In the previous chapter, we introduced a new semantics based on an old idea. According to this semantics, what a conditional sentence conveys is the existence of an inferential connection between its antecedent and its consequent (3.4). This semantics takes into account a well-established philosophical tradition of distinguishing between deductive, inductive and abductive inferences whose relevance for conditionals has been first observed by Douven and Verbrugge (2010). In the following, we aim to provide further support for the typology of conditionals introduced in section 3.3 of the previous chapter and for the semantics based on it. We do this by relating the typology 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 4.2 and 4.3, 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.

Parts of this chapter, including the reported experiment, have been published as a joint paper with Sylvia Wenmackers and Igor Douven (Krzyżanowska et al. 2013).

4.1 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, 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. Note that, 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.1 Here, the terms “evidentiality” and “evidential markers” are used in a broader sense that extends to lexical items used in evidential strategies.

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

---

1 For a discussion of Aikhenvald’s position see, e.g., Diewald and Smirnova (2010, p. 3-6).
4.1 EVIDENTIAL MARKERS

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 (also referred to as inferences “from results”) and those that are based on reasoning (“inferences from reasoning;” see, e.g., Willett 1988; Faller 2002; Matthewson et al. 2007). 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, previous experience or conjectures. Linguists working on evidentiality have, to our knowledge, made no attempt to relate this distinction to the types of inference commonly distinguished in philosophy and logic with a notable exception of Kwon (2012), who identified the Korean evidential -napo- as signalling the presence of an inductive inference. Principally, 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. For instance, an inferential evidential in Bulgarian, discussed by Smirnova (2012), clearly indicates an abductive inference, as illustrated by the following example:

Context: You and your sister were out of touch for a couple of years. Today you visit her for the first time. As she shows you around her apartment, you see that there is a piano in her daughter Maria’s room. You infer that Maria plays the piano. Later, you tell your husband:

(83) “Maria plays the piano, [I inferred].”

By contrast, if the speaker concludes that Maria plays the piano on the basis of what she knows about Maria’s mother’s plans for her daughter (conjoined perhaps with a high degree of certainty that the mother achieved what she wanted), which would be an example of inductive inference, (83) with the evidential form of the verb “to play” is not felicitous (p. 13).
On the other hand, the evidential system of St’át’imcets is equipped with two inferential markers (Matthewson et al. 2007), which both seem to be appropriate when the reasoning is abductive, yet one of them can only be used when the premise from which the conclusion is drawn, is provided by a direct perception. Matthewson illustrated the difference by the following example (p. 206):

**Context:** You are a teacher and you come into your classroom and find a caricature of you drawn on the blackboard. You know that Sylvia likes to draw caricatures.

(84) a. “It was [must have been] Sylvia who did it.”

b. “[Apparently], it was Sylvia who did it.”

In the above context, (84b), which involves an evidential marker signalling reasoning from what has been directly perceived, is not felicitous since the speaker lacks perceptual evidence for their conclusion. However, if the speaker saw Sylvia hiding behind the door, they could express their conclusion by means of any of the two inferential evidentials. Here the distinction between inferences “from results” and inferences “from reasoning” clearly applies to how the premises are justified, rather than the type of reasoning itself.

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 are hardly ever 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.

---

2 A language of indigenous inhabitants of British Columbia, Canada, also known as Lilooet.
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). In fact, Verbrugge (2007b, Ch. 6) established a close connection between Dutch modal expressions “zal wel” (which translates as “may well”) and “waarschijnlijk,” and inferential conditionals. Specifically, in an elicitation task in which participants were requested to complete conditionals whose antecedents were given, they tended to come up with a significantly higher number of inferential conditionals (as opposed to content conditionals) when they were in addition requested to use “zal wel” in the consequents than when they were not (see, however, the note on content conditionals in 3.3.2).

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, e.g., Karttunen (1972, p. 12), Groendijk and Stokhof (1975, p. 69), Veltman (1985, p. 161 ff), and Kratzer (1991, p. 645), have argued that insertion of this modal auxiliary verb makes an assertion weaker, and that “It must be that p” does not entail p. They noted that, for instance, when an English speaker notices a fresh cup of coffee and a pile of new papers on a desk of a colleague who is not present at the moment and has been on vacation for the last couple of days, they may infer that their colleague is back at work and express the resulting belief by saying:

(85) She must be back.

Were they to see their colleague sitting at her desk, their assertion of (85) 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:

(86) The bottle of wine must be in the fridge.
As noticed by von Fintel and Gillies (2007, 2010), what the uses of “must” in (85) and (86) have in common is that both signal the presence of an inference.3

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:

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

“Should” here seems to signal an uncertain inference: An assertion of (87) 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 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 (85) in our earlier example is most plausibly thought of as being abductive, that is, as an inference to the best explanation: that the colleague is back is the best explanation for the evidence that the speaker has, to wit, that there is a fresh cup of coffee and a pile of new papers laying on her desk. In the example of Susan, it rather seems to be some form of inductive reasoning that warrants the assertion of (87): 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 (86).

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

---

3 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).
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 comes across as being the closest counterparts of these markers in Dutch, the native language of two of the researchers involved in the study, to wit, “moet” and “zou moeten.” 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.

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.

4.1.1 Note on the methodology

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

4 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.
other authors, mainly based on intuitive judgements of a handful of examples (such as (85)–(87) 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).
4.2 EXPERIMENT 1. ENGLISH MARKERS: “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.

4.2.1 Method

Participants

Participants were recruited via the CrowdFlower interface,\(^5\) which directed them to the Qualtrics platform\(^6\) 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

\(^5\) \text{http://www.crowdflower.com.}\n\(^6\) \text{http://www.qualtrics.com.}
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.

An example of an item involving an AI conditional is given in Figure 3. See the Appendix A for the rest of the materials.

4.2.2 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.
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.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>unassertable</td>
<td>Highly assertable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>unassertable</td>
<td>Highly assertable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>unassertable</td>
<td>Highly assertable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>unassertable</td>
<td>Highly assertable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: An example of an AI item used in the Experiment 1.

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 signific-
antly 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 $ps < .0001$).

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

<table>
<thead>
<tr>
<th></th>
<th>DI</th>
<th></th>
<th>AI</th>
<th></th>
<th>II</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>no marker</td>
<td>6.52</td>
<td>0.87</td>
<td>5.01</td>
<td>1.32</td>
<td>4.71</td>
<td>1.59</td>
</tr>
<tr>
<td>should</td>
<td>6.11</td>
<td>1.13</td>
<td>4.13</td>
<td>1.56</td>
<td>5.38</td>
<td>1.32</td>
</tr>
<tr>
<td>must</td>
<td>6.29</td>
<td>1.07</td>
<td>5.14</td>
<td>1.51</td>
<td>4.06</td>
<td>1.65</td>
</tr>
<tr>
<td>probably</td>
<td>6.32</td>
<td>1.01</td>
<td>6.01</td>
<td>1.02</td>
<td>6.15</td>
<td>0.96</td>
</tr>
</tbody>
</table>

In order to make the impact that the insertion of the markers has on the different types of conditionals easier to see, Figure 4 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 somewhat, respectively strong, negative impact on the assertability of DI and II conditionals. “Probably”
Figure 4: Effect of the various English markers on relative assertability for the different types of conditionals. Error bars represent 95% confidence intervals.

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 (86).

4.3 **EXPERIMENT 2. DUTCH MARKERS: “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.
4.3.1 Method

Participants

We recruited Dutch and Flemish participants via CrowdFlower (N = 25) and via social media (N = 19), which directed them to the Qualtrics platform 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 manipulated within subjects.

Materials and Procedure

We used the same materials as in Experiment 1 translated into Dutch, the participants’ mother tongue (see the Appendix B). 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.

4.3.2 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

<table>
<thead>
<tr>
<th></th>
<th>DI</th>
<th></th>
<th>AI</th>
<th></th>
<th>II</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>no marker</td>
<td>6.05</td>
<td>1.22</td>
<td>4.71</td>
<td>1.63</td>
<td>4.04</td>
<td>1.68</td>
</tr>
<tr>
<td>zou moeten</td>
<td>5.19</td>
<td>1.63</td>
<td>3.65</td>
<td>1.52</td>
<td>4.53</td>
<td>1.66</td>
</tr>
<tr>
<td>moet</td>
<td>5.40</td>
<td>1.34</td>
<td>5.07</td>
<td>1.40</td>
<td>3.69</td>
<td>1.72</td>
</tr>
<tr>
<td>waarschijnlijk</td>
<td>4.80</td>
<td>1.73</td>
<td>6.16</td>
<td>0.92</td>
<td>6.08</td>
<td>1.10</td>
</tr>
</tbody>
</table>

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 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 5). The patterns are qualitatively almost identical to the ones shown in Figure 4 (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 neg-

Figure 5: Effect of the various Dutch markers on relative assertability for the different types of conditionals. Error bars represent 95% confidence intervals.
ative impact 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.

4.4 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 judgements 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 4 and 5 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,” 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.”

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. Given that this is the typology of conditionals discussed in chapter 3 which allows us to explain the systematic differences observed in our experiments, we take these results to be a piece of evidence in favour of our proposal. In the next chapter, we will show that the semantics built upon the said typology helps to evade certain challenges that many theories of indicative conditionals have to face. Additionally, we will see that our proposal’s ability to distinguish between the types of inferential conditionals allows us to reevaluate the role indicative conditionals play in the process of decision making.

7 http://dictionary.cambridge.org/dictionary/british/should_2.