stokhof, M. J. (1975). Modality and conversational information. *Theoretical Linguistics*, 2:61–112.

de Haan, F. (1999). Evidentiality and epistemic modality: Setting boundaries. *Southwest Journal of Linguistics*, 18:83–101.

Haegeman, L. (2003). Conditional clauses: External and internal syntax. *Mind & Language*, 18:317–339.

Haßler, G. (2010). Epistemic modality and evidentiality and their determination on a deictic basis: the case of Romance languages. In Diewald, G. and Smirnova, E., editors, *Linguistic Realization of Evidentiality in European Languages*, pages 223–248. Mouton de Gruyter, Berlin/New York.

Huitink, J. (2008). *Modals, Conditionals, and Compositionality*. PhD thesis, Radboud Universiteit Nijmegen.

Karttunen, L. (1972). "Possible" and "Must". In Kimball, J., editor, *Syntax and Semantics*, volume 1, pages 1–20. Academic Press, New York.

Kratzer, A. (1991). Modality. In von Stechow, A. and Wunderlich, D., editors, *Semantics: An International Handbook of Contemporary Research*, pages 38–74. de Gruyter, Berlin.

Krzyżanowska, K. (2012). Ambiguous conditionals. In Stalmaszczyk, P., editor, *Philosophical and Formal Approaches to Linguistic Analysis*, pages 315–332. Ontos Verlag, Frankfurt.

Krzyżanowska, K., Wenmackers, S., and Douven, I. (2013). Rethinking Gibbard's riverboat argument. Forthcoming in *Studia Logica*.

Kwon, I. (2012). Please confirm what I inferred: On the Korean inferential-evidential marker *-napo-*. *Journal of Pragmatics*, 44:958–969.

Kyburg, H. and Teng, C. M. (2001). *Uncertain Inference*. Cambridge University Press, Cambridge.

Levinson, S. (1983). *Pragmatics*. Cambridge University Press, Cambridge.

Matthewson, L., Davis, H., and Rullmann, H. (2007). Evidentials as epistemic modals: Evidence from St'át'imcets. In Van Craenenbroeck, J. and Rooryck, J., editors, *Linguistic Variation Yearbook*, volume 7, pages 201–254. John Benjamins, Amsterdam.

Mortelmans, T. (2012). Epistemic MUST and its cognates in German and Dutch. *Journal of Pragmatics*, 44:2150–2164.

Nuyts, J. and Vonk, W. (1999). Epistemic modality and focus in Dutch. *Linguistics*, 37:699–737.

Papafragou, A. (1998). Inference and word meaning: The case of modal auxiliaries. *Lingua*, 105:1–47.

Salmon, W. (2011). Conventional implicature, presupposition, and the meaning of *must*. *Journal of Pragmatics*, 43:3416–3430.

Sanford, D. H. (1989). *If P, Then Q: Conditionals and the Foundations of Reasoning*. Routledge, London.

Smirnova, A. (2012). Evidentiality in Bulgarian: Temporality, epistemic modality, and information source. *Journal of Semantics*, doi: 10.1093/jos/ffs017.

Veltman, F. (1985). *Logics for Conditionals*. PhD thesis, University of Amsterdam.

Verbrugge, S. (2007). *A Psycholinguistic Analysis of Inferential Conditional Sentences*. PhD thesis, Katholieke Universiteit Leuven.

Willett, T. (1988). A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language*, 12:51–97.

Woods, M. (2003). *Conditionals*. Oxford University Press, Oxford.