On the Linguistic Encoding of the Notion of Inference

A dissertation presented
by

Giuseppe Ricciardi

to

The Department of Linguistics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Linguistics

Harvard University
Cambridge, Massachusetts

September 2022

Dissertation Advisor: Kathryn Davidson   Giuseppe Ricciardi

On The Linguistic Encoding of the Notion of Inference

Abstract

In our everyday life we often talk about events that we have not witnessed but just inferred based on some evidence available to us. This dissertation investigates two types of linguistic expressions that people use when talking about inferred events: inferentials and epistemics.

Chapter 2 investigates the meaning of inferential expressions through the case study of the Italian non-temporal future, e.g., *Susanna sarà malata ora* ('Susan will be sick now'). I argue - based on introspection - that the semantic contribution of inferentials in a clause – shortly 'INFER-p' – is to convey as a not-at-issue proposition that the contextually defined individual who is understood as undergoing a commitment to the truth of 'p' (called here 'the judge of p') has obtained 'p' through an inference. Moreover, I argue that inferentials do not encode an epistemic evaluation of 'p': the perception that an inferential weakens the judge's commitment to the truth of 'p' is due to the combination of an extralinguistic considerations pertaining to how people perceive the reliability of information obtained through an inference and the competition with bare 'p' which is the natural way to convey knowledge.

Chapters 3 and 4 investigate the meaning of epistemic necessity auxiliary verbs through the case study of the English word 'must' and the Italian word 'dovere', e.g., *Susan must be sick now*. I argue - based on a combination of data coming from experimental findings, corpus-searched utterances, and introspection - that epistemic 'must' and 'dovere' are polysemous between two meanings: the meaning of an epistemic necessity operator (defining an event as certain based on some evidence) and the meaning of

an inferential (defining an event as having been obtained by the relevant judge through an inference). Furthermore, I speculate that the inferential sense of the 'must' and 'dovere' is a derivative meaning of the words stemming from their overuse as a marker of epistemic necessity: the original meaning indicating certain conclusions is weakened to indicate just conclusions (i.e., opinions) of the judge, after reiterated and implausible exaggerated uses of the words in their original sense.

Chapter 5 offers an assessment of the debate about the relevant perspective for defining the truth-value of bare epistemic possibility statements: the utterer or the assessor. I offer findings from two studies suggesting that in scenarios where the statement is appropriately asserted from the point of view of the speaker but does not correspond to how in reality things are English speakers disagree on the truth-value not only of 'might p' but also, surprisingly, of bare 'p'. I argue that underlying this behavior is not a disagreement on the semantic content of the statements but a disagreement on how to interpret the meaning of the adjectives 'true' and 'false' when applied to a statement uttered by another agent. Specifically, I suggest that 'true' and 'false' have an inherent relative nature – to decide if a statement as uttered by A is true one has to first decide a perspective (A or the assessor).

Contents

Acknowledgments

Many people have contributed to making me reach this stage.

Andrea Moro and Gennaro Chierchia, who first believed in me and without whom all of this would not have been possible.

Kate Davidson, who has been acting as my advisor throughout the process of writing the thesis and has been providing me with an example of the type of scholar that I could be.

Ted Gibson, who has been such an unvaluable mentor during these years where excitement was mixed with frustration. He has pushed me to communicate more efficiently, an ability that lies at the heart of being a great professor and researcher.

Susi Wurmbrand, who has helped me to define the structural counterpart to the semantic insights underlying the work presented here.

Manfred Krifka, whose work has been an inspiration for me and who has enthusiastically contributed to improve the arguments that I present here.

Helen Lewis, who has been constantly reminding me that I should not overthink and learn to compromise.

My cohort peers Yingtong, Zach, and Shannon, who have endured this burden with me for 6 years or so by providing emotional support and mutual help. Without Yingtong I would not be able to go over the first courses in statistics. With Shannon I have been having some of the most engaging conversations about semantics and life in general; I only regret that we have not had the chance to collaborate on a project: perhaps in the future.

The other amazing graduate students that I have encountered during my Ph.D. One special mention for Yuhan, who I have seen growing in self-awareness and become a great friend.

My Italian friends Gabriele, Elvira, Elisabetta, Carlo Alberto, Niccolò, Alberto who have been there when I needed to have fun and talk about other things than linguistics.

Yijing, who I met at the very beginning of this experience when we were both young (she much more than me) and who has been a constant point of reference for me when I needed to put myself together after a crisis.

My siblings Roberto, Antonella, Francesco, who have seen me before I started all of this and will be seeing who I became after.

My parents Maria and Tommaso, whose care has made me reach this point in my life safely. They taught me the attitude which I have been adopting in approaching this work: be true to yourself and don't go with the wind.

Chapter 1. Introduction

1.0 Preliminaries: how we reason, and we talk about the conclusions of our reasonings

In our everyday life we often form beliefs about the existence of an event reasoning in a condition of uncertainty. Beginning with Tversky & Kanheman (1974), studies in the psychology of decision making have begun to understand the cognitive mechanisms underlying how we reason to form such beliefs under conditions of uncertainty (e.g., Tversky & Kanheman 1992; Gigerenzer et al. 2011; Preuschoff, Mohr, & Hsu 2013; Gonzalez 2016). Overall, this line of work has shown that people tend to violate laws of logic or probability in determining their confidence in the possible occurrence of an event, and rely, instead, on heuristics which simplify the likelihood estimation task. By relying on these heuristics, people rely on cognitive biases that make them adopt beliefs about the existence of an event based on a subjective probability of that event more or less independently from its objective probability.

For example, there is extensive experimental evidence suggesting the existence of an *explanation effect* such that "an explanation's quality is used as a guide to the probability of that explanation" (Lombrozo 2012: 15; see also Ross et al. 1977; Anderson, Lepper & Ross, 1980; Chapman & Chapman, 1967, 1969; Koehler 1991; Brem and Rips 2000; Lombrozo 2007). For instance, Lombrozo (2007) shows that the degree of simplicity of an explanation is used as a cue for its likelihood: when participants learned about a patient with two symptoms, they overwhelmingly considered more likely that those two symptoms were caused by a single disease (simple explanation) rather than by the conjunction of two diseases (complex explanation) in the absence of base rates for the diseases. Other properties of explanations that have been shown to increase their estimated likelihood are their breath, coherence, consistency with prior knowledge (see Lombrozo 2012 for an overview of this literature)

Human languages offer many expressions which modulate our degree of confidence about the occurrence of an event such as 'I think that…', 'I guess that …', 'probably …', 'it is certain that…', 'necessarily…', 'I know that …' 'perhaps …' etc. However, not much work has been done to connect the investigation conducted by psychologists into how humans reason under uncertainty and the investigations conducted by linguists into the meaning of expressions describing the output of those reasonings.

1.1 The general goal of this dissertation

In this project, I aim at filling this gap by showing that a proper understanding of how people reason and form beliefs under conditions of uncertainty can help in shedding light on the interpretive properties of functional words whose meaning refers explicitly to a reasoning performed by the speaker (or some other relevant individual involved in the utterance where those words are used). Specifically, in this work, I report on three projects investigating the meaning of inferential evidentials and epistemics. When used in a statement both types of expressions convey that the speaker is basing their statement on an act of inference. Therefore, it is critical to include in the investigation of the meaning of these expression considerations about how speakers conceive acts of inferences.

For example, I argue here that knowing how people form their beliefs is critical in understanding when people would say that they have concluded that an event happened, hence predicting when they would present that event in a sentence with an inferential.

Or such considerations are crucial in explaining why a statement with an inferential is typically understood as not communicating a piece of knowledge but just opinions of the

speaker: a statement based on an act of inference is bound to be interpreted in daily life as a weak statement.

Based on the same reason, I argue that a statement with a necessity epistemic – i.e., a statement presenting a conclusion as certain – is bound to be met with skepticism by listeners and bound to be understood as making a weaker claim. Interestingly, I show that this bias has an effect on the semantics of the verbal necessity epistemics such that currently verbs like 'must' are interpreted as an inferential more than as necessity operators.

Another interesting issue is whether, upon hearing someone uttering a statement with an epistemic expression reporting an assessment of the likelihood of a conclusion, people evaluate such a likelihood based on the premises of the speaker or their own. It turns out that this is not a real alternative: people seem to not agree on which perspective is the right one. This, I argue, suggests that perhaps the premises (i.e., the evidence) which a speaker bases their statement on are not included in the semantics of epistemics.

This was just a preview of some the issues that I'm going to address in this dissertation and the solutions that I provide. The dissertation is composed of three projects, which all address specific issues concerning the meaning of inferential and epistemic expressions. Before to detail what each of these three projects consists of, I offer a preliminary descriptive overview of inferential evidentials (Section 1.2) and epistemics (Section 1.3). The reader who is familiar with these types of expressions can skip them and go directly to section 1.4 where I offer a detailed overview of the content of this dissertation.

1.2 Inferential evidentials: a descriptive overview

Evidentials can be defined as those linguistic expressions indicating the '**mode of acquisition of a proposition**', namely how the speaker has obtained the proposition that she is presenting (cf. Boas 1938; Chafe and Nichols eds. 1986; Willett 1988; de Haan 1999; Aikhenvald 2004, 2018; Izvorski, 1997; Garrett, 2001; Aikhenvald, 2004, 2018; Faller, 2002, 2019; Matthewson et al., 2007; McCready & Ogata, 2007; Sauerland & Schenner, 2007; Schenner, 2010; McCready, 2010; Murray, 2010, 2017; Korotkova, 2016, 2020, 2021). Scholars conducting typological studies (see Aikhenvald 2004, 2018) have identified that languages specify three main modes of information acquisition:

- **direct** (the speaker has witnessed through one of their senses the piece of information described by the proposition)
- **reportative** (the speaker has received the information from an external source)
- **inferential** (the speaker has obtained the information as a conclusion of a reasoning)

An example of the variates of evidential morphemes that can occur in a language is offered by Tariana, an Arawak language spoken in northwestern Brazil, where evidentials are expressed as autonomous verbal suffixes. The convention in the descriptive literature is to translate evidentials as parentheticals occurring at the end of the utterance.

(1)   Aikhenvald 2003: 134-5

a. *Ceci*   *ʧinu-nuku*   *du-kwisa-ka*   (TAE)
   Cecília dog-TOP.NON.A/S 3SGF-scold-REC.P.VIS
   'Cecília scolded the dog' (I saw it: VISUAL)

b. *Ceci*   *ʧinu-nuku*   *du-kwisa-mahka*
   Cecília dog-TOP.NON.A/S 3SGF-scold-REC.P.NONVIS
   'Cecília scolded the dog' (I heard it: NONVISUAL)

c. *Ceci*   *ʧinu-nuku*   *du-kwisa-pidaka*
   Cecília dog-TOP.NON.A/S 3SGF-scold-REC.P.REP
   'Cecília scolded the dog' (I have learnt it from someone else:
   REPORTED)

d. *Ceci*   *ʧinu-nuku*   *du-kwisa-sika*
   Cecília dog-TOP.NON.A/S 3SGF-scold-REC.P.INFR
   'Cecília scolded the dog' (I inferred it: INFERRED)

e. *Ceci*   *ʧinu-nuku*   *du-kwisa-nihka*
   Cecília dog-TOP.NON.A/S 3SGF-scold-SPEC.INFR.REC.P
   'Cecília scolded the dog' (I infer it on the basis of obvious
   evidence: SPECIFIC INFERRED)

Concerning their morphosyntactic realization, it is useful to distinguish between lexical evidential and grammatical evidentials. For example, in English lexical evidentials can come in the form of attitude verbs or adverbs.

(2)   Lexical evidentials in English

   I.   *Evidential attitude verbs*

   a.   I saw that Mary was singing /I saw Mary sing

   b.   I heard that Mary was singing/I heard Mary sing

   c.   I learned/was told that Mary was singing

   d.   They say that Mary was singing

   e.   I suppose(d) that Mary was singing

   f.   I deduce(d) that Mary was singing

   g.   I conclude(d) that Mary was singing

   h.   I guess(ed) that Mary was singing

6

II.   Evidential adverbs

    i.   Apparently, Mary was singing

    j.   Allegedly/reportedly, Mary was singing

    k.   Evidently, Mary was singing

Grammatical evidentials can be of different types: dedicated verbal suffixes (e.g., Tariana), verbal suffixes where the evidential meaning is fused with other meanings like tense (e.g., Turkish) or mood (e.g., Cheyenne), enclitic suffixes attaching to any syntactic category (e.g., Cuzco Quechua), second position clitics (e.g., St'át'imcets), particles (e.g., German). Focusing on the inferential type here is a sample of morphemes that have been argued in the literature to behave like grammatical inferentials.

(3)

Italian non-temporal future   (my example)

Susanna   **sarà**   malata ora
Susan   be.**INFER**.3sg sick   now
'Susan is sick now, I infer'

German particle wohl (example from Zimmermann 2008)

Hein ist **wohl** auf See.

Hein is at sea,

'Hein is at the sea, I infer'

Cuzco Quechua enclitic suffix -cha (example from Faller 2002)

'Para-sha-n-**cha**'

rain-prog-3-cha

'It is raining, I infer'

St'át'imcets second position clitic k'a (example from Matthewson et al. 2007)

ts'aqw-an'-ás **k'a** tu7 k Lenny ti kíks-a

eat-DIR-3ERG **INFER** then DET Lenny DET cake-EXIS

'Lenny has eaten the cake, I infer'

Cheyenne verbal suffix (example from Murray 2017)

Mó-hó'taheva-he-**he** Annie

Q+3-win-negan-**inf** Annie

'Annie won, I infer'

1.3 Epistemics: a descriptive overview

In this work, I define as **sentential epistemics** 'those linguistic expressions that take as their complement a clause and whose semantic contribution is to convey an estimation of the likelihood that the proposition denoted by the complement clause is true based on some other known facts' (cf. Bybee et al. 1994; Coates 1983; van der Auwera and Plungian 1998; Palmer 2001; Nuyts 2016). In the literature, the same class of expressions is also referred to as the class of '**probability operators**' (e.g., Yalcin 2010) or '**expressions of subjective uncertainty**' (e.g., Swanson 2011, 2016). Here, I stick to the more traditional terminology of 'epistemics' but the reader should be reminded that the other two labels are equally valid for referring to the same group of expressions, namely those **expressions encoding the notion of likelihood estimation of the truth of a proposition**.

Concerning its grammatical encoding, the epistemic notion is not tied to any specific syntactic realization. Nonetheless, it is useful to distinguish between lexical epistemics and grammatical epistemics. In English, for instance, one can identify at least three different morphosyntactic types of lexical epistemics - nouns followed by a *that*-clause, adjectives in the impersonal form followed by a *that*-clause, adverbs – and one type of grammatical epistemics in the form of auxiliary verbs.

(4) Morphosyntactic types of epistemics in English

        I) *epistemic nouns (+ that-clause)*

        a. **There is certainty that** Mary is home

        b**. There is a high/low/20% probability that** Mary is home

        c. **There is a (good) possibility that** Mary is home

        d. **There is a (30%) chance that** Mary is home

        II) *epistemic adjectives (+ that-clause)*

e. **It is (absolutely/almost) certain that** Mary is home

f. **It is (highly) probable/likely that** Mary is home

g. **It is possible that** Mary is home

III) *epistemic adverbs*

h. Mary is **certainly** home

i. Mary is **necessarily** home

j. Mary is **probably** home

k. Mary is **possibly** home

l. **Perhaps/maybe** Mary is home

IV) *epistemic auxiliaries*

m. Mary **must** be home

n. Mary **might/may** be home

In other languages, one can identify several other types of grammatical epistemics like clitics, affixes, copulas, complex constructions, or complementizers (see Boye 2016 for a survey of how epistemics are encoded across languages).

It is common to include the epistemics in the class of modal expressions. The literature on the notion of modality is enormous because it has been investigated by philosophers, logicians, and linguists adopting different paradigms. This variety of perspectives has determined that the same label 'modality' has been used in quite different ways, thereby making the notion of modality an "exceptionally complex one in the sense that it encompasses a semantically diverse set of functions, which have been categorized in terms of different dimensions" (Squartini 2016: 50). To simplify, it is possible to identify in the literature at least two main senses of 'modal'. Some scholars adopt a broad sense of 'modal' which can be defined as 'any linguistic device that allows speakers to express descriptions of events that need not be real'

(e.g., Palmer 2001; Narrog 2005; Portner 2009). In this broad sense, all the following linguistic devices are presented as modals: epistemics, attitude verbs (e.g., *think, doubt, believe*, etc.), deontic and volitional verbs or adjectives (e.g., *need to, allow to, oblige to*, *may*, *can*, *want*, *hope*, *able to*, etc.), past and future tense, progressive and perfect aspect, generics (e.g., *Dogs are good friends*), habituals (e.g., *Mary goes to school by train*), individual-level predicates (e.g., *Mary is smart*), conditionals (*if... then ...*), verbal mood (e.g., indicative, subjunctive).

Some other scholars, instead, adopt a narrow sense of 'modal' which is more closely connected to what logicians and philosophers mean when they talk about modal logic: "modality has to do with necessity and possibility" (Kratzer 2012: 28). More precisely, for Kratzer modal expressions are those that involve a quantification over possible worlds. So, in this narrow sense, 'sentential modal' can be defined as 'any linguistic device corresponding to an quantifier over worlds which asserts the truth of a proposition in the set of worlds quantified over by the quantifier'. According to this definition, the adjectives *possible* and *necessary*, the adverbs *possibly* and *necessarily* represent the prototypical sentential modals but also attitude verbs and other intensional expressions. In this work, I'll use '**sentential modal**' in this narrow sense: 'any expression that takes as a complement a clause and whose core semantic contribution is to label the proposition denoted by the complement clause as true in a specified set of possible worlds'.

So, what is the relation between epistemics (i.e., expressions indicating the likelihood of an inferred proposition) and modals in the narrow sense (i.e., expressions indicating a proposition as necessary or possible)? The short answer is: epistemics are considered as a subset of modals. In this sense, estimating the likelihood of an inferred proposition is regarded as one of the ways of talking about possible or necessary propositions. What are other modals in this sense? In the literature, there is no convergence on their exact number (see Portner 2009; Nuyts

11

2016 for an overview of this debate). Perhaps the most accepted classification comprises three main types of modals: the already mentioned epistemic modality, dynamic modality, and deontic modality. I'll briefly illustrate the dynamic and deontic type. **Dynamic modality** is involved when an expression ascribes to one of the event participants an ability or a need.

(5) a. Mary **is able to** eat the entire cake

   b. Mary **needs to** eat the entire cake

Deontic modality is involved when an expression refers to permissions, obligations, and, more generally, the degree of desirability of an event.

(6) a. Mary **is allowed/permitted to** eat the entire cake

   b. Mary **is obliged to/has to** eat the entire cake

   c. Mary **should** eat the entire cake

Some scholars group together the concepts of dynamic and deontic modality under the label '**root**' (e.g., Coates 1983; Sweetser 1990; Hacquard 2010) or '**event modality**' (e.g., Palmer 1986) and argue that, based on syntactic reasons, the most relevant distinction is between the root modality and the epistemic modality. Here, I'll use the three-way categorization.

Thus, the epistemic notion of estimating the likelihood of an inferred proposition is customarily grouped together with the dynamic notion of indicating abilities/needs and the deontic notion of indicating permissions/obligations: all are conceived as being about possible and necessary propositions. The strongest support for this three-way subcategorization of the modal notion comes from the robust observation that across different languages there exists a couple of auxiliaries that are systematically ambiguous between those three types of meanings.

The English auxiliary 'must' exemplifies this pattern: the exact same utterance containing 'must' can express any of the three modal concepts.

(7) Mary must be home now

     =   Mary needs to be home now   (dynamic)

     =   Mary is obliged to be home now   (deontic)[1]

     =   It is 100% likely that Mary is home now   (epistemic)

The same pattern can be observed for the Italian auxiliaries 'dovere' and 'potere'.

(8) a.   Maria deve   essere in casa   ora

     Mary   must.3sg be   in home now

     'Mary must be home now'

     =   Mary needs to be home now   (dynamic)

     =   Mary is obliged to be home now   (deontic)

     =   It is 100% likely that Mary is home now   (epistemic)

    b. Maria può   essere in casa   ora

     Mary   may.3sg be   in home now

     'Mary may be home now'

     =   Mary is able to be home now   (dynamic)

     =   Mary is allowed to be home now   (deontic)

     =   There is a chance that Mary is home now   (epistemic)

Some other English auxiliaries are ambiguous only between two of the three modal concepts like 'may' (deontic or epistemic) or 'might' (epistemic or dynamic). But how common is this feature

---

[1] The difference between the dynamic and the deontic reading is subtle, which also explains why some scholars blur the distinction by defining one category of non-epistemic reading. In my intuition, both the dynamic and the deontic reading define an obligation for an individual, but the dynamic obligation is due to a physical need affecting the individual while the deontic obligation is due to an external principle independent from the individual.

across languages? Which languages exactly exhibit auxiliaries that are ambiguous between epistemic, deontic, and dynamic concepts? Typological studies have shown that this systematic ambiguity seems to be an areal feature "characteristic of Europe. Here, and only here, nearly all languages have highly overlapping systems, and those that do not still show at least some overlap. [...] Outside of this area, languages without any overlap are the majority" (van der Auwera & Ammann, 2013: Section 2).

Summarizing, it is robustly attested that across European languages there is a class of auxiliaries whose meaning has to do with the concepts of necessity/possibility and which are systematically ambiguous between the epistemic, the deontic, and the dynamic meaning[2]. This inherent ambiguity of auxiliaries like 'must' and 'might' has been used as the strongest piece of evidence supporting the hypothesis that the epistemic meaning (which has to do with estimating the likelihood of a conclusion), the deontic meaning (which has to do with defining obligations / permissions), the dynamic meaning (which has to do with defining abilities/needs) form a natural category, namely the category of the modal meanings (i.e., those that have to do with defining propositions as possible/necessary).

---

[2] Notice that this ambiguity is far less systematic for the adjectives 'possible' and 'necessary' and their Italian counterpart. For reasons of space, I'll not further investigate this issue here.

1.4 Overview of the dissertation

In this dissertation I'm going to explore how the notion of inference is encoded in language. More specifically, I investigate the encoding of the notion of inference in evidential expressions and epistemic expressions.

I address three related questions:

• What is an inferential? (Chapter 2)

• What is an epistemic necessity verb? (Chapters 3-4)

• From which perspective are epistemic statements truth-evaluated? (Chapter 5)

Chapters 2-4 represents a coherent unit: in chapters 3-4 I defend the thesis that epistemic necessity verbs (in one of their senses) are grammatical inferentials, which are investigated in chapter 2. Chapter 5 can be read as an independent project which addresses a different aspect of the semantics of statements containing epistemics.

1.4.1 First project: What is an inferential? (Chapter 2)

This project addresses the problem of defining what is exactly an inferential expression. As stated in section 1.1, at a descriptive level, inferentials are the sub-type of evidentials indicating that the speaker got acquainted with the described event through a reasoning, i.e., the proposition describing the event is the conclusion of an act of inference. Although there is agreement on this initial description of their meaning, more controversial is the exact semantic analysis of clauses containing an inferential, which I'll refer to as 'Infer-p' (e.g, Izvorski, 1997; Garrett, 2001;

Faller, 2002, 2019; Matthewson et al. 2007; Schenner 2008, 2010; McCready, 2010; Murray, 2010, 2014, 2017; Korotkova, 2016, 2020; Aikhenvald, 2004, 2018). I focused on the following four open issues in the literature concerning the meaning of 'INFER-p'.

(I)   Is always the speaker the individual who is indicating has having inferred 'p'?

(Section 2.1)

(II)   Does an INFER morpheme contribute an epistemic meaning in addition to the evidential meaning?   *(Section 2.2)*

(III)   Is the evidential meaning contributed to the at-issue or the not-at-issue level?

(Section 2.3)

(IV)   Can assertion of 'INFER-p' be truth-evaluated?   *(Section 2.4)*

I address these questions through the case study of Italian clauses containing the non-temporal future (shortly, 'INFER-p') based on introspective data (see example below).

(9) Susanna **sarà**   malata
    Susan   be.**INFER**.3sg   sick
    'Susan is sick, I infer'

To preview, I answer the four questions raised above as follows.

(I)   The individual who is indicated as having inferred 'p' is not the speaker but the contextually defined individual who is understood as undergoing a commitment to the truth of 'p', called here the 'judge of p'

(II)   INFER does not encode an epistemic evaluation of 'p' but the weak commitment to the truth of 'p' is due to an extralinguistic considerations pertaining to how people perceive the reliability of information obtained through an inference

16

(III)   The inferential meaning is contributed to the not-at-issue level

(IV)   Assertion of INFER-p are neither true nor false, i.e., the question of their truth cannot be raised.

The general output of my investigation is that a clause of the type INFER-p is a bidimensional object with 'p' as the at-issue proposition and 'the judge of p inferred that p' as a not-at-issue proposition. Syntactically 'INFER' is a functional head sitting in the left periphery of a clause that takes a TP as a complement.

(10)

[$_{EvidP}$ INFER [$_{TP}$ P]]

Semantically, it is an operator that takes as an input the proposition 'p' denoted by the TP and returns as an output 'p' itself amended with a not-at-issue proposition (i.e., a proposition sitting in a separate dimension where it doesn't interact with other operators in the clause) ascribing to the relevant judge of 'p' an inferential mode of acquisition of 'p'.

(11)

$[\![INFER]\!]^{M,g,c} = \lambda p_{<s,t>}$: the judge of p has inferred that p. p

"INFER-p" = "p, which its current judge has obtained through a reasoning"

1.4.2 Second project: What is an epistemic necessity verb? (Chapters 3-4)

The second project investigates the meaning of English and Italian statements containing the epistemic necessity auxiliary verb 'must' / 'dovere', a topic of a longstanding debate in the philosophical and linguistics literature. In chapter 3, I offer a detailed overview of this debate and I advance a proposal.

First I review the standard analysis of statements containing epistemic necessity auxiliary verbs (which I refer to as 'must p') adopted in logical semantics: a speaker asserting 'must p' is communicating that they know some facts which make 'p' a necessarily true proposition, which predicts that speakers asserting 'must p' are in such a strong epistemic position towards 'p' that they can also assert 'it is certain that p', 'I know that p', or simply 'p'.   (*Sections 3.1-3.2*)

However, the prediction of this hypothesis is at odds with the intuition of semanticists and with naturalistic occurrences of 'must p' which suggest that speakers overwhelmingly use 'must p' without being in such a strong epistemic position towards 'p', i.e., when they would not say that they are certain that 'p' or they know that 'p'   (*Section 3.3*).

Next, I review the debate that was triggered by this conflict between the standard logical hypothesis and the common use of 'must p', where I identified these four positions: (i) people uttering 'must p' do assert that they know some facts which make 'p' a necessarily true proposition as predicted by the standard hypothesis but tend to overuse these assertions in an exaggerated fashion, which I refer to as the 'hyperbolic logical must hypothesis'; (ii) people uttering 'must p' assert that they know some facts and hold some believes which together make 'p' a necessarily true proposition, which I refer to as the 'weak logical must hypothesis'; (iii) people uttering 'must p' assert that they know some facts which make highly probable that 'p' is

true, which I refer to as the 'probabilistic must hypothesis'; (iv) people in uttering 'must p' use 'must' as an inferential evidential, namely they assert that they know some facts which make them conclude (i.e., form the belief) that 'p' is true, which I refer to as the 'evidential must hypothesis'. (*Section 3.4*).

Lastly, I offer my hypothesis in this debate. Following a suggestion in Lyons (1977), I argue that the standard logical hypothesis is not wrong: it defines correctly the original epistemic meaning of the words 'must' and 'dovere', which is still available to some speakers. However, typically what we call epistemic use of 'must' and 'dovere' is not based on this meaning but on the meaning defined by the evidential hypothesis. So, I argue that currently epistemic 'must p' is polysemous between two senses: the logical sense (p is necessarily true given some salient facts) and the evidential sense (the speaker concludes that p given some salient facts). Furthermore, I speculate that this evidential sense of the words 'must' and 'dovere' is a derivative meaning of the words stemming from their overuse as a marker of epistemic necessity: the original meaning indicating logical conclusions (i.e., certainties) is weakened/bleached to indicate just conclusions (i.e., opinions), after reiterated and implausible exaggerated uses of the words in their logical sense. So, we think that the hyperbolic logical hypothesis is right in identifying a process of exaggeration as the key component in the epistemic 'must' puzzle; but that hypothesis is wrong in assuming that at the current stage of the language speakers exaggerate when they use the epistemic 'must'. In fact, after reiterated exaggerated uses as necessity operators, currently, the words 'must' and 'dovere' in their epistemic uses have turned into inferential markers: in using them, speakers typically are not communicating exaggerated confidence in the certainty of an event but just that they have formed the belief that the event happened based on their reasoning. (*Section 3.5*)

In Chapter 4, I offer data supporting my proposal advanced in Chapter 3.

First, I offer findings from four comprehension tasks requiring participants to decide whether they endorse a statement based on the given contextual information (Experiment 1-3 with English speakers, Experiment 4 with Italian speakers), which support the evidential hypothesis over the hyperbolic logical hypothesis and the probabilistic hypothesis for both English 'must' and Italian 'dovere': English and Italian speakers typically utter epistemic 'must p' to assert that a contextually salient reasoning made them conclude (i.e., form the belief that 'p'). These findings provide further confirmation for the claim in the psychology of decision making that people form the belief that an uncertain event happened relying more on their subjective probability of that event than on its objective probability.   (*Section 4.1*)

Next, I offer non-quantitative data which further show that epistemic necessity verbs in English and in Italian behave similarly to inferential evidentials like the non-temporal future investigated in Chapter 2: 'must p' like 'INFER-p' can co-occur with epistemic adverbs of any degree; when 'must p' is embedded under a doxastic predicate, the contribution of 'must' is not at-issue like that of 'INFER' when 'INFER-p' is embedded under a doxastic predicate; 'must p' like 'INFER-p' doesn't felicitously occur in the antecedent of conditionals.   (*Section 4.2*)

Then, I offer data based on introspection suggesting the existence of the original epistemic meaning of the words 'must' and 'dovere'. While the contribution of INFER in 'INFER-p' cannot be targeted by a truth-value judgement or by negation, the contribution of 'must' in 'must p' can.   (*Section 4.3*)

Lastly, I discuss two problematic cases: the use of 'must p' embedded under 'know' and the utterance of 'must p' in situations where 'p' is a logical conclusion from true premises. In both

cases, it is not clear if 'must' is interpreted with its original meaning of necessity or as an

inferential.   (*Section 4.4*)

1.4.3 Third project: From what perspective are epistemic statements truth-evaluated? (Chapter 5)

The issue at stake in this chapter is: if someone utters in front of us an epistemic statement

without making explicit their evidence and we are asked to assess the truth of the statement, do

we inquiry into the speaker's evidence for making the assessment? In other words, do we

evaluate the truth of an epistemic statement assuming the speaker's or our perspective? This

question has given rise to a hot debate in the formal semantic literature (e.g., MacFarlane 2011;

Egan et al. 2005; Yalcin 2011; von Fintel & Gillies 2011; Bach 2011; Yanovich 2014;

Mandelkern 2019; Rudin 2021). Central to this debate are findings on truth-value judgments of

'might p' statements in "eavesdropping" scenarios where the statement is appropriately asserted

from the point of view of the speaker but does not correspond to how in reality things are.

First, I review the debate   (*Section 5.1*)

Next, I review previous experimental findings on truth-value judgments of 'might p' and 'p'

statements in "eavesdropping" scenarios where the statement is appropriately asserted from the

point of view of the speaker but does not correspond to how in reality things are. (Knobe &

Yalcin 2014; Beddor & Egan 2018; Phillips & Mandelkern, 2021; Reuter & Brun 2021).

                                                                              (*Section 5.2*).

Lastly, I offer findings from two studies suggesting that in these scenarios English speakers disagree on the truth-value not only of 'might p' but also, surprisingly, of bare 'p'.

(*Section 5.3*).

I argue that underlying this behavior is not a disagreement on the semantic content of the statements but a disagreement on how to interpret the meaning of the adjectives 'true' and 'false' when applied to a statement uttered by another agent. Specifically, I compare two potential accounts: the first account ascribes to the adjective 'true' an ambiguity between a 'coherent' sense – a statement uttered by A is true as long as the statement coheres with A's evidence – and a 'correspondentist' sense - a statement as uttered by A is true as long as the statement correspond to reality; the second account treats 'true' as having only the correspondentist sense but ascribes to it an inherent relative nature – to decide if a statement as uttered by A is true one has to first decide a perspective (A or the assessor). Lastly, we discuss the implications of the findings for the semantics of epistemic 'might p' statements.

(*Section 5.4*)

Chapter 2.   Inferentials: the case study of the Italian non-temporal future

2.0 Introduction

The first project (Chapter 2) addresses the problem of defining what is exactly an inferential expression. As stated in section 1.1, at a descriptive level, inferentials are the sub-type of evidentials indicating that the speaker got acquainted with the described event through a reasoning, i.e., the proposition describing the event is the conclusion of an act of inference. Although there is agreement on this initial description of their meaning, more controversial is the exact semantic analysis of clauses containing an inferential, which I'll refer to as 'Infer-p' (e.g, Izvorski, 1997; Garrett, 2001; Faller, 2002, 2019; Matthewson et al. 2007; Schenner 2008, 2010; McCready, 2010; Murray, 2010, 2014, 2017; Korotkova, 2016, 2020; Aikhenvald, 2004, 2018). I focused on the following four open issues in the literature concerning the meaning of 'INFER-p'.

(I)   Is always the speaker the individual who is indicating has having inferred 'p'?

(Section 2.1)

(II)   Does an INFER morpheme contribute an epistemic meaning in addition to the evidential meaning?   (Section 2.2)

(III)   Is the evidential meaning contributed to the at-issue or the not-at-issue level?

(Section 2.3)

(IV)   Can assertion of 'INFER-p' be truth-evaluated?   (Section 2.4)

I address these questions through the case study of Italian clauses containing the non-temporal future (shortly, 'INFER-p') based on introspective data (see example below).

(12)

Susanna **sarà**   malata adesso
Susan   be.**INFER**.3sg   sick   now
'Susan is sick now, I infer'

To preview, I answer the four questions raised above as follows.

(I)   The individual who is indicated as having inferred 'p' is not the speaker but the contextually defined individual who is understood as undergoing a commitment to the truth of 'p', called here the 'judge of p'

(II)   INFER does not encode an epistemic evaluation of 'p' but the weak commitment to the truth of 'p' is due to the combination of an extralinguistic considerations pertaining to how people perceive the reliability of information obtained through an inference and the competition with bare 'p' which is the natural way to convey knowledge

(III)   The inferential meaning is contributed to the not-at-issue level

(IV)   Assertion of INFER-p are neither true nor false, i.e., the question of their truth cannot be raised.

The general output of my investigation is that a clause of the type INFER-p is a bidimensional object with 'p' as the at-issue proposition and 'the judge of p inferred that p' as a not-at-issue proposition. Syntactically 'INFER' is a functional head sitting in the left periphery of a clause that takes a TP as a complement.

(13)

    [$_{EvidP}$ INFER [$_{TP}$ P]]

Semantically, it is an operator that takes as an input the proposition 'p' denoted by the TP and returns as an output 'p' itself amended with a not-at-issue proposition (i.e., a proposition sitting in a separate dimension where it doesn't interact with other operators in the clause) ascribing to the relevant judge of 'p' an inferential mode of acquisition of 'p'. This not-at-issue proposition

can be conceived as a non-restrictive relative clause specifying the source of information of the relevant judge.

(14)

$[\![INFER]\!]^{M,g,c} = \lambda p_{<s,t>}$: the judge of p has inferred that p. p

"INFER-p" = "p, which its current judge has obtained through a reasoning"

In its unembedded uses, the meaning of 'INFER-p' can be paraphrased as 'I'm committed to the truth of p, which I obtained through a reasoning (given some facts A)'. Now, although in principle it is possible to be in a situation where a piece of information obtained through an inference counts as knowledge, in daily life there is not much room for logical inferences, therefore people are biased to consider inferred propositions as describing opinions of the agent who draws the inference. As a consequence, people are biased to interpret someone communicating 'I'm committed to believe in the truth of p which I obtained through a reasoning (given some facts A)' as communicating 'I formed the opinion that p through a reasoning (given some facts A)'. So, in daily communication 'INFER-p' is understood as 'I formed the opinion that p based on some relevant evidence'.

2.1 Who is the relevant individual whom 'INFER' defines as having inferred that 'p'?

Consider the following utterance of an Italian bare indicative matrix clause.

(15)   Susanna   è   malata

         Susan   PRES.be.3sg   sick

         'Susan is sick'

When a speaker asserts a bare indicative matrix clause like 'Susan is sick' they are projecting in the conversational space the proposition describing the situation of Susan being sick at the current time. In an influential theory of assertive acts - the so-called Table Theory – this action is described as 'presenting the proposition on the conversational table' (e.g., Farkas & Bruce 2010; Farkas & Roelofsen 2017; Faller 2019; Rudin 2022). But it has long been recognized that a speaker asserting a declarative clause is not just presenting the proposition denoted by that clause but is also telling something about the relation between them and the presented proposition, i.e., what they are doing with that proposition. According to a prominent view, what the speaker is doing is undergoing a **commitment to hold an epistemic stance** towards the presented proposition (cf. Hamblin 1973; Brandom 1983; Gulongson 2008; Farkas & Bruce 2010; MacFarlane 2011; Shapiro 2018; Geurts 2019; Krifka 2021). It is possible to identify two basic types of epistemic stances towards a proposition:

- the **factive stance** whereby an agent judges a proposition as a piece of knowledge
- the **doxastic stance** whereby an agent judges a proposition as an opinion of theirs

Intuitively, a speaker asserting a bare indicative clause is understood to be committed to hold a factive stance towards the proposition p denoted by that clause. To exemplify, a speaker

asserting a clause like (15) is understood to be committed to be in the epistemic position of also asserting 'So che Susanna e' malata' ('I know that Susan is sick') or 'E' vero che Susanna e' malata' ('It is true that Susan is sick').

So far, I have assumed that the speaker is the relevant individual undergoing an epistemic commitment towards the proposition denoted by a bare indicative clause. But languages offer perspective-shifting expressions that allow to ascribe the epistemic commitment to another individual. Consider the following cases.

(16)   Secondo Maria,   Susanna   è   malata

        According to Mary,   Susan   IND-PRES.be.3sg   sick

        'According to Mary, Susan is sick'

(17)   Maria pensa che Susanna   è   malata

        Mary thinks that Susan   IND-PRES.be.3sg   sick

        'Mary thinks that Susan is sick'

(18)   Maria è convinta che Susanna   è   malata

        Mary is convinced that Susan   IND-PRES.be.3sg   sick

        'Mary is convinced that Susan is sick'

In all cases (4-6) the individual who is understood to be committed to hold an epistemic stance (specifically, a doxastic stance) towards the proposition denoted by the clause 'Susanna è malata' is Mary and not the speaker.

Generalizing, for any given declarative clause S that is used in an utterance it is possible to identify two relevant individuals: the individual who is using S in an utterance and the individual who is committed to hold an epistemic stance towards the proposition denoted by S. Let's call the first individual the **speaker of S** and the second individual the **judge of S** (cf. Lasersohn 2005; Stephenson 2007; Charnavel 2019; Krifka 2021). Other terms adopted in the literature to refer to the individual that here I call 'judge' are 'epistemic reference point' (Zimmermann 2008), 'evaluator' (Patel-Grosz, 2012), 'perspectival center' (Harris 2021), 'origo' (Korotkova 2020). When a clause S is uttered as a matrix clause the speaker of S and the judge of S typically coincide; but when S includes a perspective-shifting expression the speaker of S and the judge of S may differ.

Now the question at stake in this section is: when a INFER-p clause is used in an utterance, who between the speaker of INFER-p and the judge of INFER-p is relevant for the interpretation of INFER? In other words, to whom is INFER ascribing an inferential mode of acquisition of p? The answer to this question can be provided by looking at the cases when an INFER-p clause is embedded under expressions that shift the judge to another individual different from the speaker.

(19)   Secondo Maria,   Susanna   sarà   malata

         According to Mary,   Susan   INFER.be.3sg sick

         'According to Mary, Susan will be sick'

(20)   Maria ha detto che Susanna   sarà   malata

         Mary has said that Susan   INFER.be.3sg sick

'Mary has said that Susan will be sick'

(21)   Maria pensa che Susanna   sarà   malata

               Mary   thinks that Susan   INFERbe.3sg sick

               'Mary thinks that Susan will be sick'

(22)   Maria è convinta che Susanna   sarà   malata

               Mary is convinced that Susan   INFER.be.3sg   sick

               'Mary is convinced that Susan will be sick'

In all these cases, the individual to whom INFER ascribe an inferential mode of acquisition of p can only be Mary (i.e., the relevant judge of p) and not the speaker. Consider this other case.

(23)   Maria ha sentito dire che Susanna   sarà   malata

               Mary has heard say that Susan   INFER.be.3sg sick

               'Mary has heard that Susan will be sick'

When a clause is embedded under a reportative verb like 'sentir dire' (to hear), the individual committed to the truth of the denoted proposition is the originator of the report and not the receiver of the report (i.e., the subject of the reportative verb). Accordingly, INFER ascribes to this unspecified individual an inferential mode of acquisition of the proposition.

So, it seems that we can formulate the following generalization: in a declarative INFER-p, INFER keeps track of p's judge in the specific context of utterance and not who the speaker is.

The inferential future can also occur in an interrogative clause, both in a yes-no question and in a wh-question.

(24)

    a. Susanna sarà   malata?

        Susan   INFER.be.3sg   sick

        'Will Susan be sick?'

    b. Dove   sarà   Susanna?

        Where INFER.be.3sg Susan

        'Where will Susan be?'

To whom is INFER referring to in these cases? In my intuition, the speaker is inquiring about whether Susan is sick (14a) or where Susan is (14b) and in both cases the addressee is invited to provide the answer based on an act of reasoning. So, when used in a question the inferential future points to the addressee, namely the individual who is invited to take an epistemic commitment towards one of the alternative propositions defined by the question operator. We can call this individual the **judge of the answer to the question**.

So, for the Italian INFER the following generalization can be made:

> *Italian INFER-p always ascribes an inferential mode of acquisition of p to the relevant judge of the semantic content denoted by p (the speaker in matrix contexts, the attitude holder in most embedding contexts, the original source of the message in reportative embedding contexts, the addressee in interrogative contexts)*

The question arises naturally: is this a general property of INFER morphemes across languages or is it a specific property of Italian INFER? Zimmermann (2011) reports the same judgements for the particle *wohl*. *Wohl* can be embedded under a verb like *glauben* ("think") and can only refer to the mode of acquisition of the attitude holder and not of the speaker.

(25)   Zimmermann (2011: 2023)

Tom glaubt,   dass es wohl   Erdbeeren gibt.

Tom thinks   that it   prt   strawberries gives

'Tom thinks that WOHL there will be strawberries.'

= Tom has acquired through an inference that there will be strawberries

≠ speaker has acquired through an inference that there will be strawberries

Wohl can occur in a yes-no question and ascribes the inferential mode of acquisition to the addressee, namely the individual who is invited to take an epistemic commitment towards the answer.

(26)   Zimmermann (2011: 2020)

Hat Hans wohl Maria eingeladen?

has Hans prt   Mary invited

'Has Hans WOHL invited Mary?'

So, both the Italian and the German INFER-p behave identically with respect to the individual whom they refer to: it is always the relevant judge of p.

What do we know about other languages? More specifically, is there a language where the INFER-morpheme either in an embedding context or in a question still refers to the speaker? Relative to this issue, in the literature on evidentials, it is claimed that there is cross-linguistic

variation as to whether an evidential morpheme in embedded and interrogative contexts

undergoes evidential shift, i.e., whether the role of evidential holder shifts from the speaker to

the subject of the attitude predicate or to the addressee (Valenzuela, 2003; Matthewson et al.,

2007; McCready & Ogata, 2007; Sauerland & Schenner, 2007; Schenner, 2010; McCready,

2010; Murray, 2017; Korotkova 2021; Frana & Menendez-Benitez 2019). Specifically,

(Korotkova 2016) classifies languages in three categories depending on their evidential shift

pattern.

- **Languages where evidential shift doesn't occur** (E.g., Georgian, Bulgarian)
- **Languages where evidential shift does occur and is obligatory (**E.g., Korean, Japanese, St'át'imcets, Tibetan, Zazaki)
- **Languages where evidential shift does occur but is optional** (E.g., German, Turkish)

However, what seems to be missing in the literature is a more systematic assessment of

evidential shift across different types of evidentials and different types of attitude predicates. In

particular, the discussion of the reportative evidentials and inferential evidentials seems to

assume that both evidentials should have a uniform shifting pattern. However, the two types of

evidentials may behave differently. For example, I predict that there is no language that allows 'x

think that p-inferential' and the individual who has concluded that p is the speaker instead of x,

because that would not make sense from a communicative point of view: if the speaker is

committed to have concluded that 'p' there is not much sense of ascribing the belief in 'p' to

someone else. So, the crucial question is: is there any data point in the literature about a use of an

inferential evidential where the judge of the at-issue proposition is not the speaker but the

inferential still refers to the speaker? After overviewing the sparse literature that investigates the

embedding of evidential (Floyd 1999, Valenzuela 2003, La Polla 2003, de Reuse 2003, Faller

2006, McCready & Ogata 2007, Matthewson et al. 2007, Sauerland & Schenner 2006, Schenner 2010, Korotkova 2016, 2019, Murray 2017) I didn't find any reported case of non-shifted inferential evidential.

2.2 Does INFER-p encode an epistemic evaluation of 'p'?

In the previous section, I have concluded that INFER-p clause indicates that the contextually defined judge of 'p' has acquired 'p' through an inference. In this section, I aim to answer the question: does INFER-p encode in its semantics an epistemic component besides the evidential component? Before to address this question, I'll make a couple of extra-linguistic considerations pertaining to the reliability of information acquired through reasoning.

Conceptually, the mode of acquisition of a proposition is independent from the epistemic stance towards that proposition. In other words, at a conceptual level, whether an individual has acquired a proposition through direct perception, a report from someone else, or an inference is independent from whether that individual holds a factive or a doxastic stance towards the proposition. However, in practice, how an individual has acquired a proposition influences whether that individual considers the proposition as a piece of knowledge or an opinion. Seeing a situation happening is usually enough for knowing that the situation happened, unless someone holds a general skepticism towards the perceptions coming from the senses. Acquiring a proposition through a report leads to a factive stance towards that proposition depending on whether the report is considered trustworthy. What about acquiring a proposition through an inference? Is an inferred proposition a piece of knowledge or an opinion of the individual who draw the inference? Philosophers and mathematicians would answer: "It depends on the nature of the inference drawn". Indeed, philosophers distinguish between two main types of inferences: logical (the conclusion is necessarily true if all the premises are true) and non-logical (the conclusion could be false even if all the premises are true). The non-logical conclusions are further divided in inductive (the conclusion is drawn from the premises based on purely

statistical considerations) and abductive (the conclusion is drawn from the premises as the best

explanation for those premises). Based on this picture, there exists one type of inference - logical

inference based on true premises – which entitles the agent who draws the inference to hold a

factive stance towards the conclusion, i.e., to claim that the conclusion is a piece of knowledge.

Thus, in principle, it is possible to be in a situation where the same individual has acquired a

proposition through an inference and holds a factive stance towards that proposition. As an

example of such a case (a proposition inferred and known) consider the following.

(27)

Premise 1:	You place a ball in one of three boxes positioned in a specific order.
Premise 2:	The order of the three boxes is changed by another individual while you cannot see.
Premise 3:	The individual doesn't use any trick.
Premise 4:	You open the first two boxes, and they are empty.
Conclusion:	The ball is in the third box

With these preliminary considerations about the relation between the notion of inference

and the notion of knowledge, let's go back to our linguistic questions. In theory, if a clause with

an inferential evidential indicates simply that its judge acquired a proposition 'p' through an

inference it should be possible to use the inferential morpheme even when the judge knows 'p'.

However, in the literature it has been observed that across different languages inferential

morphemes don't combine well with a factive stance towards a proposition.

German *wohl*. Zimmerman (2008: 3) reports that "an utterance containing *wohl* is infelicitous in contexts expressing absolute certainty, as shown in [24a]. Nor can it be embedded under a verb expressing absolute certainty, as shown in [24b].

(28)   Zimmermann (2008: 3)

    a.   #Ich weiß genau, wo Hein ist. Er ist wohl auf See.

       'I know for sure where Hein is. He is WOHL at the sea.'

    b.   #Ich weiß genau, dass Hein wohl auf See ist.

       I know for sure that Hein is WOHL at the sea

Cuzco Quechua *-cha*. Faller (2002: 173) reports that "a speaker who uses *-cha* is not committed to the truth of the embedded proposition".

St'át'imcets *k'a*. Matthewson et al. (2007: 24) reports that an utterance of a clause with the clitic k'a is "not felicitous if the speaker is sure that the embedded proposition is true".

Cheyenne *-he-he*. Murray (2017: 74) reports that "with the inferential evidential, the speaker is not committed to the truth of the at-issue proposition, though she believes it to be a good possibility".

In my intuition, the Italian future seems to behave like these morphemes described in the literature: it feels weird when used in clause whose relevant judge is assumed to have knowledge of the described event.

(29)

    a.   #So che cosa è successo a Susanna. Sarà malata

       'I know what happened to Susan. She will be sick'.

    b.   #Maria sa che Susanna sarà malata

'Mary knows that Susan will be sick'

Moreover, in my intuition and in the intuition of other Italian semanticists (cf. Mari 2010; Eckart & Beltrama 2018; Giannakidou & Mari 2018; Ippolito & Farkas 2020) INFER-p feels weird if used to report a logical conclusion based on true premises which leaves no doubt about the truth of 'p'.

(30)

Premise 1:	You place a ball in one of three boxes positioned in a specific order.
Premise 2:	The order of the three boxes is changed by another individual while you cannot see.
Premise 3:	The individual doesn't use any trick.
Premise 4:	You open the first two boxes, and they are empty.
Conclusion:	#Quindi, la palla sarà nella terza scatola.
	'So, the ball will be in the third box'

Based on these judgements, scholars agree that inferential morphemes convey, beside the evidential claim, a non-factive epistemic commitment by the judge. To account for the weak epistemic commitment expressed by inferentials Faller (2002) - followed by several other scholars - suggest that inferential evidentials encode an epistemic component beside the evidential component (e.g., Faller 2002; Matthewson et a. 2007; McCready 2010; Korotkova 2016; Murray 2017). The idea suggested by Faller (2002) is to assume that a sentence with an inferential evidential encodes the message 'the relevant judge asserts the possibility of a proposition p based on a reasoning' such that an inferential evidential "behaves like English possibility modals" (Faller 2002: 178). Thus, according to Faller, an Italian sentence with the

inferential future behaves semantically like an Italian sentence with an epistemic possibility modal ('potere').

(31)   <u>Faller (2002): 'INFER-p' = 'puo' p'</u>

    a. Susanna sarà   a casa

        Susan   INFER.be.3sg   at home

        'Susan will be home'

    b. Susanna può   essere a casa

        Susan   might be   at home

        'Susan might be home'

However, this account faces multiple challenges. First, it predicts that, like possibility epistemics, an inferential evidential should be not felicitously co-occurring with epistemic adverbs expressing a stronger epistemic stance than mere possibility. But this is not the case at least with respect to the Italian future: while an Italian possibility epistemic can felicitously combine only with a possibility adverb like 'forse' ('perhaps'), the Italian future can felicitously combine with modal adverbs expressing any degree of epistemic commitment from maximal commitment ("certainly") to minimum commitment ("perhaps").

(32)   <u>Possibility epistemic and modal adverbs</u>

    a. Forse Susanna può essere malata

        'Perhaps, Susan might be sick'

    b. #Probabilmente Susanna può essere malata

        'Probably, Susan might be sick'

    c. #Certamente Susanna può essere malata

        'Certainly, Susan might be sick'

(33)   Future and modal adverbs

        a. Forse Susanna sarà malata

          'Perhaps, Susan will be sick'

        b. Probabilmente Susanna sarà malata

          'Probably, Susan will be sick'

        c. Certamente Susanna sarà malata

          'Certainly, Susan will be sick'

Second, while one can use the epistemic possibility verb in two conjoined statements reporting non-compatible events, this is not the case for the Italian future.

(34)   a. Susanna può essere a casa, ma può anche essere in università **(no contradiction)**

        'Susan might be home, but she also might be in university'

        b. Susanna sarà in casa ma sarà anche in università'   **(contradiction)**

        'Susan will be home, but she will also be in the university'

Based on these considerations, I argue that there is no epistemic component encoded in INFER-p.

Instead, I speculate that the feeling that 'INFER-p' is not fully felicitous in situation where the relevant judge knows that 'p' is a conversational implicature triggered by the combination of two factors. First, an extralinguistic factor: in daily life we are biased to consider an inference/reasoning as a source of information whose strength is generally weak and highly dependent on the authority and the competence of the reasoner; typically, in daily life communication, most of our reasonings are far from reliable. Second, an intralinguistic factor: as soon as one believes that a proposition 'p' describes a piece of knowledge there is a communicative pressure on the speaker to utter 'p' as a bare statement; adding words or

morphemes that introduce additional pieces of meaning other than just 'p' – like the inferential meaning added by 'INFER' – invites the listeners to infer that the speaker is not in the position of presenting 'p' as piece of 'knowledge'. To put it in other words, in daily communication, a speaker who is fully convinced that 'p' describes a piece of knowledge is pushed, for economy of communication, to just say 'p' even if the source of information is inferential.

To further support the hypothesis that the incompatibility between knowledge of 'p' and INFER-p is just an implicature – i.e., it is only suggested – and is not encoded in the semantics of INFER, I show that it is possible to find naturalistic occurrence of INFER-p embedded under "know", which suggests that the incompatibility between 'know' and 'INFER-p' is not absolute.

(35)   Eve però vuole ritrovare Adam perché **sa che**, senza lei, **ora lui sarà molto triste**[3]

         'Eve wants to find Adam because **she knows that he will be very sad now** without her'

In this example, Eve is presented as knowing the proposition that Adam is very sad without her and at the same time as having acquired that proposition through an inference. One might react by questioning whether Eve really knows that Adam is sad without her, hence questioning the appropriateness of using 'know' here. Nonetheless, no matter our perspective on the choice of 'know' here, it remains undisputable that the speaker of this utterance felt comfortable using 'INFER-p' embedded under 'know that', suggesting that at least from their perspective 'know' and 'INFER-p' can go well together.

---

[3] Source: http://it.wikipedia.org/wiki/Ultimi_raggi_di_luna

2.3 What is the at-issueness status of INFER?

In this section I'm going to address the question: what is the at-issueness status of INFER in INFER-p? First, I review what at-issueness is about (section 2.3.1); next I address the syntactic aspect of at-issueness (section 2.3.2); lastly I discuss the at-issueness profile of Italian INFER (section 2.3.3).

2.3.1 What is at-issueness?

In its basic form, the formal semantics approach to the meaning of declarative clauses is based on the following assumption: the whole meaning of a declarative clause is obtained from the combination of the meanings of each of the morphemes occurring in the clause according to the syntactic arrangement of those morphemes. However, this picture has turned out to be too simplistic: it has been observed that some expressions convey a meaning that doesn't combine with the meaning of the other morphemes. More specifically, these expressions convey a proposition that feels like a separate piece of information from the proposition obtained by the combination of the other morphemes occurring in the clause. For example, consider the clauses below. Each of them encodes two propositions which are independent from each other: one proposition is clearly conveyed by the presence of a single word (in red here) and the other is conveyed by combining all the other words (in blue here).

(36)   a. Susan is **still** sick
   'Susan is sick'   'Susan was sick recently'

       b. Susan is sick **too**
   'Susan is sick'   'Someone else besides Susan is sick' or
                          'Susan is sick beside being something else'

42

c. Susan is **already** sick

'Susan is sick'   'Susan was expected to be sick'

d. Susan is sick **again**

'Susan is sick'   'Susan was sick in the recent past'

e. Susan is not sick **yet**

'Susan is not sick'   'Susan is likely to get sick in the future'

Intuitively, words like "still", "again", "yet", "too" feel special in that they convey an extra-proposition in addition to that conveyed by the combination of all the other words. The current consensus is to refer to the proposition conveyed by the combination of the other words as the "**at-issue proposition**" and to the proposition conveyed by those special words as the "**not-at-issue proposition**" (e.g., Potts 2005; Simons et al. 2010; Tonhauser 2012; Tonhauser et al. 2013; Murray 2017; Koev 2018; Murray & Starr 2021). The intuition underlying the terminology "at-issue and not-at-issue" is that the two propositions have a different status: the at-issue proposition represents the part of the utterance message that is meant to address the question under discussion in the discourse where the utterance is made while the not-at-issue proposition represents a sort of complementary message not relevant for the question under discussion.

Traditionally, not-at-issue propositions are further divided in two categories: presuppositions and conventional implicatures. However, this distinction has never been completely clarified and there has been continued uncertainty about how to apply the label in specific cases. Here, I'm not interested in what differentiates presuppositions from conventional implicature but in what they have in common. For my purpose here, the crucial point is that not-

at-issue propositions are special in that they project another dimension of meaning parallel to the at-issue dimension. So, to simplify, I'll adopt the following definitions throughout the work.

At-issue meaning	any meaning that is able to interact with other operators and compose regularly with the meaning of the other morphemes in the clause, thereby preserving semantic one-dimensionality
Not-at-issue meaning	any meaning that doesn't compose regularly with the meaning of the other morphemes in the clause and defines a second semantic dimension

Table 1. Definitions of at-issue and not-at-issue meaning

Whether an expression triggers not-at-issue meaning is diagnosed in several ways. I'll divide the diagnostics in two groups: the **projection diagnostics** and the **anaphora diagnostics**.

The projection diagnostics

The projection diagnostics are used to test whether a given proposition semantically scopes under other linguistic operators like modals, questions, conditionals, embedding predicates, and negation. All these tests work as follows: if a proposition is not semantically under (i.e., projects out of) the scope of any of these operators, it is a not-at-issue proposition.

Not-at-issue proposition escape the scope of modal expressions like 'perhaps' or 'may'. When, for example, 'perhaps' is added to any of the declarative clauses above only the at-issue proposition falls under the scope of the adverb.

(37)   <u>Interaction with modals</u>

   a. Perhaps Susan is **still** sick

   = Perhaps Susan is sick now + it is true that Susan has been sick for some time up
                                    to now

   ≠ Perhaps Susan is sick now + Perhaps Susan has been sick for some time up to
                                    now

   b. Perhaps Susan is sick **too**

   = Perhaps Susan is sick now + it is true that someone else besides Susan is sick

   ≠ Perhaps Susan is sick now + perhaps someone else besides Susan is sick

   c. Perhaps Susan is **already** sick

   = Perhaps Susan is sick now + it is true that Susan was expected to be sick

   ≠ Perhaps Susan is sick now + Perhaps Susan was expected to be sick

   d. Perhaps Susan is sick **again**

   = Perhaps Susan is sick now + it is true that Susan was sick in the past

   ≠ Perhaps Susan is sick now + Perhaps Susan was sick in the past

   e. Perhaps Susan is not sick **yet**

   = Perhaps Susan is not sick now + It is true that Susan is likely to get sick

   ≠ Perhaps Susan is not sick now + Perhaps Susan is likely to get sick

Not-at-issue propositions escape the scope of question operators. Indeed, when any of the declarative clauses above is turned into a yes-no question the speaker remains committed to the truth of the not-at-issue proposition which is, thus, not part of the information that the speaker is inquiring about.

(38)   Interaction with questions

   a. Is Susan **still** sick?
   = Is Susan sick now? + it is true that Susan has been sick for some time up to now
   ≠ Is Susan sick now and has she been sick for some time up to now?
   b. Is Susan sick **too?**
   = Is Susan sick now? + it is true that someone else besides Susan is sick
   ≠ Is Susan sick now and is someone else besides Susan sick now?
   c. Is Susan **already** sick?
   = Is Susan sick now? + it is true that Susan is expected to be sick
   ≠ Is Susan sick now and is Susan expected to be sick?
   d. Is Susan sick **again?**
   = Is Susan sick now? + it is true that Susan was sick in the past
   ≠ Is Susan sick now and was Susan sick in the past?
   e. Is Susan not sick **yet?**
   = Is Susan sick now? + it is true that Susan is likely to get sick in the future
   ≠ Is Susan not sick now and is Susan likely to get sick in the future?

Not-at-issue propositions escape the scope of conditional operators. When a declarative clause is inserted as the antecedent of an if-clause, the speaker is not committed to the truth of the at-issue proposition but remains committed to the truth of the not-at-issue proposition.

(39)   Interaction with if-clauses

   a. If Susan is **still** sick, then …

= If Susan is sick now, then … + it is true that Susan has been sick for some time

up to now

≠ If Susan is sick now and if Susan has been sick for some time up to now,

then …

b. If Susan is sick **too**, then …

= If Susan is sick, then… + it is true that someone else besides Susan is sick

≠If Susan is sick now and if someone else beside Susan is sick, then …

c. If Susan is **already** sick, then …

= If Susan is sick, then … + it is true that Susan is expected to be sick

≠ If Susan is sick now and if she is expected to be sick, then …

d. If Susan is sick **again**, then…?

= If Susan is sick, then … + it is true that Susan was sick in the past

≠ If Susan is sick now and if she was sick in the past, then …

e. If Susan not sick **yet**, then …

= If Susan is not sick, then … + it is true that Susan is likely to get sick

≠ If Susan is not sick now and if she is likely to get sick, then …

When a clause S denoting an at-issue proposition 'p' and a not-at-issue proposition 'q' is embedded under a doxastic predicate like in 'x thinks that S', x but not the speaker is understood as committed to the truth of 'p', whereas the speaker is understood as to be committed to the truth of 'q' (x may or may not believe that 'q' is true).

(40)   Interaction with embedding predicates

a. John thinks that Susan is **still** sick

= John thinks that Susan is sick now + the speaker is committed to the truth of
'Susan has been sick for some time up to now'.

≠ John thinks that Susan is sick now and that she has been sick for some time up to now

b. John thinks that Susan is sick **too**

= John thinks that Susan is sick now + the speaker is committed to the truth of 'someone else besides Susan is sick now'.

≠ John thinks that Susan is sick now and that someone else besides Susan is sick now

c. John thinks that Susan is **already** sick

= John thinks that Susan is sick now + the speaker is committed to the truth of 'Susan was expected to be sick'

≠ John thinks that Susan is sick now and that she was expected to be sick

d. John thinks that Susan is sick **again**

= John thinks that Susan is sick now + the speaker is committed to the truth of 'Susan was sick in the past'

≠ John thinks that Susan is sick now and that she was sick in the past

e. John thinks that Susan is not sick **yet**

= John thinks that Susan is not sick now + the speaker is committed to the truth of 'Susan is likely to get sick in the future'

≠ John thinks that Susan is not sick now and that she is likely to get sick'

Not-at-issue propositions escape the scope of sentential negations ("not"). I left this diagnostic for last because adding negation to a sentence could not be always straightforward. For example, some of the words considered here cannot co-occur with negation and are substituted by other words: e.g., "Susan is still sick" > "Susan is not sick anymore"; "Susan is sick too" > "Susan is not sick either". Let's apply this diagnostic only to "already" and "again":

while the at-issue proposition undergoes the effect of negation (i.e., it is under the semantic

scope of the negation operator) the not-at-issue proposition is not negated.

(41)   <u>Interaction with negation</u>

    a. "Susan is not already sick"

       Not (Susan sick) + *Susan is expected to be sick*

    b. "Susan is not sick again"

       Not (Susan sick) + *Susan was sick in the past*

Anaphora diagnostics ("direct challenge test" or "assent/dissent test")

This diagnostic is based on the assumption that not-at-issue propositions differ from at-

issue proposition in the way they are available for a following propositional anaphora: only the

at-issue proposition can be the referent, for example, of 'that' in a follow-up like 'That's right'.

(42)   <u>Availability for propositional anaphoras</u>

a. Susan is **still** sick

'Susan is sick'   'Susan was sick recently'

X

That's right

b. Susan is sick **too**

'Susan is sick'   'Someone else besides Susan is sick' or

'Susan is sick beside being something else'

X

That's right

c. Susan is **already** sick

'Susan is sick' ➤ 'Susan was expected to be sick'

X

That's right

d. Susan is sick **again**

'Susan is sick' 'Susan was sick in the recent past'

X

That's right

e. Susan is not sick **yet**

'Susan is not sick' 'Susan is likely to get sick in the future'

X

That's right

2.3.2 Syntax and compositional semantics of clauses with not-at-issue meaning

A crucial question is: **is the semantic independence of expressions triggering not-at-issue propositions mirrored in a syntactic independence as well**? To put it more explicitly, do words like "again" sit in their regular synctactic position or they project a second structural dimension mirroring the semantic bi-dimensionality? The standard approach in formal semantics is to maintain the syntactic one-dimensionality and introduce the bi-dimensionality only at the semantic level. Consider for example the word "again". Without entering in the details of the analysis of this word, we can assume that a natural syntactic position for an adverb like "again" is that of VP-modifier.

(43) [TP PRES [VP **again** [VP be Susan sick]]]

So, from a structural point of view, the distinction between at-issue and not-at-issue meaning doesn't play a role. The distinction is activated in the semantic component. The at-issue meaning of an expression is projected in the output of the compositional operation with the sister phrase, whereas the not-at-issue meaning doesn't appear in the output of the compositional operation but is, instead, treated as **an independent pre-condition on the well-formedness of the compositional operation**. To exemplify, for 'again' to successfully compose with its VP sister it has to be the case that the event denoted by the VP already occurred sometime in the past: if that condition is satisfied the event denoted by the VP (the meaning of the VP) is projected in the output of the compositional operation; if the condition is not satisfied the compositional operation clashes. The common practice based on Heim & Kratzer (1998) is to represent this not-at-issue condition in the lexical entry of the expression triggering the not-at-issue meaning as a clause restricting the domain of application of the function.

(44) $[\![\text{Again}]\!]^{M.g.c} = \lambda P_{<v,t>}$: $\exists\, t' < t$, at least an instance of the type of event described by P occurred at t'. P

Blue: Input type
Red: condition on the input not projected in the output
Green: output type

So, the idea is that 'again' can combine with the event denoted by the VP only if the condition that the event already happened in the past is satisfied; otherwise, the interpretation process clashes. Now, when the tense operator combines with [VP again [VP be Susan sick], the meaning of 'again' is not visible because it lies on another dimension.

(45)　　　　　$(\llbracket [_{TP} \text{ PRES}[_{VP} \text{ again } [_{VP} \text{ be Susan sick}]]] \rrbracket^{M.g.c} =$

$[\lambda e: \quad \exists t' < t, \text{ sick}(e) \text{ and agent}(e) = \text{Susan and time}(e) = t'.$

$\text{sick}(e) \text{ and agent}(e) = \text{Susan and time}(e) = t$

Here I will assume this standard view. So, expressions triggering not-at-issue meaning:

- have a regular syntax

- the not-at-issue part of their meaning doesn't project in the output of the compositional operation but on a parallel secondary semantic dimension

2.3.3 What is the at-issueness status of INFER?

Consider the following Italian clause containing the inferential future INFER-p.

(46)　　　Susanna **sarà**　　　　malata

　　　　　Susan　**INFER**.be.3sg　　　sick

　　　　　'Susan will be sick'

The first question I'm going to address is: what is the syntactic position of INFER in a clause like (38)? According to a line of thought initiated in Cinque (1999), clauses exhibit several functional projections above the VP level (e.g., Cinque 1999, 2006; Speas & Tenny 2003; Speas 2004, 2008; Wurmbrand 2004; Krifka 2021). Cinque (1999) proposes - based on the relative order of functional morphemes and adverbs – that for each functional morpheme that can occur in a clause there is a dedicated projection in the left periphery of the clause above the VP projection; moreover, he suggests that these heads are arranged in a rigid hierarchical order for all languages. One of these functional projections is dedicated to host evidential morphemes. Below is a subset of these functional projections (the evidential is highlighted).

(47)   Subset of functional projections (Cinque 2006: 12)

MoodP$_{speech\ act}$ > MoodP$_{evaluative}$ > MoodP$_{evidential}$ > ModP$_{epistemic}$ > TP(Past) > TP(Future) > MoodP$_{irrealis}$ > ModP$_{alethic}$ > AspP$_{habitual}$ > AspP$_{repetitive(I)}$ > AspP$_{frequentative(I)}$ > ModP$_{volitional}$ AspP$_{celerative(I)}$ > TP(Anterior) > AspP$_{terminative}$ > AspP$_{continuative}$ > AspP$_{retrospective}$ AspP$_{proximative}$ > AspP$_{durative}$ > AspP$_{generic/progressive}$ > AspP$_{prospective}$ > ModP$_{obligation}$ ModP$_{permission/ability}$ > AspP$_{Completive}$ > VoiceP > AspP$_{celerative(II)}$ > AspP$_{repetitive(II)}$ > AspP$_{frequentative(II)}$

Building on this tradition, Speas (2004) and Krifka (2021) provide further detailed analyses of the evidential functional head. Krifka (2021) conflates the evidential phrase and the epistemic phrase in one single phrase that he calls 'judgement phrase' (JP) "whose basic role is to make the judge parameter j available for […] the epistemic and evidential modifications of the proposition that the speaker is committed to" (Krifka 2021: 10).

Beside the details of the analysis, from a syntactic point of view, there is agreement that evidentials morphemes are operators that sit in a dedicated functional projection above the TP level. Assume the following simplified syntactic analysis for the clause.

(49)   [$_{EvidP/JudgP}$ INFER [$_{TP}$ PRES [$_{VP}$ be Susan sick]]]]

The next question is: is the evidential claim encoded in INFER at-issue or not-at-issue? Two hypotheses can be formulated. The first hypothesis states that the evidential claim projects in the output of the semantic composition of INFER with the proposition denoted by the sister TP, yielding a regular one-dimensional at-issue proposition: 'the relevant judge concludes that p'. The second hypothesis states that the evidential claim projects on a secondary independent

semantic dimension, yielding a two-dimension semantic content consisting of the proposition p denoted by the TP and the evidential proposition 'the relevant judge has concluded that p'.

(50)

    a. <u>One-dimension hypothesis (INFER is at-issue)</u>

$\llbracket \text{INFER} \rrbracket^{M,g,c} = \lambda p_{<s,t>}.$

Judge of p has inferred that p

e.g.,

$\llbracket [_{\text{EvidP/JudgP}} \text{ INFER } [_{\text{TP}} \text{ PRES } [_{\text{VP}} \text{ be Susan sick}]]]] \rrbracket^{M,g,c} =$

At- issue proposition:

'The relevant judge has acquired through an inference that Susan is sick at t'

    b. <u>Two-dimensions hypothesis (INFER is not-at-issue)</u>

$\llbracket \text{INFER} \rrbracket^{M,g,c} = \lambda p_{<s,t>}:$ judge of p has inferred that p.

p

e.g.,

$\llbracket [_{\text{EvidP/JudgP}} \text{ INFER } [_{\text{TP}} \text{ PRES } [_{\text{VP}} \text{ be Susan sick}]]]] \rrbracket^{M,g,c} =$

At-issue proposition:

'Susan is sick at t'

Not-at-issue proposition:

'The relevant judge has acquired through an inference that Susan is sick at t'

The near-consensus in the literature on evidentials is that evidential propositions are not-at-issue and thus should be analyzed along the lines of the two-dimensions hypothesis (see a.o., Izvorski, 1997; Garreth 2001; Faller, 2002, 2019; Matthewson et al., 2007; Sauerland & Schenner 2007; Schenner 2008, 2010; McCready 2010; Murray 2010, 2017; Matthewson 2015; Pancheva & Rudin 2019; Murray & Starr 2021). However, recently, Korotkova (2016, 2020) suggests that evidential should be analyzed as an at-issue expression, along the lines of the one-dimension hypothesis. A similar debate has been conducted in the specific literature about Italian INFER. For example, Ippolito & Farkas (2020) argue that the Italian INFER is an at-issue sentential operator whereas Frana & Menendez-Benitez (2019) argue that the Italian INFER is a not-at-issue operator. Here, I offer arguments supporting the two-dimension hypothesis that Italian INFER is a not-at-issue operator.

Matrix uses of INFER-p

Notice that the matrix uses of INFER-p don't discriminate between the one-dimension hypothesis and the two-dimension hypothesis. To see why, first consider an utterance of a bare indicative present tense matrix clause.

(51)   Susanna   **è**   malata

          Susan   **PRES**.be.3sg sick

          'Susan is sick'

Here, the judge of the proposition coincides with the speaker. I argue that the epistemic stance of the relevant judge is factive but is not encoded linguistically: whenever a speaker presents a

piece of information (i.e., a proposition) through a bare declarative clause, it is expected from a cooperative speaker that such information is a piece of knowledge and not an opinion of the speaker themselves. But, I suggest, this is not forced by the linguistic form of a bare declarative: in fact a speaker uttering a declarative clause like in (51) can follow up this utterance with a parenthetic 'I believe' which substitute the default factive stance with a doxastic stance.

(52)   Susanna   **è**   malata, credo

        Susan   **PRES**.be.3sg sick,   believe.1sg

        'Susan is sick, I believe'

Now, what happens when INFER is added to a matrix declarative clause?

(53)   Susanna **sarà**   malata

        Susan   **INFER**.be.3sg   sick

        'Susan will be sick'

Both the one-dimension hypothesis and the two-dimension hypothesis predict the same final message communicated through an utterance of a matrix INFER-p: the speaker (in the role of relevant judge) is committed to the truth of 'I conclude that p'. Where the two hypotheses differ is in the internal compositional mechanism through which that final message is achieved. According to the one-dimension hypothesis, the speaker is expressing a commitment to the truth of the proposition 'I conclude that p' which implies that the speaker is committed to the truth of 'p'. According to the two-dimension hypothesis, a speaker uttering INFER-p is committed to the truth of two propositions: the at-issue proposition 'p' denoted by the TP and the not-at-issue proposition 'I have concluded that p' contributed by INFER. This can be paraphrased as 'I'm committed to the truth of 'p', which I have obtained as the conclusion of a reasoning'

To discriminate between these two views, one needs to apply the diagnostics of not-at-issueness described in section 2.4.1. However, when applied to the Italian future not all the diagnostics of (not-)at-issueness are equally applicable or equally valid as test of at-issueness. In the rest of this section, I discuss the application of the projection diagnostics and the anaphora diagnostics to the inferential future. To preview, I argue that only the diagnostic based on the embedding under doxastic predicate and the diagnostic based on the embedding under negation are valid tests of at-issueness of INFER and the output of their application supports the two-dimension view over the one-dimension view.

The Inferential future and the projection diagnostics

       Inferential future and negation

In the literature on evidentials it is often claimed that one strong argument for the two-dimension hypothesis derives from the robust cross-linguistic observation that evidential propositions never semantically scope under negation, therefore they are not-at-issue (see Murray 2017 for a review of this diagnostic). The inferential future is no different: when a clause with an inferential future also contains a sentential negation the sentence can never be interpreted as 'the speaker has not concluded that p' but only as 'the speaker has concluded that Susan is not sick'.

(54)   Susanna **non sarà** malata

              Susan   **not INFER**.be.3sg sick

              'Susan will not be sick'

= The speaker has acquired through an inference that Susan is not sick

(INFER>not)

≠ The speaker has not acquired through an inference that Susan is sick

(not>INFER)

However, as pointed out by Korotkova (2020), the unavailability of the reading not>INFER doesn't necessarily support the two-dimension hypothesis over the one-dimension hypothesis. In fact, the one-dimension hypothesis is compatible with not>INFER if one makes the additional assumption that the syntactic position of evidential operators is rigidly higher than the position of negation. In this case, the lack of the reading not> INFER would be just telling us that negation occupies a syntactic position lower than that occupied by the inferential future.

(55) $[[_{EvidP} \text{INFER} [_{NegP} \text{NEG} [_{TP} \text{PRES} [_{vP} \text{be Susan sick}]]]]]$

So, the interaction between the inferential future and negation doesn't necessarily offer a decisive test of the at-issueness status of the inferential future; it may be just offering an indication of the relative structural arrangement of the inferential operator and the negation operator.

Inferential future and modal adverbs

The same consideration applies to the diagnostic that relies on the interaction between modal adverbs and inferential future. Let's start from the observation that the inferential future can felicitously occur with a modal adverb expressing any degree of epistemic commitment. Like in the case of negation, the evidential proposition is never semantically under the scope of the modal adverb.

(56)

    a. Certamente   Susanna sarà   malata

       Certainly   Susan   INFER.be.3pl.   sick

       'Certainly, Susan will be sick'

       = The speaker has acquired through an inference that certainly Susan is sick

       (INFER > adv)

       ≠ Certainly, the speaker has acquired through an inference that Susan is sick

       (adv > INFER)

    b. Probabilmente Susanna sarà   malata

       Probably   Susan   INFER.be.3pl.   sick

       'Probably, Susan will be sick'

       = The speaker has acquired through an inference that probably Susan is sick

       (INFER > adv)

       ≠ Probably, the speaker has acquired through an inference that Susan is sick

       (adv > INFER)

    c. Forse   Susanna sarà   malata

       Perhaps   Susan   INFER.be.3pl.   sick

       'Perhaps, Susan will be sick'

       = The speaker has acquired through an inference that perhaps Susan is sick

       (INFER > adv)

       ≠ Perhaps, the speaker has acquired through an inference that Susan is sick

       (adv > INFER)

Again, the unavailability of the reading where the inferential future scopes under the modal adverb doesn't offer a decisive reason in favor of the two-dimension hypothesis: it could be an indication of the relative arrangement of the inferential future and the modal adverbs in the syntactic structure of the clause.

Inferential future and conditional clauses

The diagnostic involving the conditional clauses cannot be applied to Italian INFER-p clauses because those clauses cannot felicitously occur in the antecedent of a conditional clause.

(57)   #Se Susanna sarà   malata , allora la   andiamo a visitare

         If   Susan   INFER.be3sg sick   then   her we-go   to visit

         'If Susan will be sick, then we'll go to visit her'

Why? I suggest that this unacceptability has to do with the fact that the antecedent of a if-clause does not have a judge. When a conditional clause is produced neither the speaker nor any other contextually defined agent is committed to have an epistemic stance towards the proposition expressed by the antecedent clause. When then the inferential future points to the judge of the antecedent clause, there is no such judge, hence yielding a semantic clash. I speculate that the inferential future is not acceptable in the antecedent of a conditional for this very reason: there is no judge which an inferential mode of acquisition of the at-issue proposition can be ascribed to.

Recall the logic of the diagnostic of at-issueness based on embedding a clause under a doxastic predicate: when a clause S denoting an at-issue proposition 'p' and a not-at-issue proposition 'q' is embedded under a doxastic predicate like in 'x thinks that S', x but not the speaker is understood as committed to the truth of 'p', whereas the speaker is understood to be committed to the truth of 'q' (x may or may not believe that 'q' is true).

(58)   John thinks that Susan is sick **again**

         = John thinks that Susan is sick now + the speaker is committed to the truth of 'Susan was sick in the past'

         ≠ John thinks that Susan is sick now and that she was sick in the past

Let's apply this diagnostic to the case of a clause with the inferential future. First of all, I observe that the inferential future can be syntactically embedded under a doxastic verb like "pensare" (think).

(59)   Maria pensa che Susanna   sarà   malata

         Mary   thinks that Susan   INFER.be.3sg sick

         'Mary thinks that Susan will be sick'

But what is the interpretation of a sentence like (59)? In this case, the one-dimension view and the two-dimension view make different predictions. The one-dimension view predicts that the meaning of (59) should be paraphrased as 'Mary thinks that Mary has concluded that Susan is sick' with the subject of the attitude verb but not the speaker committed to the truth of the

evidential proposition. The two-dimension view predicts that that the meaning of (59) should be paraphrased as a two-dimensional object: "Mary thinks that Susan is sick" (at-issue) plus the not-at-at-issue evidential proposition 'Mary has obtained the proposition that Susan is sick through an act of reasoning', with the speaker committed to the truth of the evidential proposition (i.e., it is the speaker who ascribes to Mary an inferential mode of acquisition of 'p').

(60)   One-dimension view: predicted meaning for (59)

         Mary thinks that Mary has acquired through an inference that Susan is sick

         (*the evidential information is semantically embedded under 'think' with the
         speaker not committed to the truth of that information*)

         Two-dimension view: predicted meaning for (59)

         Mary thinks that Susan is sick, which Mary has acquired through an inference

         (*the evidential information is not semantically embedded under 'think' with the
         speaker committed to the truth of that information*)

In my intuition, the interpretation of (59) is as predicted by the two-dimension view: the proposition towards which Mary has a mental attitude is 'Susan is sick' and not 'Mary has acquired through an inference that Susan is sick'. More specifically, the interpretation goes as follows: the speaker states that Mary thinks that Susan is sick now (at-issue message) and the speaker ascribes to Mary an inferential mode of acquisition of the proposition that Susan is sick now (not-at-issue message).

(61)   Maria   pensa che Susanna   sarà   malata

         Mary   thinks that Susan   INFER.be.3sg   sick

62

'Mary thinks that Susan will be sick'

Interpretation:

=   "Mary is of the opinion that Susan is sick, which she has inferred"

≠   'Mary is of the opinion that she has inferred that Susan is sick'

To help to see more clearly the interpretation of 'Mary thinks that Susan will be sick' it is useful

to compare it to 'Mary thinks that Susan is sick'. What is the semantic difference between the

two sentences? With the inferential future the speaker is explicitly presenting Mary as having in

her mind the opinion that Susan is sick and as having obtained such an opinion through an act of

reasoning, whereas with the simple indicative clause the speaker is just presenting Mary as

having in her mind the opinion that Susan is sick without specifying how Mary obtained such

opinion.

(62)

     a.   Maria   pensa che Susanna   **è**   malata

     Mary   thinks that Susan   **PRES.be.3sg**   sick

     'Mary thinks that Susan will be sick'

     Interpretation:

     =   'Mary is of the opinion that Susan is sick'

The inferential future behaves in the same way when it is embedded under other doxastic

predicates like 'believe', 'being convinced that', 'guess', 'suppose', etc.

(63)   Maria è convinta che Susanna   sarà   malata

Mary is convinced that Susan   INFER.be.3sg   sick

'Mary is convinced that Susan will be sick'

Interpretation:

=   "Mary is convinced that Susan is sick, which she has concluded"

≠   Mary is convinced that she has inferred that Susan is sick

The inferential future can also be syntactically embedded under impersonal construction like 'it is probable that' or 'it is evident that'. In this case, there is no new potential judge introduced by the construction, so the relevant judge of the embedded proposition 'Susan is sick' remains the speaker. The one-dimension view predicts that the interpretation of the utterance should be 'it is probable/evident that the speaker has acquired through a reasoning that Susan is sick' whereas the two-dimension view predicts that the utterance is interpreted as conveying separately the at-issue proposition 'it is probable/evident that Susan is sick' and the not-at-issue proposition 'the speaker has acquired through a reasoning that Susan is sick'.

(64)   È probabile/evidente che   Susanna   sarà   malata

Is probable/evidente   that   Susan   INFER.be.3sg sick

'It is probable/evident that Susan will be sick'

Interpretation:

= It is probable/evident that Susan is sick, which the speaker has inferred

≠ It is probable/evident that the speaker has concluded that Susan is sick

64

So, the diagnostic based on the behavior under embedding predicates supports the two-dimension hypothesis over the one-dimension hypothesis: whenever a clause with the inferential future is syntactically embeddable under a predicate, the inferential proposition is never under the semantic scope of the embedding predicate.

In the literature there is no discussion of the cases when INFER-p is embedded under a negated impersonal construction like 'it is not probable that'. I expect this combination to not occur because then the speaker would be committed to judge as not probable that 'p' and at the same time to have concluded that 'p' yielding two incompatible attitudes.

(65)   #Non e' probabile/evidente che   Susanna   sarà   malata

         Not is probable/evidente   that   Susan   INFER.be.3sg sick

         'It is not probable/evident that Susan will be sick'

         Interpretation:

         = ?!It is not probable/evident that Susan is sick, which the speaker has concluded

'X does not think that INFER-p' is acceptable only under the NEG-raising reading of the utterance where 'not' is interpreted in the embedded clause.

(66)   Maria   non pensa che Susanna   sarà   malata

         Mary   not   thinks that Susan   INFER.be.3sg sick

         'Mary thinks that Susan will not be sick'

         Interpretation:

         =   Mary thinks that **Susan is not sick**, which Mary has inferred

Inferential future and questions

The inferential future can syntactically occur in a question, both in a yes-no question and in a wh-question.

(67)

    a. Susanna sarà   malata?

        Susan   INFER.be.3sg   sick

        'Will Susan be sick?'

    b. Dove   sarà   Susanna?

        Where INFER.be.3sg Susan

        'Where will Susan be?'

But what is the interpretation of these questions? The one-dimension hypothesis predicts that the interpretation goes as follows: the speaker is asking the addressee if they are in the position of making an epistemic commitment to infer one of the alternative propositions defined by the question operator. The two-dimension hypothesis predicts this interpretation: the speaker is asking the addressee if they are in the position of making an epistemic commitment to one of the alternative propositions defined by the question operator and assumes that the addressee can make such a commitment only based on an inference. In my intuition, the interpretation of the questions goes as predicted by the two-dimension view (the same observation has been made for the German particle wohl; see Zimmermann 2008, 2011). So, I conclude that also the question diagnostic supports the two-dimension hypothesis over the one-dimension hypothesis: the semantic contribution of the Italian future takes scope over the alternative answers invoked by the interrogative operator.

Output of the projection diagnostics: summary

After having applied the projection diagnostics to INFER-p, I reach the following conclusion: **the evidential proposition encoded in the Italian inferential INFER is not-at-issue**. In other words, in a INFER-p clause, the proposition introduced by INFER that the relevant judge has obtained p through an inference does not interact regularly with the meaning of the other operators in the clause and is projected at another independent level of meaning. Therefore, I take these observations as supporting the two-dimensions hypothesis over the one-dimension hypothesis.

The inferential future and the anaphora diagnostic

Here, I challenge the validity of the anaphora diagnostics as a test of at-issueness. But first I'll summarize the working mechanism of the anaphora diagnostics as presented in the literature. Consider an utterance of a clause containing a trigger of not-at-issue meaning like 'Susan got married again' which projects two propositions in the discourse space - the at-issue proposition (*Susan is married now*) and the not-at-issue proposition (*Susan was married in the past*). Now, the standard story in the literature goes by saying that only the at-issue proposition is available to be picked up as the referent of a propositional anaphora in a follow-up reply.

(68) Susan is married **again**
      'Susan is married now'   'Susan was married in the past'

                        X
      That's right
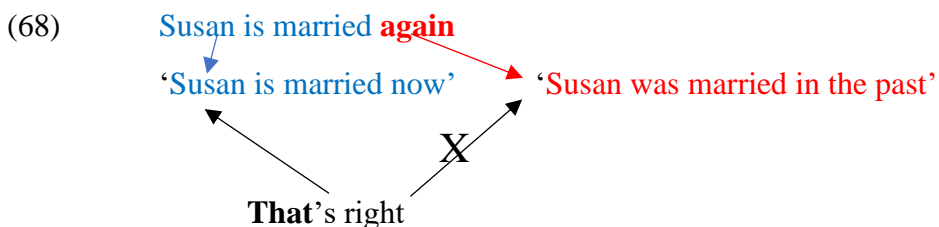
Is this view right? In my opinion it is not, and I will show why. Before I do that, I need to make some preliminary terminological remarks: a distinction needs to be made between the 'referent of an anaphora' and the 'antecedent of an anaphora'. The **referent of an anaphora** is the semantic object which the anaphora refers to and the **antecedent of an anaphora** is the syntactic object whose referent defines the referent of the anaphora. In the case of a propositional anaphora the referent is a proposition, and the antecedent is a clausal phrase produced in the proximal discourse context that refers to that proposition.

Assuming this terminological distinction, let's see the working mechanism of a propositional anaphora like 'that'. A propositional anaphora points to a clausal phrase (the antecedent) and inherits as its own semantic content the semantic content of that phrase. To exemplify, consider the example of a negative sentence. Krifka (2013) shows that the anaphora 'it' is used in follow-up to a negated sentence, it can pick-up both the whole negated clause or the positive clause embedded under negation.

(69)   Example from Krifka 2013: 5

         Ede didn't steal the cookie,

         i.   and he actually can prove **it** (= **that he didn't steal the cookie**)

         ii. even though people believed **it** (= **that he stole that cookie**)

The positive antecedent of the anaphora has to be retrieved by inspecting the syntactic structure of the sentence beyond the surface level. There one can see two phrases that are responsible for introducing the two potential referents of the following anaphora: a TP introducing a discourse referent for the positive proposition and a NegP introducing a discourse referent for the negative proposition.

(70)   Krifka (2013)'s analysis of propositional anaphoras with negated sentence

Ede didn't steal the cookie.

[_{NegP} Ede did-n't [_{TP} t_{Ede} t_{did} steal the cookie]]

    i.   and he actually can prove **it (antecedent = NegP)**

    ii.   even though people believed **it (antecedent =TP)**

With this discussion in the background, let's go back to the case of 'Susan is married again'. When one replies to an utterance of this sentence with 'That's right' what are the potential antecedents of 'that'? By inspecting the syntactic structure of the sentence, it seems that there is only one potential clausal antecedent, namely the whole TP which include the adverbial modifier 'again'.

(71)   John: "Susan is married again"

        [_{TP} PRES [_{VP} again [_{VP} be Susan married]]]

    Mary: "**That (antecedent = TP)**'s right"

        (= It is right that Susan is married again)

So, when Mary says 'That's right' the anaphora 'that' points to the only available clausal antecedent and inherits its semantic content which includes, on a parallel dimension, the not-at-issue proposition: in saying 'That's right' Mary is committed to consider correct both the information that Susan is married now and the information that Susan was married in the past. Indeed, in a scenario where Mary assumes that Susan is married now but was never married before, Mary cannot felicitously reply to John with just 'That's right'. Instead, Mary needs to

articulate more her reply by saying, for example, 'It is right that Susan is married but why do you say 'again'? She has never been married before'

(72)   John: "Susan is married again"

        Mary: (knowing that Susan is married now but was never married before):

                 #That's right

                 "It is right that Susan is married but why do you say 'again'? She has never been married before'

This is predicted under the assumption that in the sentence 'Susan is married again' there is no potential clausal antecedent that encodes the at-issue proposition to the exclusion of the not-at-issue proposition; the only available clausal antecedent of a following propositional anaphora is the structure [$_{TP}$ PRES [$_{VP}$ again [$_{VP}$ be Susan married]]] which includes the semantic contribution of 'again'. So, as soon as one replies to an utterance of 'Susan is married again' with simply 'That's right', one is automatically committed to judge correct both that Susan is married now and that she was married in the past.

So, I claim that the anaphora diagnostics as presented in the literature is not a well-formed test: it is not true that a propositional anaphora can always pick up an at-issue proposition to the exclusion of the not-at-issue proposition. Instead, the working mechanism of propositional anaphora is based on the availability of potential clausal antecedents: a propositional anaphora points to any of those available clausal antecedents and inherits their denotation which may include a not-at-issue proposition. In this sense, the anaphora diagnostics doesn't detect whether a proposition is at issue but whether a proposition is mapped onto a clausal phrase that is available to be the antecedent of a propositional anaphora. In this respect, it is interesting to see

70

how the anaphora diagnostic works in the case a not-at-issue proposition introduced by a clausal phrase, like in the case of relative appositive clauses. For example, Syrett & Koev (2015) report experimental findings showing that the sentence-final appositive relative clause in sentences like 'I introduced him to my friend Sophie, *who is a classical violinist*' can be the target of a 'that's not true' reply. Based on my discussion above, this finding suggests that a sentence-final appositive relative clause can serve as the antecedent of a propositional anaphora, hence making its denoted proposition available to be the referent of the anaphora, no matter the at-issueness status of such a proposition.

To summarize, I have argued that the anaphora diagnostic does not detect whether a proposition is at-issue but whether a proposition is mapped to the clausal object that serve as the antecedent of a propositional anaphora. And if a clausal object encodes both an at-issue proposition and a not-at-issue proposition, then a propositional anaphora taking the clause as an antecedent cannot selectively refer to the at-issue proposition to the exclusion of the not-at-issue one: a propositional anaphora inherits from its antecedent clause the whole semantic content denoted by the clause.

Let's consider now the case of evidentially marked clauses. No matter its actual morphosyntactic realization, the evidential morpheme is widely claimed to be interpreted as a sentential operator taking as an argument the TP (responsible for introducing the scope proposition) and projecting a clausal EvidP (responsible for introducing the evidential proposition) (see Matthewson et al. 2007; Murray, 2017; Korotkova, 2020). Assuming this syntactic analysis then clauses with an INFER morpheme have two potential antecedents for a following propositional anaphora: the TP and the EvidP.

(73)   Susanna be-INFER sick.

[$_{EvidP}$ INFER [$_{TP}$ Susan be sick at t]]

The question that arises is: can both the TP and the EvidP be the antecedent of a following propositional anaphora? The answer is yes. One can reply to an utterance of INFER-p with a propositional anaphora that targets either the TP or the EvidP.

(74)   John: "Susanna sara' malata"

[$_{EvidP}$ INFER [$_{TP}$ Susan be sick at t]]

Claudia: "Se è **così (antecedent=TP)**, mi dispiace per lei

'If it is **so (=TP)**, I'm sorry for her'

(= If it is the case that she is sick, …)

Claudia: "Perche' dici **questo (antecedent=EvidP)?"**

'Why do you say **that (=EvidP)**?'

(=Why do you say that Susan be-INFER sick?)

Claudia: "Anch'io penso **questo** (**antecedent** = **EvidP** or **TP**)"

'I also think **so (=EvidP or TP)**'

(=I also think that she will be sick or I also think that she is sick)

Does this test reveal anything about the at-issueness of the proposition conveyed by INFER? No. Consider the case of Claudia replying to John by saying 'I also think so', where 'so' can either take the whole EvidP or just the TP as an antecedent. What do we learn from this? We learn two things: i) the anaphora 'questo' in Italian can take the clausal EvidP as an antecedent and inherit its semantic content; ii) EvidP can be embedded under 'I think' as shown by spelling out Claudia's reply: "Anch'io penso che Susanna sara' malata" (I also think that Susan will be sick). This diagnostic only shows that the semantic content of an Evidential Phrase can be embedded under the predicate 'think' but it doesn't say anything about the at-issueness status of that semantic content. As I discussed above, the interpretation of INFER-p under 'think' favors a not-at-issue analysis of INFER but that's information comes from the application of the embedding under doxastic predicates diagnostic.

To summarize, the application of the anaphora diagnostic to a sentence INFER-p shows that INFER-p is a clausal phrase which a propositional anaphora can take as an antecedent and whose semantic content can be inherited by that anaphora; but I argue that it doesn't say anything about the at-issueness status of that semantic content.

2.4 The truth-evaluation of INFER-p

Does it make sense for an utterance of an Italian INFER-p to raise the question whether it is true or false? Imagine Gianni uttering 'Susanna sarà malata' ('Susan will be sick'). In my intuition, it doesn't make sense to ask the question "Is what Gianni said true or false?". Indeed, to answer this question, I would take the sentence S uttered by Gianni and insert them under 'it is true (false) that'. But this embedding operation yields a weird effect when I perform it with 'INFER-p'. Let me explain why.

Let's start by considering this scenario. Claudia knows that Gianni has just visited Susan, so he has seen Susan sick. In this case, Claudia cannot judge Gianni's utterance false but only wonder why he used the INFER morpheme.

(75)   Gianni: "Susanna sarà malata"

                        'Susanna will be sick'

        Claudia: #Non è vero che sarà malata. Tu l'hai vista malata.

                        'It is not true that Susan will be sick. You saw her sick'

        Claudia: Perche' dici "sarà" se l'hai vista malata?

                        'Why do you say "sarà" if you have seen her sick?'

Why can't Claudia deem Gianni's utterance 'false'? Korotkova (2016) suggest that, since the information conveyed by Gianni's statement (I have concluded that Susan is sick) was acquired by the speaker through introspection on which the speaker herself is the only epistemic authority no third person can correct that statement by deeming it false or wrong. However, if we consider an explicit version of the INFER-p statement like 'I have concluded that Susan is sick', I observe

that it is acceptable for the interlocutor to deem that utterance false if the interlocutor has reason to believe that the speaker has actually seen the described situation.

(76)   Gianni: "Ho concluso che Susanna è malata"

                            'I have concluded that Susanna is sick'

              Claudia: Non è vero che hai concluso che Susanna è malata. Tu l'hai vista malata.

                            'It is not true that you concluded that Susanna is sick. You saw her sick'

              Claudia: Perché' dici "ho concluso" se l'hai vista malata?

                            'Why do you say "ho concluso" if you have seen her sick?'

So, why can a listener deem as 'not true' a speaker's statement 'I conclude that p' but not 'INFER-p' if they are assumed to communicate more or less the same message? I suggest that this difference has to do with the semantic status of INFER. Notice that to state the falsity of the sentence uttered by Gianni 'I conclude that Susan is sick', Claudia had to change the indexical occurring in the sentence from 'I' to 'you' in order to maintain the referent of the indexical right (i.e., John); if Claudia keeps 'I' in the sentence and says 'It is not true that I have concluded that Susan is sick' she would automatically be denying another proposition than the original one meant by Gianni. This is a key point: while the indexical in 'I conclude that p' is overt and available for being manipulated, the indexical in 'INFER-p' is covert and not manipulable. INFER always refers to the current judge of 'p'. Now, when one is asked to perform a truth-value assessment of an utterance of a sentence S, one is producing another utterance where a

75

sentence denoting the same proposition as S is embedded under 'it is true (false) that'. Crucially, when you do that with 'INFER-p' you obtain a weird semantic effect.

Recall what happens when INFER-p with these features is embedded under 'it is not probable that': the negated epistemic operator takes as its complement the at-issue proposition yielding the at-issue message 'it is not probable that Susan is sick' and INFER conveys as a not-at-issue message that 'the speaker as the current judge of 'Susan is sick' has concluded that Susan is sick.'

(77)   #Non è probabile/evidente che   Susanna   sarà   malata

          Not is probable/evidente   that   Susan   INFER.be.3sg sick

          'It is not probable/evident that Susan will be sick'

          Interpretation:

          = ?!It is not probable/evident that Susan is sick, which the speaker has inferred

A similar problem arises when anyone tries to embed INFER-p under 'it is not true that'.

(78)   #Non è vero che   Susanna   sarà   malata

          Not is   true   that   Susan   INFER.be.3sg sick

          'It is not true that Susan will be sick'

          Interpretation:

          = ?!It is not true that Susan is sick, which the speaker (Claudia) has inferred

          ≠ It is not true that you (Gianni) have concluded that Susan is sick

So, one cannot judge an utterance of INFER-p 'not true' because to do so you will have to perform the utterance of 'it is not true that INFER-p' which, by virtue of the inherent indexical nature of INFER, would commit oneself to have concluded that 'p'. To keep the INFER to refer to the original judge you need to reply by anchoring the judge to the original speaker through something like 'You are wrong in saying that INFER-p'.

But maybe you can judge an utterance of INFER-p 'true'? No. To see why, let's consider what happens when you say 'it is true that INFER-p': you would not be committed to the intended message 'it is true that you (the original speaker) have concluded that p' but you would be committed to the truth of 'it is true that p, which I (the current speaker) have inferred'.

(79)   Gianni: "Susanna sarà malata"

         'Susanna will be sick'

         Claudia: È vero che sarà malata.

         'It's true that Susan will be sick'

         Interpretation:

         = ?!It is true that Susan is sick, which the speaker (Claudia) has inferred

         ≠ It is not true that you (Gianni) have concluded that Susan is sick

But can you deem the utterance of INFER-p 'not true' if you know that 'p' is false? No because of the same reason: by saying it is not true that Susan will be sick you will be committed to say 'it is not true that Susan is sick' which is your intended message, but you would also be committed to have concluded that Susan is sick. The natural reply in this case is to say to the speaker 'you are wrong'.

(80)   Giovanni: "Susanna sarà malata"

'Susanna will be sick'

Claudia: #Non è vero che sarà malata. L'ho appena incontrata ed era in

ottima salute.

'It is not true that she will be sick. I just met her, and she was in

great shape'

Claudia: Ti sbagli (nel dire che sara' malata). L'ho appena incontrata ed era

in ottima salute

'You are wrong (in saying that she will be sick). I just met her, and

she was in great shape'

Similarly, if you know that 'p' is true, usually you know that not based on an inference. But if you say 'it is true that INFER-p' you are committed to have inferred that 'p'. To reply, one can say 'You are right'.

(81)   Giovanni: "Susanna sarà malata"

'Susanna will be sick'

Claudia: #È vero che sarà malata. Mi ha appena telefonato dicendomi di

avere la febbre.

'It is true that she will be sick. She just told me by phone that she

has a fever'

78

Claudia: Giusto / Hai ragione (nel dire che sarà malata). Mi ha appena telefonato dicendomi di avere la febbre.

'Right / You are right (in saying that she will be sick). She just told me by phone that she is sick'

To summarize, in my intuition an utterance of the Italian INFER-p can never be judged either 'vero' ('true') or 'falso' ('false') because it cannot be felicitously embedded under 'it is (not) true that'. As soon as a speaker utters (or says in their mind) 'it is not true that INFER-p', the speaker would be committed to be the relevant judge of 'p' whom INFER points to.

2.5 Summary

In this chapter, I have investigated some critical interpretive properties of Italian clauses containing the non-temporal future ('INFER-p'). The output of this investigation can be summarized as follows:

(I)   INFER ascribes an inferential mode of acquisition of "p" always to the relevant judge of "p" (the speaker in matrix contexts, the attitude holder in embedding contexts, the original source of a report in reportative contexts, the addressee in interrogative contexts).

(II)   INFER does not encode an epistemic evaluation of 'p' but the weak commitment to the truth of 'p' is due to the combination of an extralinguistic considerations pertaining to how people perceive the reliability of information obtained through an inference and the competition with bare 'p' which is the natural way to convey knowledge

(III)   INFER is a sentential operator triggering a not-at-issue proposition (it projects out of the scope of other operators)

(IV)   An utterance of INFER-p cannot be judge either true or false because it cannot be felicitously embedded under 'it is (not) true that'

So, what is INFER-morpheme? Following a suggestion in Pancheva & Rudin (2019) about the reportative evidentials, I propose to analyze INFER-morphemes like the gender morpheme on pronouns. The semantic job of gender morphemes on pronouns is to introduce a not-at-issue restriction on the gender of the referent of the pronoun. Analogously, the semantic job of an INFER-morpheme is to introduce a not-at-issue restriction on how the contextually defined judge of the at-issue proposition acquired the at-issue proposition.

$[\![INFER]\!]^{M,g,c} = \lambda p_{<s,t>}:$ judge of p has acquired through an inference that p.

"INFER-p" = "p, which its current judge has acquired through a reasoning"

In its matrix uses, the meaning of INFER-p can be paraphrased as 'I'm committed to the truth of p which I obtained through a reasoning (given some facts A)'. Now, although in principle it is possible to be in a situation where a piece of information obtained through an inference counts as knowledge, in daily life we are biased to consider an inference/reasoning as a source of information whose strength is generally weak and highly dependent on the authority and the competence of the reasoner; typically, in daily life communication, most of our reasonings are far from reliable. Moreover, as soon as one believes that a proposition 'p' describes a piece of knowledge there is a communicative pressure on the speaker to utter 'p' as a bare statement; adding words or morphemes that introduce additional pieces of meaning other than just 'p' – like the inferential meaning added by 'INFER' – invites the listeners to infer that the speaker is not in the position of presenting 'p' as piece of 'knowledge'.

As a consequence, people are biased to interpret someone communicating 'I'm committed to the truth of p which I obtained through a reasoning (given some facts A)' as communicating 'I formed the opinion that p through a reasoning (given some facts A)' So, in daily communication 'INFER-p' is understood as strong as 'I formed the opinion that p based on some relevant evidence'.

In the next two chapters, I'm going to argue that epistemic necessity verbs like 'must' are used in daily life with the same meaning as 'INFER', namely as an inferential morpheme.

Chapter 3.   Epistemic necessity verbs: the theories

3.0 Introduction

This chapter addresses the following question: what is the semantic analysis of a necessity verb in its epistemic interpretation? To simplify the task, I focus here on the English word 'must' and the Italian word 'dovere'[4] in their use as epistemic necessity verbs, exemplified below.

(82)   *English*

John: "I'm wondering where Susan is"

Mary: "She **must** be in her office. I see the light on"

Italian

Gianni: "Mi   chiedo dove   sia   Susanna"

         Myself ask.1sg where be.Subj.3sg Susan

         'I wonder where Susan is'

Maria: "**Deve** essere nel   suo ufficio. Vedo   la luce   accesa"

         'Must be   in+the her office   See.1sg the light on

         'She must be in her office. I see the light on'

---

[4] Here I focus on the uses of the word 'dovere' in the indicative mood, leaving aside the cases when it is inflected in the conditional mood.

The meaning of epistemic necessity verbs has been a matter of lively debate in the semantic and philosophical literature at least since the 1970s (see a.o. Kartunnen, 1972; Jackendoff, 1972; Lyons, 1977; Kratzer, 1981, 1991; Stone, 1994; Westmoreland, 1998; Palmer, 2001; Drubig, 2001; Papafragou, 2006; Stephenson, 2007; Matthewson, 2015; von Fintel & Gillies, 2010, 2021; Lassiter, 2016, 2017; Giannakidou & Mari, 2016, 2018; Goodhue, 2017; Mandelkern, 2019; Roberts, 2019; Del Pinal & Waldon, 2019; Degen et al. 2019; Waldon 2020).

First, I review the trigger of such debate: the standard view adopted in logical semantics of statements containing epistemic necessity auxiliary verbs (which I refer to as 'must p'). Under this view, a speaker asserting 'must p' is communicating that they know some facts which make 'p' a necessarily true proposition, which predicts that speakers asserting 'must p' are in such a strong epistemic position towards 'p' that they can also assert 'it is certain that p', 'I know that p', or simply 'p'. (Sections 3.1-3.2). However, the prediction of this hypothesis is at odds with the intuition of semanticists and with naturalistic occurrences of 'must p' which suggest that speakers overwhelmingly use 'must p' without being in such a strong epistemic position towards 'p', i.e., when they would not say that they are certain that 'p' or they know that 'p'. (Section 3.3).

Next, I review the debate that was triggered by this conflict between the standard logical hypothesis and the common use of 'must p', where I identified these four positions: (i) people utttering 'must p' do assert that they know some facts which make 'p' a necessarily true proposition as predicted by the standard hypothesis but tend to overuse these assertions in an exaggerated fashion, which I refer to as the 'hyperbolic logical must hypothesis'; (ii) people uttering 'must p' assert that they know some facts and hold some believes which together make 'p' a necessarily true proposition, which I refer to as the 'weak logical must hypothesis'; (iii)

people uttering 'must p' assert that they know some facts which make highly probable that 'p' is true, which I refer to as the 'probabilistic must hypothesis'; (iv) people in uttering 'must p' use 'must' as an inferential evidential, namely they assert that they know some facts which make them conclude (i.e., form the belief) that 'p' is true, which I refer to as the 'evidential must hypothesis'. (Section 3.4).

Lastly, I offer my hypothesis in this debate. Following a suggestion in Lyons (1977), I argue that the standard logical hypothesis is not wrong: it defines correctly the original epistemic meaning of the words 'must' and 'dovere', which is still available to some speakers. However, typically what we call epistemic use of 'must' and 'dovere' is not based on this meaning but on the meaning defined by the evidential hypothesis. So, I argue that currently epistemic 'must p' is polysemous between two senses: the logical sense (p is necessarily true given some salient facts) and the evidential sense (the speaker concludes that p given some salient facts). Furthermore, I speculate that this evidential sense of the words 'must' and 'dovere' is a derivative meaning of the words stemming from their overuse as a marker of epistemic necessity: the original meaning indicating logical conclusions (i.e., certainties) is weakened/bleached to indicate just conclusions (i.e., opinions), after reiterated and implausible exaggerated uses of the words in their logical sense.

So, we think that the hyperbolic logical hypothesis is right in identifying a process of exaggeration as the key component in the epistemic 'must' puzzle; but that hypothesis is wrong in assuming that at the current stage of the language speakers exaggerate when they use the epistemic 'must'. In fact, after reiterated exaggerated uses as necessity operators, currently, the words 'must' and 'dovere' in their epistemic uses have turned into inferential markers. (Section 3.5)

3.1 The meaning of epistemics

In section 1.4, I have claimed that the inherent ambiguity of auxiliaries like 'must' and 'deve' across European languages has been used as the strongest piece of evidence supporting the hypothesis that the epistemic meaning (which has to do with estimating a conclusion's likelihood), the deontic meaning (which has to do with defining obligations/permissions), the dynamic meaning (which has to do with defining abilities/needs) form a natural category, namely the category of the modal meanings (i.e., those that have to do with defining propositions as possible/necessary).

This observation has led some scholars to the conclusion that one interested in modelling the meaning of the epistemic expressions is bound to also explain how the same model applies to the deontic and the dynamic expressions. For instance, this is the line of reasoning initiated by Kratzer (1977) and which has become the standard view among formal semanticists. Kratzer models the meaning of all modal expressions with the mathematical tools developed by logicians for analyzing reasonings involving the concepts of necessity and possibility: the tools of modal logic based on the concepts of possible world and of quantifiers over possible worlds. The main advantage of such a framework is that it can be easily used for the analysis of the deontic and dynamic expressions. Here is a brief overview of how this account works (cf. Portner 2009; Fintel & Heim 2011; Kratzer 2012).

From a syntactic point of view, a modal is analyzed as a raising verb that occupies one of the functional heads in the middle field of clauses (cf. Wurmbrand 1999 for arguments supporting an analysis of modal verbs as raising verbs no matter their interpretation). From a semantic point of view, a modal expression is an at-issue operator which takes as an input the proposition expressed by the complement clause (let's call it 'p') and yields as an output another

proposition 'q' labelling 'p' either as 'necessary' or 'possible'. In the framework of possible worlds semantics these concepts of 'necessary proposition' and 'possible proposition' translates into the concept of '**proposition true in all possible worlds**' and '**proposition true in at least one possible world**':

$[[[_{CP}$ necessity modal$] [_{TP}$ p$]]]^{M,g} = $ in all the possible worlds w', p is true in w'

$[[[_{CP}$ possibility modal$] [_{TP}$ p$]]]^{M,g} = $ in at least one possible worlds w', p is true in w'

What determines, then, the different interpretations (dynamic, deontic, and epistemic)? Kratzer suggests that modal expressions come with a hidden variable providing a set of assumptions (which she calls '**conversational background**') restricting the domain of the possible worlds which the modal operator quantifies over. This variable can be conceived as part of the context parameter (c).

$[[[_{CP}$ necessity modal$] [A] [_{TP}$ p$]]]^{M,g,c} = $ 'in all the possible worlds w' where the A-assumptions holds, p is true in w'.

$[[[_{CP}$ possibility modal$] [A] [_{TP}$ p$]]]^{M,g,c} = $ 'in at least one possible worlds w' where the A-assumptions hold, p is true in w'.

Depending on the type of assumptions listed in A, one obtains the different interpretations. For instance, the epistemic interpretation is obtained when the implicit variable A contains assumptions describing some fact(s) which count as evidence supporting the conclusion 'p'.

(83)   *Possible worlds analysis of epistemics* (basic version)

    a. ⟦ [[$_{CP}$ necessity epistemic] [A-facts] [$_{TP}$ p]] ⟧^{M,g,c} =

        'in all the possible worlds w' where the A-facts occur, p is true in w'

    b. ⟦ [[$_{CP}$ possibility epistemic] [A-facts] [$_{TP}$ p]] ⟧^{M,g,c} =

        'in at least one possible worlds w' where the A-facts occur, p is true in w'

However, the possible world analysis comes with a shortcoming. The tool is based on a discrete binary distinction between necessary propositions and possible propositions while the epistemic dimension is based on a gradient scale ranging from 1 (maximal likelihood) to 0 (zero likelihood). Therefore, by applying the possible world machinery to the epistemic domain one must fit an inherently discrete conceptual system to an inherently continuous conceptual system. For example, explicitly probabilistic epistemic expressions like 'there is a 30% chance that' cannot be analyzed in Kratzer's theory: the possible world machinery is inherently not suitable for such fine distinctions in probability.

For this reason, some scholars have argued that it would be more natural to model the epistemic dimension with a mathematical tool that is inherently meant for continuous dimensions like probability theory (e.g., Swanson 2011, 2015; Yalcin 2010, 2012; Lassiter 2016, 2017; Moss 2015, 2018; Rudin 2016, 2018; Santorio & Romoli 2017; Charlow 2020). It is important to emphasize that the contrast between probability-based approaches and possible worlds-based approaches to epistemics is a matter not only of theoretical preferences but also of which stance one has on the systematic ambiguity between the epistemic meaning and the deontic/dynamic meanings observed in the use of modal auxiliaries. In this sense, scholars who choose the

probability analysis of epistemics show that they are primarily interested in properly capturing the notion of epistemicity without the constraint of having to account for how this notion relates to the deontic/dynamic notions; instead, scholars who choose the possible world model of epistemics show that they are primarily interested in explaining the notion of epistemicity as a variant of the deontic/dynamic notions.

Here, I'm not interested in this contrast between the possible world analysis and the probabilistic one because for all I'm concerned here the two models are commensurable: Kratzer's possible world analysis of necessity modals and possibility modals can be easily translated into a probabilistic analysis and vice versa. In fact, the possible worlds notion of 'proposition true in all possible worlds' and 'proposition true in at least one possible world' can be translated into the notions of '**proposition with probability 1**' and '**proposition with probability above 0**'.

(84) *Probabilistic analysis of epistemics* (basic version)

   a. ⟦ [[$_{CP}$ necessity epistemic] [A-facts] [$_{TP}$ p]] ⟧^{M,g,c}

       'P (p | A-facts) = 1'

       'the probability that p is true given the facts in A is maximal'

   b. ⟦ [[$_{CP}$ possibility epistemic] [A-facts] [$_{TP}$ p]] ⟧^{M,g,c}

       'P (p | A-facts) > 0'

       'the probability that p is true given the facts in A is above 0'

In this work, I adopt the probabilistic analysis.

To summarize, in the view assumed here epistemics are analyzed as at-issue sentential operator that take a proposition 'p' as their input and return another proposition 'q' stating the **conditional probability of 'p' given a contextually salient set of facts**.

3.2 The standard meaning of epistemic necessity verbs

Epistemic verbs like 'must' and 'dovere' have most commonly been taken by logically oriented semanticists to behave like epistemic necessity expressions, namely as expressions assigning to a proposition the maximal conditional likelihood given some salient facts. To be more explicit, according to this view, 'must p given A' not only asserts that 'p' is a logical conclusion from the premises in A but also asserts that all the premises in A are true. I'll refer to this as the **standard logical analysis** of epistemic necessity verbs.

(85)   *Logical analysis of epistemic necessity verbs*

$[\![$ $[[_{CP}$ epistemic must/deve] [A] [$_{TP}$ p]] $]\!]^{M,g,c}$

=

'All the propositions in A are true and P (p | A) = 1'

Let's apply this analysis to the example at the beginning of the chapter.

(86)   John: "I'm wondering where Susan is"

Mary: "She must be in her office. I see the light on"

= P (Susan is in her office at t | The light is on in her office at t) = 1

Under this account, thus Mary is understood to be communicating that it is true that the light in Susan's office is on and that the event of Susan being in her office at the utterance time is 100% likely given the fact that the light in her office is on at the utterance time.

This analysis predicts that speakers asserting 'must p (because A)' are basically communicating that 'p' is a fact which follows by logic from other facts, thereby committing themselves to be in the epistemic position to also assert 'it is certain that p because A', 'I know that p because A', and 'p because A'.

To summarize, the standard logical analysis of words like 'must' or 'deve' assumes that words like 'must' and 'deve' carry with them the concept of 'necessary proposition given a set of true assumptions' which is retained across all the conceptual domains where the words are used. Specifically, when applied to the epistemic domain, this concept naturally translates into the concept of 'maximally likely proposition given a set of contextually salient facts (true propositions)'. Assuming this analysis, speakers asserting 'must p' and 'deve p' are predicted to consider 'p' as a fact necessitated by other facts, thereby being committed to be able to also assert 'it is certain that p', 'I know that p', or simply 'p'.

3.3 The standard meaning of epistemic necessity verbs is too strong

The standard logical analysis of epistemic 'must', however, is at odds with the intuition which was first expressed by Kartunnen (1972): "In general one would use [epistemic 'must p'] only in circumstances where it is not yet an established fact that ['p']" (Kartunnen 1972: 12). This intuition is shared virtually by every semanticist: "Confronted with Karttunen's Problem, semanticists have reacted with an overwhelming consensus that the meaning of epistemic must needs to be weaker than classically predicted [...] a consensus that has mantra status" (von Fintel & Gillies 2010: 352). To corroborate this intuition, Lassiter (2016) reported several corpus examples suggesting that "speakers use *must p* when they are not maximally certain of *p*; when they explicitly consider *non-p* to be a possibility; and when their stated grounds for concluding that *p* must be the case are less than fully compelling, and even explicitly stated to be so by the speaker" (Lassiter, 2016: 121). Below are some of the examples reported by Lassiter (2016) showing that English speakers naturally produce epistemic must p statements followed by explicit denial of knowledge of 'p' and explicit denial of certainty.

(87)   a. *Must plus explicit denial of knowledge* (Lassiter 2016: 123; Examples 11-17)

This is a very early, very correct Mustang that has been in a private collection for a long time. …The speedo[meter] shows 38,000 miles and it must be 138,000, but I don't know for sure.

They don't repaint the helmets, I don't know for sure why, but it must be too much of a hassle to do.

Spencer's "Native Soil" prints that I've seen are traditional darkroom prints. I've heard he is currently working with inkjet printing which must mean PS work too, but I don't know that for sure.

This is not non-stick. There must be others but I don't know for sure.

In Despeado34's post the picture from JAXTRIUMPHGUY shows a nice little rack on the gas tank. I've seen these before and they must be aftermarket but I don't know for sure.

I don't know for sure, sweetie, but she must have been very depressed. A person doesn't do something like that lightly.

It's not a bad old thing. :) Must be '60s but I don't know for sure.

b. *Must plus explicit denial of certainty* (Lassiter 2016: 125; Examples 26-31)

That face…was it not the same face as the shepherd he had met on the way? The one who gave him the straw? It must be, he was almost certain.

That disturbing thorax, jointed into those long legs, casting a greater shadow than they have any right to. Yes, it must be the legs. I'm almost certain. It's in the name after all–Daddy Long Legs.

This spot might be good fishing, I've always thought, though I haven't seen a soul out there trying. The land must be private, I'm almost certain.

Sophia trailed off. Another epiphany exploded in her mind. "The man must be Bren. I'm almost sure of it. You can see the way he looks at her...like she's his whole world."

I just deactivated the email routing, and same happens, the fax stays as "retrying routing". This must be a bug, I'm almost sure!

There must be some sort of limitation. I'm almost sure in that I cannot redeem this free night voucher towards my $3k/night stay.

Beside semanticists' intuition and naturally occurring examples, in some recent literature there have been reported a few experimental findings further suggesting that the expected meaning of epistemic 'must' is too strong.

Lassiter (2016) asked participants to read a lottery scenario where the probability of the event of Bill having won the lottery is known and very small (1 chance out of 1000), and then to judge whether they agree or disagree with a single statement out of a list of nine statements which included 'Bill must not have won the raffle' ('must not'), 'It is certain that Bill did not win the raffle' ('certain not'), and 'We know that Bill did not win the raffle' ('know not'). Crucially, the author found that more participants agreed with 'must not' (58%) than with 'certain not' (25%) and 'know not' (21%) - suggesting that 'must p' expressed a lower confidence in the truth of 'p' than 'it is certain p' and 'we know p'.

Degen et al. (2019) designed a battery of experiments meant to test the meaning of several English and German expressions with respect to the strength of speaker commitment. In one of their experiments, participants were first told: "You are in a windowless room. Your friend X walks in and says _". Then, in the blank space each participant saw one of the following four sentence types: *p* (bare), *must p* (must), *probably p* (probably), *might p* (might). After reading one of these statements, each participant answered the question 'How likely do you think it is that *p*?' by adjusting a slider with endpoints labeled "impossible" (coded as 0) and "certain" (coded as 1). The goal of this question was to test how, depending on the expression used in the statement they read, participants were convinced of the truth of p. Crucially, the authors found that participants believed that 'p' was more likely after reading 'bare p' than after reading 'must p' and that 'p' was equally likely after reading 'must p' and 'probably p' which suggest that asserting 'must p' does not commit one to also be in the position of asserting 'p'.

Del Pinal & Waldon (2019) asked participants to rate the acceptability of conjoined statements like 'must p' but I don't know for sure that p' (must condition) and 'it is certain that p but I don't know for sure that p' (certain condition) on a scale from 1 ('completely unacceptable') to 7 ('completely acceptable'). The authors found that participants gave on average a higher rating to the 'must' condition than to the 'certain' condition, suggesting that 'must p' followed by an explicit denial of knowledge in 'p' is felt less contradictory than 'it is certain p' followed by an explicit denial of knowledge in 'p'.

To summarize, there is enough evidence suggesting that the expected standard logical analysis of epistemic 'must' doesn't fit English speakers' behavior with the word: in order to properly utter epistemic 'must p', English speakers don't need to know that 'p' or be certain that 'p'. This poses a puzzle: on one hand, there is a well-motivated expectation that the word 'must' when applied to the epistemic domain should express the speaker's maximal confidence in the truth of an inferred proposition; on the other hand, when English speakers use epistemic 'must p', they seem to express a lower than maximal confidence in the truth of 'p'. What is underlying this tension? Is the standard logical analysis wrong and should be rejected? What is then the actual meaning of epistemic 'must'? Many scholars have tried to answer these questions, thereby contributing to the on-going discussion about the strength of epistemic 'must'. In the next section, I review this debate by identifying the main claims and arguments advanced in the literature.

3.4 Previous hypotheses in the debate about the strength of epistemic necessity verbs

3.4.1 Epistemic must is used hyperbolically

The most conservative hypothesis has been advanced by von Fintel & Gillies (2010, 2021). They argue that the modal logic analysis defines well the literal meaning of 'must p': a speaker uttering 'must p' is communicating 'Given some salient facts, p's likelihood is maximal', thereby committing herself to also assert 'I know that p', 'I'm/it's certain that p'. In order to account for the perceived weakness of these statements, von Fintel & Gillies (2021) suggest that 'must p' tends to be used hyperbolically -- namely in situations where 'p' is very close to be certain based on salient facts but, strictly speaking, it is just highly probable -- which may give the impression that 'must p' is weaker than it actually is. In other words, the authors propose that the cases when the epistemic 'must' is used to talk about a very probable but not certain event should be analyzed in analogy to the cases when we say, for example, 'It's 3pm' but actually it is 2:59pm: by saying 'It's 3pm' we said something that is strictly speaking false but easily tolerable in casual conversations, where the exact time is not crucial.

Similarly, for example, when a speaker utters 'Susan must be in her office. I see the light on' they are exaggerating in presenting the event that Susan is in her office as certain given the fact that the light is on inside the office, disregarding the possibility that she might have left the office and forgotten the light on. However, as soon as someone points this possibility out, the speaker has to admit that 'Susan must be in her office because the light is on in her office' is false as much as 'It is certain that Susan is in her office because the light is on in her office'.

(88)   John: "Susan must be (= is necessarily) in her office. I see the light on"

         Mary: "No, what you just said is false. It's not the case that since the light is on in her office, Susan must be (= is necessarily) there. She may have forgotten the light on when leaving"

Moreover, von Fintel & Gillies 2021 to further illustrate their hypothesis offer the following analogy: epistemic 'must p' statements stand to 'it is certain that p' statements as generic plural statements (e.g., 'The villagers are asleep') stand to universally quantified statements (e.g., 'All the villagers are asleep'). For example, in a situation where a couple of villagers are awake, it feels fine to say, 'The villagers are asleep' but not 'All the villagers are asleep'; however, as soon as someone points out that some of the villagers are awake, then everyone should admit that 'The villagers are asleep' is a false claim as much as 'All the villagers are asleep'.

(89)   John: "The villagers are asleep (= All the villagers are asleep)"

         Mary: "No, what you just said is false. The baker is awake."

I'll will refer to this as the '**hyperbolic must hypothesis**'.

(90)   *The hyperbolic must hypothesis*

$$[\![\; [[_{CP} \text{ epistemic must/deve}] \text{ [A] } [_{TP} \text{ p}]] \;]\!]^{M,g,c}$$

> **'P (p | A-facts) = 1'**
>
> **+**
>
> **Speakers tend to use this claim hyperbolically**

To summarize, the hyperbolic must hypothesis defended by von Fintel & Gillies holds that epistemic 'must p' statements are as strong as expected from the standard logical analysis – their semantics consists of presenting an inferred proposition as maximally likely given some other known facts - but they tend to be hyperbolically used in situation where a proposition is judged probable enough to be presented as certain.

3.4.2 Epistemic must relies on defeasible assumptions

Another hypothesis was introduced in Kratzer 1981 (see also Kratzer 1991, 2012) and since then it has been endorsed and refined by several other scholars (Veltmann 1996; Giannakidou & Mari 2016, 2018; Goodhue 2017; Roberts 2019; Del Pinal & Waldon 2019; Waldon 2020; Del Pinal 2021). This proposal maintains from the standard logical analysis the assumption that epistemic 'must' is associated to the concept of maximal likelihood of an event but denies that this likelihood estimation is relativized to facts only: "in uttering ['must p'] I signalize that I don't reason from established facts alone. I use other sources of information which may be more or less reliable" (Kratzer 1981: 307). What are these other sources of information? Kratzer refers to them as 'normality assumptions' or 'stereotypical beliefs', i.e.,

beliefs describing reasonable expectations about what is a normal course of events. According to this hypothesis, in asserting 'must p' speakers communicate that the likelihood of 'p' is maximal given some contextually salient facts and some contextually salient normality assumptions that they believe.

To exemplify, in asserting 'Susan must be in her office. I see the light on' a speaker is communicating that they judge the event of Susan being in her office as certain if one assumes that the light in her office is on and that, if the light is on in one's office, one's inside.

(91)   'Susan must be in her office. I see the light on'

                        =

            It is maximally likely that Mary is in her office given that

            -    the light is on in her office   (relevant fact)

            -    if the light is on in someone's office, they are inside   (normality

                                                                            assumption)

We refer to this as the '**weak logical must hypothesis**'.

(92)   *The weak logical must hypothesis*

            $[\![\ [[_{CP} \text{ epistemic must/deve] [A] } [_{TP} \text{ p]] }]\!]^{M,g,c}$

<div style="border:1px solid">

'P (p | A-facts + B-normality beliefs) = 1'

</div>

So, in summary, Kratzer speculates that the standard logical analysis is not wrong in postulating that the epistemic 'must' expresses the maximal (strongest) degree on an event's likelihood scale but is wrong in postulating that this maximal likelihood estimation is relativized to facts only: people include their beliefs among the assumptions relative to which they judge an event as certain.

3.4.3 Epistemic must is a high probability operator

The third position that I review here was first advanced in Swanson (2011) and re-elaborated in Lassiter (2016, 2017). This proposal departs more radically from the standard logical analysis than the two proposals described before: it abandons the assumption that 'must' is linked to the concept of maximal likelihood. Instead, Lassiter suggests that 'epistemic must p' indicates high probability of 'p' (or, more precisely, probability of 'p' higher than a contextually defined threshold of high probability) given some other known facts. Let's see how this works with our example. In asserting 'Susan must be in her office. I see the light on' a speaker is communicating that it is highly probable (almost certain) that Susan is in her office since the light in her office is on.

(93)   'Susan must be in her office. I see the light on'

=

It is highly probable that Mary is in her office given that the light is on in her office

I'll refer to this as the **probabilistic must hypothesis**.

(94)   *The probabilistic must hypothesis*

$$[\![[[_{CP} \text{ epistemic must/deve}] [A] [_{TP} p]]]\!]^{M,g,c}$$

<div style="border:1px solid black; padding:1em;">

'P (p | A-facts) > T

Where T is a contextually
defined threshold of high
probability

</div>

To summarize, the high probability must hypothesis defended by Swanson and Lassiter holds that epistemic 'must p' statements don't express maximal likelihood of 'p' – contrary to what the standard logical analysis says – but just high probability of 'p' relative to some contextually salient facts.

3.4.4 Epistemic must is a grammatical inferential

A fourth hypothesis was advanced in Stone (1994), Westmoreland (1998), Drubig (2001) and more recently defended in Mandelkern (2019). According to this proposal, epistemic 'must' is an inferential marker, i.e., an expression indicating that the speaker's source of information for the claim is an inference. But what is an inferential marker? In the formal semantic literature, there are several proposals about the meaning of inferential markers and evidential markers more generally (see a.o. Izvorski, 1997; Faller, 2002, 2019; Matthewson et al., 2007; McCready, 2010; Murray 2010, 2017; Matthewson 2015; Korotkova 2016, 2020). Leaving aside differences in analytical details across the proposals, all of them share the following core meaning: a statement

containing an inferential marker is used to communicate a piece of information that the speaker (or some other relevant agent in embedding contexts) has acquired through an act of reasoning.

Thus, according to this hypothesis, the communicative import of (matrix) 'must p' is roughly equivalent to the communicative import of statements containing an attitude verb of inference in the first person like 'I conclude that p' or 'I deduce that p' or 'I infer that p'. To exemplify, in asserting 'Susan must be in her office. I see the light on' a speaker is communicating that they have concluded that Susan is in her office from the fact that the light in her office is on now. So, this proposal consists in assuming that people say 'must p' in situations where some salient evidence has made them draw the inference that the event described by 'p' occurred or is occurring. We will refer to this proposal as the 'evidential must hypothesis'.

(95)   *The evidential must hypothesis*

$$\llbracket\ [[_{\text{EvidP}}\ \text{epistemic must/deve}]\ [_{\text{TP}}\ \text{p}]]\ \rrbracket^{M,g,c}$$

> **At-issue:**
>
> **p**
>
> **Not-at-issue:**
>
> **the contextually defined judge of p has acquired through an inference that p**

To summarize, the evidential must hypothesis holds that epistemic 'must p' statements don't express a commitment to a specific likelihood of 'p' – i.e., they are not epistemic statements – but they indicate that the contextually defined judge of 'p' has acquired 'p' through an act of reasoning, hence subsuming epistemic 'must' under the class of the grammatical inferential markers.

3.4.5 There are two epistemic musts: the logical one and the evidential one

The fifth position that I review here was formulated by Lyons (1977) and more recently critically discussed in Papafragou (2006), Rett (2012), and Krifka (2021). This proposal departs from the other ones in that it abandons the assumption that there is one single epistemic 'must'. Indeed, Lyons (1977) suggests that epistemic 'must' is polysemous between two senses. In one sense, "the English verb 'must' has the same function as the modal operator of logical necessity" (Lyons 1977: 789), which he refers to as the 'objective epistemic must'. In the other sense the English verb 'must' has a meaning that can be paraphrased as 'I (confidently) infer that', which he refers to as the 'subjective epistemic must'. Straightforwardly, the objective epistemic 'must' corresponds to the epistemic 'must' under the standard logical hypothesis, whereas the subjective epistemic 'must' corresponds to the epistemic 'must' under the evidential hypothesis.

Now, although Lyons suggests that in principle the two meanings are available, he also makes the remark that the subjective epistemic (evidential) must "in the everyday use of language is of more frequent occurrence" (Lyons 1977: 798). Based on this remark, we think we are justified in assuming that Lyons' belongs to the group of scholars who suggest the evidential must hypothesis to account for the perceived weakness of 'must p', with the awareness that with this hypothesis he doesn't rule out the standard logical hypothesis as an accurate explanation of some less common uses of the epistemic 'must'.

3.4.6 Summary of the proposals and critical considerations

I think that it is fair to assume that not only Lyons (1977) but virtually all the scholars mentioned above would agree that the standard logical hypothesis is basically right in assuming that the words 'must' and 'dovere' are originally associated to the concept of 'necessity': at least in the deontic realm, it is undisputed that the use of 'must' - e.g. 'Mary must wash the dishes' - signals "morally necessary" events. However, at the same time, it is also intuitively undeniable that, in their typical epistemic uses, those words seem to convey a weaker meaning than that of an epistemic necessity operator expressing maximal likelihood (certainty) of an event given other facts. In light of this consideration, the proposals advanced in this debate can all be conceived as attempts to define what happens in daily language to the strong epistemic necessity meaning of 'must' and 'dovere'. From this perspective, thus, it is possible to identify four hypotheses.

Hyperbolic logical must: People uttering 'must p' do present 'p' as certain relative to some contextually salient facts but, in many cases, they do so in an exaggerated fashion, i.e., 'must p' is as strong as logicians have postulated but people rhetorically overuse it in contexts where the likelihood of 'p' is judged near to certain.

Weak logical must: People uttering 'must p' do present 'p' as certain relative to some contextually salient facts but also to some stereotypical beliefs, i.e., 'must p' is as strong as logicians have postulated with respect to the degree of likelihood assessment (maximal) but people relativize this assessment to a set of premises including non-factual information.

Probabilistic must: People uttering 'must p' have relaxed the degree of likelihood estimation relative to salient facts conveyed by 'must' to include near to certain (highly probable) events given those facts. The difference with the hyperbolic hypothesis is as follows:

while the hyperbolic hypothesis considers the uses of 'must p' with a near to certain 'p' as a figurative way of talking obtained by stretching the certainty meaning encoded in 'must', the probabilistic hypothesis considers the uses of 'must p' with a near to certain 'p' as a legitimate way of talking licensed by the meaning currently encoded in 'must'.

Evidential must. People uttering 'must p' have relaxed the degree of likelihood estimation relative to salient facts conveyed by 'must' to include any event whose likelihood relative to those facts passes the subjective threshold above which people believe that the event happened. Again, like in the case of the probabilistic hypothesis, this hypothesis assumes that this relaxation is encoded in the current meaning of the word, thereby making 'must' inferential.

So, the critical question is: which of the four hypotheses described above provides a better account of what people are communicating in their typical epistemic uses of 'must p'? To simplify the task of answering this question, I focus on the matrix uses of 'must p'. For concreteness, let's consider a naturally occurring use of a matrix 'must p'.

(96)   When you say that your students learned less, you must have had some
         mechanism to measure that, right?[5]

In my understanding, the hyperbolic logical hypothesis holds that the speaker of this utterance is saying '**Since you say that your students learned less (relevant fact), it is certain** that you have had some mechanism to measure that' but very likely the speaker is exaggerating in using 'must' here.

---

[5] Source: BLOG   http://blogs.chicagotribune.com/news_columnists_ezorn/2012/09/test-anxiety.html

In my understanding, the probabilistic hypothesis holds that the speaker of this utterance is saying '**Since you say that your students learned less (relevant fact), it is highly probable** that you have had some mechanism to measure that'.

In my understanding, the evidential hypothesis holds that the speaker of this utterance is saying '**Since you say that your students learned less (relevant fact), I conclude / assume / suppose** that you have had some mechanism to measure that'.

Concerning the weak logical must hypothesis, it is not clear to me how its advocate would interpret the utterance in (81). In my understanding, their interpretation goes as follows: the speaker is assuming, besides the explicit premise describing a fact 'you say that your students learned less', an implicit premise describing a belief and paraphrasable as 'when someone says that their students learn less, they say that based on a mechanism that measure how much students learn' and relative to these two premises together, the speaker is deeming a proposition as certain: '**Since you say that your students learned less (relevant fact) [and since, when one says that their students learn less, one say this based on a mechanism that measure how much students learn (normality assumption)], it is certain** that you have had some mechanism to measure how much students learned'. But what is the status of the premise describing a belief (in blue)? Does the speaker assume it is true? And what is the overall message that the speaker would be conveying?

In one sense, the message can be interpreted as empirically indistinguishable from the message defined by the hyperbolic logical hypothesis: the speaker wants to convey to me that, under a normal course of events, it is a certain that the addressee had a mechanism to measure how much student learn given that the addressee says that their students learned less, which sounds very similar to a speaker hyperbolically presenting as certain that the addressee had a

mechanism to measure how much student learn given that the addressee says that their students learned less. In another sense, the message can be interpreted as empirically indistinguishable from the one defined by the probabilistic hypothesis: the speaker wants to convey to me that based on the fact that the addressee says that their students learned less, it is highly probable that the addressee had some mechanism to measure the learning performance of the student. In another sense, the message can be interpreted as empirically indistinguishable from the one defined by the evidential hypothesis: the speaker wants to convey to me that they have reached the conclusion that the addressee has a performance test based on the fact that they said that their students learned less.

If I'm asked to take a stance on this matter, I would say that the weak logical must hypothesis is empirically equivalent to the probabilistic must hypothesis: in my interpretation, the supporters of the weak logical must argue that an utterance of 'must p' means roughly 'assuming a normal course of events, it is certain that p based on facts A' which I interpret as truth-conditionally equivalent to 'it is highly probable (almost certain) that p based on facts A'. However, Del Pinal & Waldon (2019) and Del Pinal (2021) offer an alternative interpretation of the weak logical hypothesis, in which it differs from the probabilistic hypothesis: in their view, the weak logical view predicts that 'must p' commits its speakers to believe 'p' whereas the probabilistic view "only commits the speaker to believing that 'p' has a high likelihood" (Del Pinal & Waldon 2019: 158). In this respect, Del Pinal & Waldon's version of the weak must seems to not empirically differ from the evidential must hypothesis because crucially both hypotheses imply that a speaker uttering 'must p' is committed to 'I believe that p (based on a reasoning)'.

In any case, determining which of these two interpretations is more accurate is not my goal here because I'm only interested in establishing whether the weak logical must hypothesis makes different predictions from the other three hypotheses. In this respect, it seems to me that, no matter whether we adopt my interpretation or Del Pinal and Waldon's, the weak logical must hypothesis doesn't define different predictions than any of the other three hypotheses. So, for my purposes here I simplify the issue and I assume that the weak logical hypothesis is empirically equivalent to the probabilistic hypothesis in the sense that it makes the same predictions in the tests that I'm going to present in the next chapter. So, eventually, I identify three hypotheses to account for the typical weak uses of 'must p' which I summarize as follows.

Hyperbolic logical must	Must p (given A) = it is certain that p given A + hyperbolic pragmatic enrichment
Probabilistic must **= weak logical must**	Must p (given A) = It is highly probable that p given A (= In a normal world it is certain that p given A)
Evidential must	Must p (given A) = I conclude p given A

Table 2. Schematic summary of the hypotheses about the strength of epistemic necessity operator

3.5 The hypothesis I defend in this work

3.5.1 Preview

The proposal that I'm going to defend here is inspired by Lyons (1977). To preview, I claim that:

i)   epistemic 'must' is polysemous between the sense defined by the standard logical analysis and the sense defined by the evidential hypothesis

ii)   the logical sense represents the original sense

iii)   the evidential sense is derived from the logical one through an extension process of semantic bleaching/impoverishment due to a rhetorical devaluation effect applying to the logical sense

So, with the proposal here, I put myself among the advocates of the evidential hypothesis as the right account for the weak meaning of epistemic 'must p' and offer a hypothesis about how this weak epistemic meaning is derived from the strong epistemic meaning postulated by logicians.

First, I'm going to clarify what I mean by polysemous word, original sense and extended sense (Section 3.5.2); then, I'm going to state explicitly the proposal (Section 3.5.3).

3.5.2 Polysemy, original sense, and extended senses

As a preliminary caveat, the view of polysemy that I'm going to present here is adopted from the cognitive linguistic approach to lexical semantics developed in Cruse (1986) and following work.

When one single phonological form is associated to more than one meaning, it is common practice among linguists to make a distinction between homonymy and polysemy.

Homonymy is obtained when the meanings associated to one single phonological form are felt as non-related to each other; in this case, it is assumed that the meanings are associated to separate words which are independently stored in the mental lexicon and which happen to have by accident the same phonological form. For example, a clear case of homonymy is represented by the word 'bank' which can be used to mean 'an organization that provides various financial services' as in 'I managed to obtain a loan from the bank' or 'the side of a river, canal, etc. and the land near it'[6] as in 'I swam to the other bank of the river'. In this case, it is hard to find any link between the two meanings, so it assumed that English speakers store in their mind two words with the same pronunciation 'bank'. From a diachronic point of view, homonymy is the result of having two distinct lexical items that accidentally happened to sound the same due to phonological changes. Homonymy is not a very common phenomenon.

Polysemy, instead, is obtained when the different meanings associated to the same phonological form are felt as somehow related to each other; in this case it is assumed that all the meanings are associated to one single entry in the speakers' mental lexicon. Interestingly, while homonymy is not a very common phenomenon, "polysemy is pervasive in natural languages, and affects both content and function words" (Falkum & Vicente 2015: 1). Thus, cases of words that are claimed to be polysemous are abundant in the language. All the following expressions can be considered as instantiating polysemy.

---

[6] Source: Oxford Advanced Learners' Dictionary.

(97)   Examples of polysemous words

Position

    a. The house is in a good *position* to observe the sea. 'position = physical location'

    b. His *position* on abortion is slightly controversial. 'position = stance'

    c. I got a better *position* in my company. 'position = job role'

See

    d. I *see* a dog in the street. 'see = perceive with the eyes'

    e. I *see* what you mean. 'see = understand'

Freeze

    f. The river has *frozen*. 'freeze = become ice'

    g. I *froze* with terror as the door slowly opened. 'freeze = become motionless'

Exponentially

    h. The radiation of the iron salt decreases *exponentially*

        'exponentially' = 'with a progression expressible through an exponential function'

    i. The sales of cars have been growing *exponentially*

        'exponentially' = 'with a faster and faster progression'

The intuition here is that the words 'position', 'see', 'freeze', 'exponentially' are used with different meanings in each case but not so different to be considered completely separated meanings: there is an intuitive link between those uses such that it feels natural to assume that the same word is used all the time (unlike homonymy where it feels like two different words are used in each occasion). Thus, the first key component of polysemy is "**relation among the meanings**". Moreover, it is felt that one of the meanings is more prominent than the other in that it is the best candidate to represent the original meaning of the word from which all the others are derived by some process of meaning extension. Thus, the second key component of "polysemy" is that **one meaning represents the original meaning of the word from which all the other meanings are derived by extension**.

So, when encountering a potential case of polysemy, one question to ask is: what is the original meaning? The following heuristics can be applied to answer this question:

(98)   *Heuristics for identifying the original meaning of polysemous word*

   a. Find out the earliest recorded meaning of the word

   b. Find out "the default reading of a word […] the one which first comes to mind when the word is encountered out of context, or the reading which one would assume to be operative in the absence of contextual indications to the contrary" (Cruse 2000: 200)

   c. Find out the reading that is the most plausible starting point of a semantic extension, i.e., find out which meaning is the most plausible source of the other meanings (Cruse 2000: 200)

The second question to ask is: which process of semantic extension has applied to the original meaning to give rise to the other meanings? The processes that lead to semantic

extensions – i.e., addition of a new meaning alongside the literal one – are the same that lead to

semantic change – i.e., replacement of the original meaning with a new meaning. In the literature

on semantic change (see Fortson 2017 for a recent overview of semantic change), there have

been identified recurrent patterns of semantic extensions across languages that rely on the

application of common figurative reasoning. Here are some of the most common.

(99)   *Recurrent types of semantic extensions*

    a. Metaphor: the new meaning is derived by applying an analogical reasoning to the

        original meaning.

    b. Metonymy: the new meaning is derived by establishing a contiguity between the new

        meaning and the original meaning.

    c. Broadening: the new meaning is obtained by applying the term to a superset than the

        set which the original meaning applies to.

    d. Narrowing: the new meaning is obtained by applying the term to a subset than the set

        which the original meaning applies to.

    e. Bleaching: the new meaning is derived as a less specific variant of the original

        meaning over time through repetition and overuse.

To exemplify, let's consider the case of 'freeze'. The basic intuition is that the two senses

– 'become ice' and 'become motionless' - are distinct but related which suggests that we are

facing a case of polysemy. Now, the crucial question is: between 'become ice' and 'become

motionless' which meaning is more likely to have generated the other through one of the

recurrent extension processes? Intuitively, it seems that the most plausible path of extension is

from 'become ice' to 'become motionless' through a metaphoric reasoning along these lines: one

of the salient features of a liquid that becomes ice is that it doesn't move anymore, so speakers

exploit this feature to establish an analogy with the general process of stopping moving. The

other direction is less plausible because it would imply that from the abstract meaning of

motionless state one derives a specific case of motionless state, namely that of ice. In fact, it is

more likely that a meaning about a specific phenomenon included in the basic experience of

human beings gets extended to a more general and abstract meaning than vice versa.

An important distinction needs to be made between '**established extended meaning of a**

word' and '**pragmatically extended meaning of a word**'. A meaning is an established meaning

of a word if it is permanently stored in the speakers' mental lexicon such that a listener

comprehends it by selecting it among the stored meanings of that word. In contrast, a meaning of

a word is conveyed through a pragmatic process when a listener has to grasp it as a conclusion of

a specific contextually driven reasoning, after having found no established meaning of the word

as appropriate in the context. Consider the following example.

(100)   'This is my **lunch**', said John, waving a five-pound note. (Cruse 2000: 108)

Here, the word 'lunch' is used to mean 'the money which the speaker will use to buy a lunch'.

This meaning can be hardly considered as permanently stored in the speakers' mental lexicon; it

is instead a meaning creatively associated to the word 'lunch' by the speaker in this specific

context of use and which the hearers are invited to infer. Assuming this distinction, a necessary

condition for a word to be said polysemous between an original meaning and one or more

extended meanings is that all the extended meanings are established meanings of the word. To

put it in different words, polysemy is at play when the reasoning that leads from the original

meaning to the extended meaning is not activated anymore in the minds of the language users. If

listeners must actively apply the extension reasoning to one of the stored meanings to get the

114

intended meaning in the specific context, then there is no polysemy involved but just contextual pragmatic enrichment. In this sense, polysemy is the endpoint of a repeated series of pragmatic enrichments: the point when the extended meaning gets encoded in the word, thereby becoming part of the semantics of the word.

3.5.3 The polysemy of epistemic necessity verbs

I propose that epistemic 'must' is polysemous between the standard logical sense and the evidential sense and that the logical sense is the original meaning of the word which gave rise to the evidential meaning through a process of semantic bleaching. By analogy, the logical sense of 'must' ('p' is a necessary conclusion by the relevant judge) stands to the evidential sense of 'must' ('p' is a conclusion by the relevant judge) as the mathematical sense of 'exponentially' ('with a progression expressible through an exponential function') stands to the extended sense of 'exponentially' ('with a faster and faster progression'): both the logical sense of 'must' and the mathematical sense of 'exponentially' represent the original strong and specific meanings of the words and the extended senses are derived through a process of loss of strength and specificity.

In the historical linguistic literature, it has been observed that language is affected by inflationary processes as much as economy: expressions with a very specific or strong meaning tend to be overused for rhetorical purposes which, overtime, can trigger an inflationary effect leading to a devaluation of the expression, resulting in more generic uses (cf. Keller 1989; Haspelmath 1999; Dahl 2001; Deo 2015). This phenomenon can be exemplified with the case of emphatic negation constructions which tend to develop into non-emphatic constructions

("Jespersen's Cycle"): constructions like the French *ne...pas* start their life as expressing emphatic negation and become over time the standard non-emphatic means of expressing negation in the language. Dahl (2001) uses the term **rhetorical devaluation** to refer to this type of phenomena where the information load of a linguistic expressions decreases with the increase of frequency of use. It is fair to assume, for example, that a rhetorical devaluation may have created the extended meaning of 'exponentially' from the original mathematical meaning.

I suggest that a similar process has taken place in the case of epistemic necessity verbs: the original epistemic meaning ('p' is a necessary conclusion) has given rise through a process of rhetorical devaluation to the extended inferential meaning ('p' is a belief that the relevant judge has formed through an inference). That the necessity meaning represents the original meaning of the word is clear by applying the heuristics described in (98).

The necessity meaning is the earliest recorded meaning of the word 'must' and 'dovere'. Indeed, there is agreement among scholars of diachronic modality that the modal vocabulary follows a stable path of semantic extension: modal expressions originate with root interpretations and later acquires the epistemic interpretation as well (see a.o. Traugott 1989, 2011; Sweetser, 1990; Bybeee et al. 1994; van der Auwera & Plungian, 1998). Specifically, it has been argued that modal auxiliaries underwent the sequence of extensions dynamic > deontic > epistemic (detailed accounts of this development for English 'must' can be found in Warner 1993 and Traugott and Dasher 2002). Assuming that in the deontic realm, the words 'must' and 'dovere' are unambiguously linked to the concept of necessity, it is reasonable to assume that this concept of necessity is originally encoded in the words when recruited for being used in the epistemic domain.

The transition from the epistemic necessity meaning (the relevant judge has concluded that p is certain) to the inferential meaning (the relevant judge has concluded that p) is the most plausible path of extension. As discussed in chapter 2, the notion of logical conclusion is practically of no relevance in daily conversation: the conclusions we talk about are virtually always considered opinions of the individual who draw the conclusion. As Westmoreland 1998 puts it "natural language has no practical use for an epistemic necessity operator" (page 54). Now, assuming a natural tendency for speakers to exaggerate their speech for rhetorical purposes, speakers tend to use the epistemic necessity verbs beyond the restricted boundaries of logical inferences to include non-logical inferences which they feel strongly confident about and which would be more humbly described with words like 'probably'. Over time, with the increase of such rhetorical/emphatic uses, at least in everyday communication, the verbs lose their status as a marker of the special case of logical inferences and becomes a generic marker of inference (a grammatical inferential) used by speakers to mark conclusions whose truth they are strongly convinced of.

One may ask at this point: why don't epistemic necessity operators develop into probabilistic operators as suggested by the supporters of the probabilistic hypothesis? I argue that an extension path from logical conclusion to generic conclusion represents a more parsimonious change than an extension path from logical conclusion to a highly probable conclusion. In order to understand why, it is crucial to understand how committing oneself to believe the conclusion of a reasoning differs from assessing a conclusion as highly probable. Intuitively, if one holds the opinion that 'p', then it is reasonable that she judges 'p' as very likely; therefore, judging 'p' as highly probable may be a necessary condition for forming the opinion that 'p'. But is it also a sufficient condition? Namely, if one judges 'p' highly probable does she automatically believe

that 'p' happened? My answer is no. For example, consider the issue of whether there is life in the universe other than on earth. Based on what experts say, it is very likely that there is life in the universe and many people would endorse this objective assessment. However, would the same people be ready to claim 'I think that there is life in the universe other than on earth'? Not necessarily. In fact, the mental act of judging an event very likely - although it seems quite strong - is actually weaker than the mental act of holding the opinion that the same event happened: by endorsing the opinion that 'p' one is making some bet on the truth of 'p', which implies a stronger form of commitment to the truth of 'p' than just assessing its likelihood as high: 'I conclude/suppose that p' is a stronger statement than 'it is highly probable that p'. In this respect, 'I conclude that p' resembles 'I take as certain that p' more than 'I take p as highly probable': a speaker uttering either 'I take as certain that p' or 'I conclude/suppose that p' is committed to believe that 'p', whereas a speaker uttering 'I take as highly probable that p' is not committed to believe that 'p'. So, the extension path that goes from an epistemic necessity operator to an inferential operator is more conservative than an extension path that goes from an epistemic necessity operator to a probabilistic operator: **the inferential meaning but not the probabilistic meaning preserves the original speaker's commitment to believe that the drawn conclusion is true**.

The claim I defend in this work is that at the current stage of language development the epistemic necessity verbs 'must' and 'dovere' are stored in the mind of English and Italian speakers with two established senses: the original necessity meaning and the extended inferential sense. In this respect, the view defended here diverges from the hyperbolic hypothesis advanced by von Fintel & Gillies which holds that the natural non-logical uses of the words 'must' and 'dovere' should be accounted as a pragmatically extended meaning of those words through a

hyperbolic enrichment. In the hyperbolic view, a speaker uttering 'must p' in a situation where 'p' is a non-logical conclusion is hyperbolically asserting that the conclusion 'p' is certain. In the view defended here, a speaker uttering 'must p' in a situation where 'p' is a non-logical conclusion is not hyperbolically presenting 'p' as a logical conclusion but is communicating that they have reached the conclusion that 'p', which is accessed as one of the established meanings of the words.

Moreover, I speculate that the same process of rhetorical devaluation can potentially affect any linguistic expression that encodes the concept of epistemic necessity like 'certainly', 'it is certain', 'necessarily'. But the rhetorical devaluation of an epistemic necessity expression is modulated by two factors: the frequency of use of the expression and how easy it is to forget that the expression originally encodes the concept of necessity. In this sense, the auxiliary verbs are the best candidate for undergoing an extension to the inferential meaning: they are short - hence suitable for a frequent use - and don't explicitly carry with them the concept of certainty / necessity like 'certainly', 'it is certain that', 'necessarily'.

In the next chapter, I'm going to offer evidence which falsify the probabilistic hypothesis and the hyperbolic hypothesis and corroborate the hypothesis that the words 'must' and 'dovere' have an established meaning as an inferential evidential. Moreover, I'll offer data based on my intuition supporting the existence of the original necessity meaning of those words.

3.6 Summary

In this chapter, first I have reviewed the standard analysis of statements containing epistemic necessity auxiliary verbs (which I refer to as 'must p') adopted in logical semantics: a speaker asserting 'must p' is communicating that they know some facts which make 'p' a necessarily true proposition. This analysis predicts that speakers asserting 'must p' are in such a strong epistemic position towards 'p' that they can also assert 'it is certain that p', 'I know that p', or simply 'p'. However, the prediction of this hypothesis is at odds with the intuition of semanticists and with naturalistic occurrences of 'must p' which suggest that speakers overwhelmingly use 'must p' without being in such a strong epistemic position towards 'p', i.e., when they would not say that they are certain that 'p' or they know that 'p'

Next, I have reviewed the debate that was triggered by this conflict between the standard logical hypothesis and the common use of 'must p', where I identified these four positions: (i) people utttering 'must p' do assert that they know some facts which make 'p' a necessarily true proposition as predicted by the standard hypothesis but tend to overuse these assertions in an exaggerated fashion, which I refer to as the 'hyperbolic logical must hypothesis'; (ii) people uttering 'must p' assert that they know some facts and hold some believes which together make 'p' a necessarily true proposition, which I refer to as the 'weak logical must hypothesis'; (iii) people uttering 'must p' assert that they know some facts which make highly probable that 'p' is true, which I refer to as the 'probabilistic must hypothesis'; (iv) people in uttering 'must p' use 'must' as an inferential evidential, namely they assert that they know some facts which make them conclude (i.e., form the belief) that 'p' is true, which I refer to as the 'evidential must hypothesis'.

Lastly, I have offered my hypothesis in this debate. Following a suggestion in Lyons (1977), I argue that the standard logical hypothesis is not wrong: it defines correctly the original epistemic meaning of the words 'must' and 'dovere', which is still available to some speakers. However, typically what we call epistemic use of 'must' and 'dovere' is not based on this meaning but on the meaning defined by the evidential hypothesis. So, I argue that currently epistemic 'must p' is polysemous between two senses: the logical sense (p is necessarily true given some salient facts) and the evidential sense (the speaker concludes that p given some salient facts). In the next chapter, I'm going to offer data supporting the existence of these two senses.

Furthermore, here, I have speculated that this evidential sense of the words 'must' and 'dovere' is a derivative meaning of the words stemming from their overuse as a marker of epistemic necessity: the original meaning indicating logical conclusions (i.e., certainties) is weakened/bleached to indicate just conclusions (i.e., opinions), after reiterated and implausible exaggerated uses of the words in their logical sense. So, I think that the hyperbolic logical hypothesis is right in identifying a process of exaggeration as the key component in the epistemic 'must' puzzle; but that hypothesis is wrong in assuming that at the current stage of the language speakers exaggerate when they use the epistemic 'must'. In fact, after reiterated exaggerated uses as necessity operators, currently, the words 'must' and 'dovere' in their epistemic uses have turned into inferential markers: in using them, speakers typically are not communicating exaggerated confidence in the certainty of an event but just that they have formed the belief that the event happened based on their reasoning.

Chapter 4.   Epistemic necessity verbs: the data

4.0. Introduction

In this chapter, I offer data supporting the polysemous nature of epistemic 'must' and 'dovere' which I argued for in Chapter 3.

On one hand, I offer data suggesting that the typical epistemic use of 'must' / 'dovere' corresponds to that of an inferential evidential like the Italian non-temporal future. (i) In comprehension tasks, English and Italian speakers understand matrix uses of 'must p' and 'dovere p' similarly to 'I conclude that p' (I believe that p based on a reasoning); (ii) several naturally occurring utterances show that both 'must' and 'deve' as well as INFER can co-occur with an epistemic adverb of any degree; (iii) in doxastic embedding contexts, 'must' and 'dovere' behave like expressions whose meaning is not-at-issue.

On the other hand, I offer data based on introspection suggesting that 'dovere', unlike 'INFER', can be interpreted as contributing a regular at-issue meaning which corresponds to that of an epistemic necessity operator. (i) 'Deve p' but not 'INFER-p' can be judged false in a context where 'p' is not certain; (ii) the meaning contribution of 'deve' in 'deve p' but not the meaning of 'INFER' in 'INFER-p' can be under the semantic scope of a sentential negation particle.

In details, the chapter is organized as follows.

First, I offer findings from four comprehension tasks requiring participants to decide whether they endorse a statement based on the given contextual information (Experiment 1-3 with English speakers, Experiment 4 with Italian speakers). These tasks were meant to assess the three hypotheses identified in section 3.4.6: the hyperbolic logical hypothesis, the probabilistic hypothesis, the evidential / weak logical must hypothesis. The findings of these experiments

support the evidential hypothesis over the other two for both English 'must' and Italian 'dovere':
English and Italian speakers typically utter epistemic 'must p' to assert that a contextually salient
reasoning made them conclude (i.e., form the belief that 'p' (Section 4.1)

Next, I offer non-quantitative data which further show that epistemic necessity verbs in
English and in Italian exhibit three properties that are typical of inferentials (as I showed in
Chapter 2): 'must p' can co-occur with epistemic adverbs of any degree; when 'must p' is
embedded under a doxastic predicate, the contribution of 'must' is not at-issue; 'must p' doesn't
felicitously occur in the antecedent of conditionals. (Section 4.2)

Then, I offer data based on introspection showing the original epistemic meaning of the
words 'must' and 'dovere'. While the contribution of INFER in 'INFER-p' cannot be targeted by
a truth-value judgement or by negation, the contribution of 'must' in 'must p' can. (Section 4.3)

Lastly, I discuss two problematic cases: the use of 'must p' embedded under 'know' and
the utterance of 'must p' in situations where 'p' is a logical conclusion from true premises. In
both cases, it is not clear if 'must' is interpreted with its original meaning of necessity or as an
inferential. (Section 4.4)

4.1 Experimental data

4.1.1 The hypotheses tested

In the previous chapter, I identified these three answers to the question "What happens in daily language to the strong epistemic necessity meaning of 'must' and 'dovere'?"

Hyperbolic logical must:

must p (given the facts A)

= it is certain that p given A + exaggeration

People uttering 'must p' present 'p' as certain relative to some contextually salient facts but, in many cases, they do so in an exaggerated fashion, by overusing 'must p' in contexts where the likelihood of 'p' is judged near to certain.

Probabilistic must / Weak logical must:

must p (given the facts A)

= it is very probable (certain in normal circumstances) that p given A

People uttering 'must p' have relaxed the degree of likelihood estimation relative to salient facts conveyed by 'must' to include near to certain (highly probable) events given those facts.

Evidential must.

must p (given the facts A)

= I conclude that p given A

People uttering 'must p' have relaxed the degree of likelihood estimation relative

to salient facts conveyed by 'must' to include any event whose subjective

likelihood is high enough to make speakers form the belief that the event

happened.

So, the critical question is: which of these three hypotheses provides a better account of what

people are communicating in their typical epistemic uses of 'must p'? In this section, I address

this question by assessing English and Italian speakers' behavior in comprehension tasks

requiring participants to decide whether they endorse a statement based on the given contextual

information. So, it is critical to understand how these three hypotheses differ in terms of their

predictions about the behavior of speakers in this type of comprehension tasks.

I take the hyperbolic logical must hypothesis as predicting that speakers would endorse

'must p' only in contexts where the given information prompts them to also endorse 'it is certain

that p' or 'I know that p'. But this hypothesis is also consistent with people sometimes

exaggerating (generating hyperbole) in some contexts.

I take the probabilistic must hypothesis / the weak logical must hypothesis as predicting

that speakers would endorse 'must p' in every context where the given information prompts them

to also endorse 'it is highly probable that p'. In other words, according to these two hypotheses, for a speaker to judge 'p' as highly probable is a sufficient condition for endorsing 'must p'.

To understand the predictions of the evidential must hypothesis, it is crucial to understand the meaning of 'I conclude/deduce/infer that p'. And to do so, one needs to examine the concept of inference. Inferences can be divided into two types: logical (the inferred conclusion is necessarily true if all the premises are true) and non-logical (the inferred conclusion could be false even if all the premises are true). Assuming this categorization, there exists one type of inference - logical inference based on true premises – which entitles the agent who draws the inference to claim that the conclusion describes a fact, i.e., the conclusion represents a piece of knowledge. Thus, in principle, it is possible to be in a situation where a piece of information obtained through an inference counts as knowledge. But, in daily life there is not much room for logical inferences, therefore people are biased to consider inferred propositions as describing opinions of the agent who draws the inference. As a consequence, people are biased to interpret someone uttering 'I conclude/deduce/infer that p given the set of facts A' as communicating to us 'I formed the belief that p given the set of facts A'. So, under the evidential hypothesis, 'must p' is as strong as 'I formed the belief that p based on some relevant evidence'.

Assuming this interpretation of the evidential hypothesis, we proceed now to show how the evidential hypothesis makes different predictions than the hyperbolic hypothesis and the probabilistic hypothesis in endorsement-tasks. On one hand, one can believe a conclusion without judging that conclusion as certain, i.e., judging a conclusion as certain is not a necessary condition for believing that such a conclusion is true. So, the evidential hypothesis, unlike the hyperbolical logical hypothesis, predicts that speakers can endorse 'must p' in contexts where they don't endorse 'it is certain that p' or 'I know that p'. On the other hand, judging a

conclusion as probable is not a sufficient condition for believing in its truth (if one judges p as probable, one doesn't automatically believe p). For example, recall the issue of whether there is life in the universe other than on earth. Based on what experts say, it is very likely that there is life in the universe and many people would endorse this objective assessment. However, would the same people be ready to claim, 'I think that there is life in the universe other than on earth'? Not necessarily. In fact, the mental act of judging an event very likely - although it seems quite strong - is actually weaker than the mental act of believing that the event happened: by believing a conclusion one is undertaking a commitment to the truth of that conclusion, which is not the case when one deems the conclusion's likelihood as high. So, the evidential hypothesis, unlike the probabilistic must hypothesis, predicts that speakers can decide to not endorse 'must p' in contexts where they endorse 'it is highly probable that p'. Overall, the evidential hypothesis predicts that the endorsement of 'must p' - as well as statements containing inferential attitude verbs (e.g., 'conclude') or inferential evidentials - is determined by the speakers' subjective assessment about whether they are convinced of the truth of a conclusion suggested from the relevant evidence, independently from their objective assessment of the likelihood of that conclusion given that evidence.

In summary, we have identified three hypotheses about the weak common use of 'must p' which make different predictions in an endorsement task, summarized in Table 3.

Hypotheses	Predictions in comprehension tasks where participants are asked whether they endorse a statement based on the given information
hyperbolic logical must	speakers would endorse 'must p' only in contexts where the given information prompts them to judge 'p' as certain (with some expected exceptions due to exaggerated uses)
probabilistic must *(= weak logical must)*	speakers would endorse 'must p' in every context where the given information prompts them to judge 'p' as highly probable
evidential must	speakers would endorse 'must p' in every context where the given information prompts them to conclude (= form the belief) that p

Table 3. The three hypotheses assessed in this work with their predictions in comprehension tasks prompting participants to decide whether they endorse a statement based on the given information

The debate about the strength of 'epistemic must' has been conducted using evidence mostly from the authors' intuitions but recently there have been a few studies aimed at experimentally assessing the hypotheses under discussion (Lassiter, 2016; Del Pinal & Waldon, 2019; Degen et al. 2019). In particular, the experiment reported in Lassiter (2016) represents the first attempt to test these hypotheses in a comprehension task across many participants. We review

this experiment immediately below and postpone the discussion of other relevant findings to the general discussion.

4.1.2 Lassiter (2016)

Participants in Lassiter (2016) were provided with a lottery scenario where the probability of the event of Bill having won the lottery is known and very small (1 chance out of 1000), and they judged whether they agree or disagree with a single statement out of a list of nine statements (see 102) which included 'Bill must not have won the raffle' ('must not'), 'It is certain that Bill did not win the raffle' ('certain not'), and 'We know that Bill did not win the raffle' ('know not') - these statements are displayed in bold.

(101)　　　Materials in Lassiter (2016)

Lottery scenario

Yesterday, Bill bought a single ticket in a raffle with 1000 total tickets. There were also 999 other people who bought one ticket each. That is, the tickets were distributed like this: People holding one ticket: Bill, Mary, Jane, … [997 more]. The drawing was held last night, and the winner will be announced this evening.

Sentences (factor manipulated between-subjects)

a. Bill won the raffle.	(did)	7%
b. Bill did not win the raffle.	(did not)	69%
c. It is possible that Bill won the raffle.	(possible)	92%
d. Bill possibly won the raffle.	(possibly)	74%

e. We know that Bill did not win the raffle.   **(know not)**   **21%**

f. It is certain that Bill did not win the raffle.   **(certain not)**   **25%**

g. Bill certainly did not win the raffle.   (certainly not) 54%

h. Bill might have won the raffle.   (might)   80%

i. Bill must not have won the raffle.   **(must not)**   **58%**

Lassiter (2016)'s main findings were: (i) most participants (58%) agreed with 'must not' in the lottery scenario; (ii) the proportion of participants who agreed with 'must not' (58%) was significantly higher than the proportion of participants who agreed with 'certain not' (25%) and 'know not' (21%) - suggesting that the 'it is certain p' and the 'we know p' statements are evaluated as expressing a stronger confidence in the truth of 'p' than the 'must p' statement. Lassiter took these findings as supporting the probabilistic must hypothesis over the logical must hypothesis for English 'epistemic must'.

However, we find Lassiter's (2016) interpretation of these findings unsatisfactory, because, although the probabilistic must hypothesis can account for the behavior of the 58% of participants who agreed with the statement, it does not have an explanation for the behavior of 42% of participants who disagreed with the statement. Instead, the two other hypotheses described above can account for the behavior of both groups. According to the hyperbolic logical must hypothesis, those who agreed with 'must not' in the context exaggerated in judging as certain an event which is only probable; instead, those who didn't agree with 'must not' in the context were interpreting the statement with its literal meaning. According to the evidential must hypothesis, those who agreed with 'must not' in the context judged the provided information sufficient to believe that Bill didn't win the lottery without judging such event as necessary;

instead, those who didn't agree with 'must not' in the context were more cautious and didn't want to jump to such a conclusion. Thus, Lassiter (2016)'s findings confirm the intuition that the logical must hypothesis is too strong but don't discriminate among the other three hypotheses reviewed above.

In this section, we offer findings from an attempted replication of Lassiter 2016 (Section 4.1.3), and three follow-up studies where we manipulated the original task to discriminate among the three hypotheses reviewed above: two studies with English speakers (Sections 4.1.4 and 4.1.5), and one study with Italian speakers (Section 4.1.6). Overall, our findings support the evidential hypothesis over the hyperbolic logical hypothesis and the probabilistic hypothesis for both English 'must' and Italian 'dovere' and provide further confirmation for the claim in the psychology of decision making that people form the belief that an uncertain event happened relying more on their subjective probability of that event than on its objective probability.

4.1.2 Experiment 1: Replication of Lassiter (2016)

We focused on three conditions out of the nine original ones – 'must not', 'know not', and 'certain not' – because these are the conditions relevant for Lassiter's primary conclusions.

Participants and methods.

In the pre-registered replication, 180 Amazon Mechanical Turkers (60 for each sentence/condition) were recruited. The three sentences/conditions are below (unlike Lassiter 2016, we labeled them without 'not').

(102)   Sentence (between-subjects design)

    a. We know that Bill did not win the raffle.   (know)

    b. It is certain that Bill did not win the raffle.   (certain)

    c. Bill must not have won the raffle.   (must)

The experiment used a between-subjects design. Each participant saw one critical sentence/trial. Participants read instructions, followed by the target sentence, a radio button choice between 'Agree' and 'Disagree', and a simple yes-no-question (intended to weed out participants who might not read the context carefully). As in Lassiter's original experiment the crucial measurement was the proportion of 'Agree' choices for each sentence, which again we interpreted as 'the proportion of participants who judge the sentence assertable in the given situation'; moreover, given any two statements p and q, if the proportion of 'Agree' responses for p is significantly higher than for q, we interpret it as suggesting that p is considered a weaker statement than q. An example trial is displayed in (8).

(103)   Sample trial

Please read the context and the sentence, state whether you agree or disagree with the sentence in the context and then answer the question immediately following.

Context: the same lottery scenario as in Lassiter (2016).

Target sentence: Bill must not have won the raffle.   Agree Disagree

Question: Is there anyone other than Bill who bought a ticket?   Yes   No

Because this experiment was designed as a replication of Lassiter (2016), we compared

only the two theories that he considered here: the logical must and the probabilistic must

hypothesis. We consider the other theories in the discussion and in later experiments. The logical

must hypothesis predicts participants will not agree with the "must" statement, nor the 'certain'

or 'know' statements: the 'Agree' proportions should be close to zero for all three. In contrast,

the probabilistic must hypothesis predicts that participants will agree with "must" (the 'Agree'

proportion should be close to 1), more so than for the 'certain' and 'know' statements.

Results

We excluded data from 25 participants because they didn't pass all of the following

inclusion criteria: (a) English as their native language; (b) USA as country of origin; (c) a

correct answer to the sanity check question "*Is there anyone other than Bill who bought a*

ticket?"; (d) they participated in only one condition. This left 155 participants. The data points,

mean agreement rates, and standard deviation for each of the three sentences are reported in table

4. The mean agreement ratings with error bars are plotted in Figure 1 (middle panel).

Sentence Type	count	mean	sd
"must"	47	0.28	0.46
"certain"	56	0.09	0.29
"know"	52	0.08	0.27

Table 4. Data points, mean agreement rate, standard deviation for each of the three
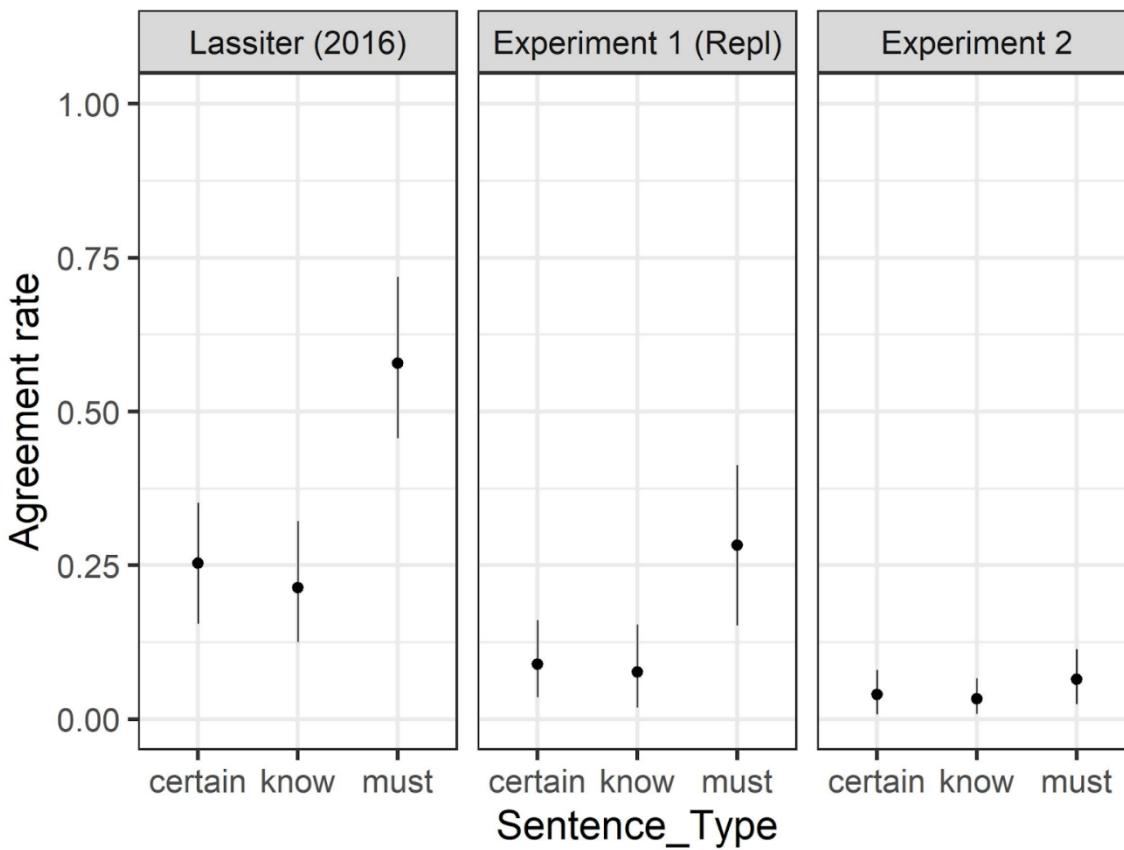
sentences in Exp 1 (Replication).

Fig. 1. Mean ratings in Lassiter (2016), Experiment 1, and Experiment 2 for *certain*, *know*, and

must. Error bars indicate bootstrapped 95% confidence intervals.

First, we observed that we obtained a lower proportion of 'Agree' responses in all three

conditions with respect to Lassiter (2016) - 0.28 for 'must', 0.09 for 'certain', and 0.08 for

'know'; we don't know for sure why that's the case. Perhaps it is because Lassiter didn't have a

comprehension question in his original design which might have increased noise (Lassiter,

p.c.) Next, we observed that the proportion of agreement for 'must' (0.28) was numerically

higher than for 'certain' (0.09) and for 'know' (0.08) and we asked whether this difference was

statistically significant. To this end, we analyzed the agreement responses of the three sentence

types with a logistic regression model with the factor 'Sentence-type' as an orthogonal contrast-

coded fixed effect (contrast 1: must = -0.66, certain = 0.33, know = 0.33; contrast 2: must = 0,

certain = -0.5, know = 0.5) by using the R function *glm*. Participants were more likely to agree

with the 'must' sentence type than with 'certain' and 'know' (β= -1.4870, z = -3.071, p =

0.00214). Agreement rates did not differ significantly between 'certain' and 'know' (β= -0.1625,

z = -0.232, p = 0.81649). Furthermore, we performed pairwise comparisons by using the R

function *emmeans* showing that there was a significant difference between 'must' and 'certain'

(β = 1.391, z = 2.4333, p = 0.04) and 'must' and 'know' (β = 1.553, z = 2.526, p = 0.03). Thus,

we successfully replicated Lassiter (2016)'s finding that 'must' is endorsed significantly more

than 'certain' and 'know', which don't differ from each other, although the observed rates of

agreement were lower overall than in Lassiter (2016). See the 'Supplemental Material' for the

full output of the statistical analyses.

Discussion

Our replication showed that in a context of almost certainty about *p*: (a) about 1/3 of

participants judged 'must p' as assertable; (b) almost no participant judged either 'it is certain p'

or 'we know p' as assertable; (c) the proportion of participants that judged 'must p' as assertable

was significantly higher than the proportion of participants that judge either 'it is certain p' or 'we know p' as assertable. We agree with Lassiter (2016) that these findings don't support the logical must hypothesis; however, we disagree with him that they support the probabilistic must hypothesis. Specifically, we consider the probabilistic hypothesis at odds with finding (a) because the probabilistic must hypothesis predicts an agreement rate very close to 1 based on the assumption that plausibly almost every participant judges 'p' very likely in the given scenario. Furthermore, finding (c) only suggests that epistemic 'must p' is weaker than predicted by the logical must hypothesis but doesn't specifically suggest that the probabilistic must hypothesis is the right account for the weakness. In fact, the other two hypotheses reviewed in the introduction are consistent with these data too. According to the hyperbolic logical must hypothesis, the minority who agreed with 'must' in the context was exaggerating in judging as certain an event which is only probable; instead, the majority who didn't agree with 'must' in the context were more careful in that they recognize that the event of Bill not winning the lottery is not absolutely warranted given the provided information. According to the evidential must hypothesis, those who agreed with 'must' in the context judged the provided information sufficient to conclude that Bill didn't win the lottery without necessarily judging such conclusion as certain; instead, those who didn't agree with 'must' in the context were more cautious and didn't want to jump to such a conclusion.

Next, we aimed at discriminating between the hypertbolic logical hypothesis and the probabilistic hypothesis (we postpone the assessment of the evidential hypothesis to Experiment 3). Notice that Lassiter (2016)'s task design (one sentence per participant without baselines) prompts uncertainty on the part of the reader as to what is intended by "agreeing" with a statement. In fact, we can think of at least two interpretations of the experimental question "Do

you agree with this sentence in the given context?": some people may interpret it as "Is this sentence something that one might say in the given context?" (This question would tap the so called assertability-conditions of a sentence) whereas some other people may interpret it as "Is this sentence true in the given context?" (This question would tap the truth-conditions of a sentence). Let's refer to the first interpretation of the task as the 'assertability task' and to the second interpretation as the 'truth-value judgment task'. Now, although in general a positive answer to the 'assertability task' (the sentence might be said in the context) implies a positive answer to the 'truth-value judgment task' (the sentence is true in the context)[7], it is possible that some types of sentences may be judged by some participants as something that one might say in a context where the same participants would judge those very sentences as false. For example, this may be the case for sentences expressing a maximal value on a given scale like the sentence "All my friends hate me": in informal talk people might say this sentence to rhetorically overstate their feeling of frustration after receiving a criticism from a couple of friends but it is very likely that the speaker themselves judges this sentence as not literally true. In this respect, the 'assertability task' is more likely to elicit answers based on this informal way of talking than the 'truth-value judgment' task, which is better suited for targeting answers based on the literal meanings of sentences. Thus, we speculate that sentences expressing a maximal value on a given scale are more likely to be agreed with under the 'assertability' than the 'truth-value judgment' interpretation of the task in contexts where a near-maximal value of the scale is defined.

But that's a crucial task feature for testing the hyperbolic logical hypothesis which does predict that if participants are induced to focus on the literal meaning of the words contained in

---

[7] But not vice versa, as showed by the case of sentences containing scalar implicature triggers like *Some of my kids went to college*: one may judge an assertion of this sentence as true in a situation where all the speaker's kids went to college but definitely not as something that one might say in the same situation.

the target sentence, then participants would converge towards the literal strong meaning of 'must p' ('it is necessary that p'). Consequently, the hyperbolic logical hypothesis predicts that the proportion of 'Agree' choices for 'must p' in an unambiguously 'truth-value judgment task' would be lower than in the original experiment (where the task is potentially ambiguous between the 'assertability task' and the 'truth-value judgment task' interpretations) and not different from that of 'certain' and 'know'. Instead, the probabilistic hypothesis ('must p' = 'the probability of p is very high') predicts no effect induced by disambiguating the task as a 'truth-value judgment' task: if anything, participants are expected to endorse 'must p' at a higher rate than 'certain p' or 'know p' after assessing that it is true that 'p' is indeed a highly probable conclusion but not 100% warranted in the given context. Therefore, to distinguish between the probabilistic and the hyperbolic logical hypothesis would be crucial to assess whether and how the endorsement rate of 'must p' is affected in a task where participants read the same scenario as in Lassiter (2016) but are induced to interpret the task as a 'truth-value judgment task'.

So, the next question is: how do we prompt participants to do so? We thought that providing participants with clearly true and clearly false statements as baseline would induce them to interpret the task as a 'truth-value judgment task': given that participants can only choose between two response options ('Agree' or 'Disagree'), they would be prompted to assign each of the three target sentences to one of two groups depending on whether the sentence is like the clearly true sentences (in this case, it would be rated with 'Agree') or like the clearly false sentences (in this case, it would be rated with 'Disagree'). In making this manipulation we were inspired by previous experimental work showing how contextual features of the task affect participants' behavior. For example, studies on scalar implicatures and presuppositions have shown that having participants explicitly evaluate the target implicature/presupposition trigger

together with relevant alternatives affects participants computation rate of the critical implication (e.g., Foppolo et al. 2012; Tonhauser et al. 2013; Skordos & Papafragou 2016; Zehr & Schwarz 2018), whereas other work has shown that participants might be inclined to reject a statement if the experimenter doesn't recreate the appropriate discourse conditions for a felicitous production of that statement (e.g., Syrett 2015; Syrett & Brasoveanu 2019).

These considerations led us to design our Experiment 2. To foreshadow the results, we found that including multiple examples to rate had a substantial effect on the results, such that participants' agreement proportion for 'must' was very similar to that for 'know' and 'certain' when multiple examples are provided to the participant to rate, including some clearly true and false items. This is in contrast to Lassiter's results and our replication. Thus, it appears that having all the three sentences for the participant to rate enables participants to evaluate 'must' similar to 'certain' and 'know' in the lottery context designed by Lassiter. In the next subsection, we present in detail the results of the experiment where we included multiple examples to rate.

4.1.3 Experiment 2: Including multiple example sentences.

In Experiment 2, we compared Lassiter's original between-subjects design to a within-subjects design, where each participant rated the 3 critical conditions of experiment 1 -- 'must', 'certain', and 'know -- together with some clearly true and clearly false sentences as baselines. This task was meant to test whether the endorsement rate of 'must' changes when participants are induced to converge on interpreting the task as a truth-value judgment task which enhances literal interpretation of the sentences: participants would be prompted to rate each of the target sentences by assigning it either to the group of the clearly true baselines (which would be rated with 'Agree') or the group of the clearly false baselines (which would be rated with 'Disagree').

This design would allow one to discriminate between the probabilistic hypothesis on one side and the hyperbolical logical must on the other side: the probabilistic hypothesis predicts that the agreement rate of 'Bill must not have won the lottery' would be like that of the clearly true sentences while the hyperbolic logical must hypothesis predicts that it would be like that of the clearly false sentences.

Notice that the evidential must hypothesis is compatible with either output. Indeed, under this hypothesis, the task would prompt participants to ask themselves whether they would conclude that 'p' based on the contextual information, which doesn't yield any quantifiable prediction: we don't possess a background theory of humans' inferential behavior which would allow us to make a clear prediction about the rate of people who would conclude that Bill did not win the lottery based on the contextual information provided in this experiment.

Participants and methods.

The story defining the scenario was the same as Lassiter (2016). The nine sentences that were seen by each participant are listed below. Notice that we included among the clearly true items the statements 'It is highly probable that Bill did not win the raffle' ('probable') and 'There is a slight chance that Bill won the raffle' ('chance'): recall that the probabilistic hypothesis predicts that 'must p' is truth-conditionally equivalent to 'it is highly probable that p' and, consequently, is compatible with 'there is a chance that not-p'.

(104)

    <u>Experimental items</u>

    a. Bill must not have won the raffle.   (must)

    b. It is certain that Bill did not win the raffle.   (certain)

    c. We know that Bill did not win the raffle.   (know)

    <u>Clearly true control items</u>

    d. It is highly probable that Bill did not win the raffle.   (probable)

    e. There is a slight chance that Bill won the raffle.   (chance)

    f. Bill bought exactly one ticket in the raffle.   (one)

    g. 1000 different people bought one lottery ticket each in the raffle.   (1000)

    <u>Clearly false control items</u>

    h. Mary bought two tickets in the raffle.   (two)

    i. The winner will be announced tomorrow.   (winner)

We recruited 180 participants on Amazon Mechanical Turk, dividing them in five groups of 36 and assigning them to one of five pseudo-randomized orders (reported in Table 5), created by varying the order of presentation of the three experimental sentences. Again, the dependent measure was the proportion of 'Agree' choices for each sentence.

Order 1	Order 2	Order 3	Order 4	Order 5
one	probable	one	certain	probable
two	winner	must	1000	must
know	one	chance	probable	winner
must	certain	know	must	two
probable	chance	probable	know	one
1000	two	two	one	1000
winner	1000	certain	chance	certain
certain	know	1000	winner	know
chance	must	winner	two	chance

Table 5. The five pseudo-randomized orders of presentation of the nine sentences in Experiment 2.

Predictions

The hyperbolic logical must hypothesis assumes that 'Bill must not have won the lottery' in its literal meaning is truth-conditionally equivalent to 'It is certain that Bill did not win the lottery' and 'We know that Bill did not win the lottery' which are expected to be judged as false in the experimental context (the conclusion that Bill did not win the lottery is not warranted given that context). Therefore, the hyperbolic hypothesis predicts that participants will agree with 'must' at a rate not significantly different from the rate of 'certain' and 'know' and numerically close to the expected rate of the clearly false baselines (i.e., very close to 0).

The probabilistic must hypothesis assumes that 'Bill must not have won the lottery' in its literal meaning is truth-conditionally equivalent to 'It is highly probable that Bill did not win the

lottery' - which is expected to be judged as true. Therefore, the probabilistic hypothesis predicts that participants will agree with 'must' at a rate significantly higher than 'certain' and 'know' and numerically close to the expected rate of the clearly true baselines (i.e., very close to 1).

The evidential must hypothesis assumes that 'Bill must not have won the lottery' in its literal meaning is truth-conditionally equivalent to 'I conclude that Bill did not win the lottery' whose agreement rate in the experimental context is not predictable. Therefore, the evidential hypothesis doesn't make any prediction in this experiment and is compatible with any output.

Results

We filtered out results from 55 participants because they didn't indicate English as their native language or USA as their country, or because they failed to correctly answer the comprehension question or because they participated in more than one condition. This left 125 participants. The data points, mean agreement rates, and standard deviation for the three experimental sentences each are reported in Table 6. The mean agreement ratings for the three experimental sentences are plotted in Figure 1 (rightmost panel). The mean agreement ratings of all nine sentences are plotted in Figure 2.

Sentence Type	count	mean	sd
must	123	0.07	0.25
certain	125	0.04	0.20
know	121	0.03	0.18

Table 6. Data points, mean agreement rate, standard deviation for 'must', 'certain', and 'know' in Exp 2.
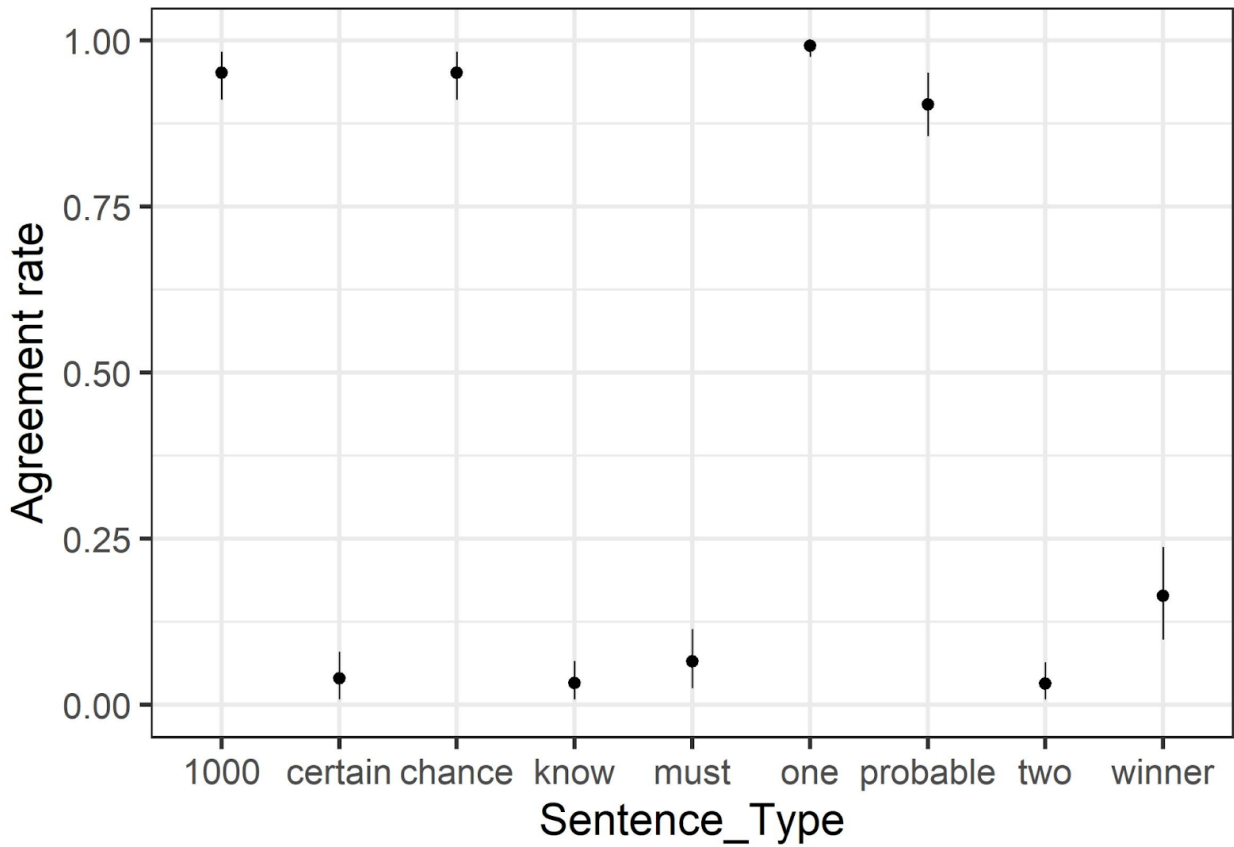

Fig. 2. Mean rates in experiment 2 for all nine sentences. Error bars indicate bootstrapped 95% confidence intervals.

First, we observed that the agreement rate of the baselines patterned as expected: the agreement rate of the four clearly true statements ('1000', 'chance', 'one', 'probable') was close to the ceiling, whereas the agreement rate of the two clearly false statements ('two', 'winner') was close to the floor. Thus, the results of the baseline statements suggest that participants were performing the task by paying attention to the literal meaning of the sentences and that nothing about the task pushed people towards lower agreement across the board. Next, we observed that the mean rating of 'must' in Experiment 2 (0.07) was numerically lower than in Experiment 1 (0.28) and very close to the mean rating of 'certain' and 'know' in Experiment 2 (0.03 and 0.04 respectively). So, we asked three questions: a) whether the mean rating of 'must' differed from the mean rating of 'certain' and 'know' in experiment 2; b) whether the probability of obtaining the observed rating decreases for 'must' from Experiment 1 to Experiment 2 was different from chance.

To answer question (a) we analyzed the agreement responses of the three critical sentences in Experiment 2 with a logistic regression model with the factor 'Sentence-type' as an orthogonal contrast-coded fixed effect (contrast 1: must = -0.66, certain = 0.33, know = 0.33; contrast 2: must = 0, certain = -0.5, know = 0.5). Agreement rates didn't differ between 'must' and 'certain' and 'know' (β= -0.6177, z = -1.222, p = 0.222). To answer question (b), we analyzed the agreement responses of the three critical sentences in Experiment 1 and Experiment 2 combined with a logistic regression model with the factor 'Sentence-type' as an orthogonal contrast-coded fixed effect and the factor 'Experiment' as an effects-coded fixed effect (Experiment 1 = -0.5; Experiment 2 = 0.5). There was a main effect of the 'Experiment' factor such that the rates of agreement in Experiment 2 were significantly lower than in Experiment 1 (β = -1.160, z = -3.180, p < 0.01); more specifically, after conducting pairwise comparisons, we found that the rates of 'must' endorsement were lower in Experiment 2 than 1 (β = 1.734, z = 3.533, p < 0.001) suggesting

that the within-subjects presentation of sentences reduced participants' agreement with the 'must' statement.

Discussion

In Experiment 2, we changed Lassiter's original one-sentence-task to a multiple-sentences task where each participants judged the three experimental statements ('Bill must not have won the raffle', 'It is certain that Bill did not win the raffle' and 'We know that Bill did not win the raffle') while these were simultaneously displayed on the screen together with some clearly true - which included 'It is highly probable that Bill did not win the lottery - and clearly false baseline statements. By making these changes, we aimed to prompt participants to assign sentences to two groups, the 'Agree' group including sentences like the clearly true baselines or the 'Disagree' group including sentences like the clearly false baselines. We found that the agreement rate for 'must' did not significantly differ from the agreement rate for 'certain' and 'know' and patterned with the agreement rate of clearly false sentences, whereas 'probable' and 'chance' patterned with clearly true statements as expected.

We take these findings as suggesting that for English speakers to judge an event as highly probable in a given scenario is not a sufficient condition for endorsing 'must p' in that same scenario, hence falsifying the probabilistic must hypothesis. However, we argue that these findings don't discriminate between the hyperbolic logical must hypothesis and the evidential hypothesis. In fact, both hypotheses can account for the data. The hyperbolic logical hypothesis would explain the results by assuming that participants converged on the literal meaning of 'must p' (p is a necessary event) and, consequently, converged on rating 'Bill must not have won the raffle' as

false based on the contextual information (it is false that it is certain that Bill did not win the lottery given the contextual information). The evidential hypothesis would explain the results by assuming that participants converged on judging the contextual information not sufficient to conclude that Bill did not win the lottery. Therefore, we take our findings from Experiment 2 as compatible with both the hyperbolic logical must hypothesis and the evidential must hypothesis.

So, how do we discriminate between the hyperbolic logical must and the evidential must hypotheses? We started by examining more closely the nature of the scenario designed by Lassiter (2016), which we kept constant across Experiment 1 and 2. Recall that the story was designed such that the probability of the event of Bill having won the lottery is known and very small (1 chance out of 1000), based on which one may be induced to conclude that he did not win the lottery. But what type of inference would one be performing in this case? This is an instance of what philosophers call 'inductive inference', namely a non-logically valid inferential pattern "based purely on statistical data, such as observed frequencies of occurrences of a particular feature in a given population" (Douven, 2017). A common practice in the philosophical literature is to distinguish within the class of non-logically valid inferences the inductive type from the abductive type (also known as 'inference to the best explanation'; cf. Harman 1965): "in abduction there is an implicit or explicit appeal to explanatory considerations, whereas in induction there is not; in induction, there is only an appeal to observed frequencies or statistics" (Douven, 2017). A good example of abductive inference is the following: "You happen to know that Tim and Harry have recently had a terrible row that ended their friendship. Now someone tells you that she just saw Tim and Harry jogging together. The best explanation for this that you can think of is that they made up. You conclude that they are friends again" (Douven, 2017). Thus, abductive conclusions and inductive conclusions are alike in being obtained through non-logically valid inferential

strategies, but they crucially differ in whether the conclusion is triggered by the need to explain some other known fact.

Now, assuming these considerations, we can restate the finding of experiment 2 as follows: English speakers converge on not agreeing with 'must p' in a context where 'p' is the conclusion of an inductive inference, i.e., a non-logical conclusion based on pure statistical reasons. Consequently, we can also restate the interpretations of these findings by the hyperbolic logical hypothesis and the evidential hypothesis: the hyperbolic hypothesis would say that participants converged on judging the conclusion of the suggested inductive inference as not certain, whereas the evidential hypothesis would say that participants converged in not drawing the suggested inductive inference, i.e. in not forming the belief that the event statistically supported by the evidence actually occurred. So, we asked ourselves: would speakers also converge on not believing that 'p' if 'p' is the conclusion of an abductive inference i.e., a non-logical inference suggested by the given evidence which provides the best explanation for that evidence?

Interestingly, concerning the answer to this question, there is extensive experimental evidence suggesting the existence of an explanation effect such that "an explanation's quality is used as a guide to the probability of that explanation" (Lombrozo 2012: 15; see also Ross et al. 1977; Anderson, Lepper & Ross, 1980; Chapman & Chapman, 1967, 1969; Koehler 1991; Brem and Rips 2000; Lombrozo 2007). For instance, Lombrozo (2007) shows that the degree of simplicity of an explanation is used as a cue for its likelihood: when participants learned about a patient with two symptoms, they overwhelmingly considered more likely that those two symptoms were caused by a single disease (simple explanation) rather than by the conjunction of two diseases (complex explanation) in the absence of base rates for the diseases. Other properties of explanations that have been shown to increase their estimated likelihood are their breath,

coherence, consistency with prior knowledge (see Lombrozo 2012 for an overview of this literature). Overall, these findings suggest that more people may jump to the conclusion that an event happened when that event represents a very good (simple, coherent, fully consistent with prior biases, etc.) explanation for some other event than when the same event is supported only by probabilistic reasoning, even if the latter is quite robust.

Based on these considerations, we expect that participants will endorse a statement like 'I conclude that p' more in a context where 'p' represents the best explanation for some other fact than in a context where the same 'p' is just a probabilistically supported conclusion. And they will do that even if they converge in judging the event as not certain in the explanatory context. Now, based on these expectations, we take the two hypotheses - hyperbolic logical must and evidential must - as differing in terms of their predictions in a task comparing an inductive-type context to an explanatory-type context. The hyperbolic logical must predicts that even in the explanatory context participants will agree with 'must p' as much as with 'it is certain p', no matter whether their agreement rate with 'I conclude that p' is higher: in the explanatory as well as in the inductive context 'p' is a non-certain event. Instead, the evidential hypothesis predicts that participants' endorsement rate of 'must p' will go up in the explanatory context and align with the agreement rate of 'I conclude that p'. These considerations led us to design Experiment 3.

4.1.4 Experiment 3: Best explanations vs inductive conclusions in English

In Experiment 3, we aimed at assessing two potential accounts of the finding in Experiment 2 that overwhelmingly participants didn't endorse 'must p' in a context where 'p' describes a event whose occurrence is very likely based on a pure statistical computation: according to the hyperbolic logical must hypothesis, participants behaved so because such an event is not certain, whereas, according to the evidential hypothesis, participants behaved so because a statistically strong support for an event is not enough to make them conclude (i.e. form the belief) that such an event happened. To discriminate between these two accounts, we compared the original context of Experiment 2 defining 'Bill did not win the raffle' as a statistically well-supported but not certain conclusion ('inductive context'; see 12a) to a context defining the same conclusion as statistically well-supported conclusion and a very good explanation for a fact described in the story but still not certain ('explanatory context'; see 12b). Moreover, besides the sentences 'Bill must not have won the lottery', 'It is certain that Bill did not win the lottery', and 'It is highly probable that Bill did not win the lottery', we added among the critical sentences seen by participants 'I conclude that Bill did not win the lottery' which would allow us to keep track of participants' willingness to form the belief that 'p' based on the context.

We expect that in the inductive context most people will disagree with both 'Bill must not have won the lottery' and 'It is certain that Bill did not win the lottery' (based on the results of Experiment 2) and also with 'I conclude that Bill did not win the lottery' (based on our guess), whereas in the explanatory contexts many more people will agree with 'I conclude that Bill did not win the lottery' than with 'It is certain that Bill did not win the lottery' (based on the existence of an explanation effect). So, under this expectations, this task would allow us to detect whether in the explanatory context the endorsement rate of 'Bill must not have won the lottery' stays

aligned to the endorsement rate of 'It is certain that Bill did not win the lottery' - as the hyperbolic logical hypothesis would predict - or goes up together with the endorsement rate of 'I conclude that Bill did not win the lottery' - as the evidential hypothesis would predict. For completeness, we also tested whether across the two contexts the endorsement rate of 'Bill must not have won the lottery' aligns to the endorsement rate of 'It is highly probable that Bill did not win the lottery' - as the probabilistic hypothesis would predict.

Participants and methods

We adopted a 2x4 mixed-design by crossing the factor 'Context' (2 levels: 'Inductive', 'Explanatory'; see 12) manipulated between-subjects and the factor 'Sentence' (8 levels: 'Must', 'Conclude', 'Certain', 'Probable', 'One', 'X1000', 'Two', 'Winner'; see 13) manipulated within-subjects as in Experiment 2. First, each participant saw either Lassiter's original context defining 'Bill did not win the raffle' as a highly plausible but not certain conclusion based on a mere probabilistic reasoning (see 90a) or a context defining the same conclusion as a very good explanation for a fact described in the story but still not certain (see 90b). Next, each participant rated saw all eight sentences in (91) simultaneously on the screen in a random order and chose between the response options 'Agree' and 'Disagree' for each of the sentences. Responses to 'Must', 'Conclude', 'Certain', and 'Probable' are critical and the clearly true and false baselines are there to provide sanity checks: we analyzed only the data from participants who rated as expected all four clearly true and clearly false baselines.

(105)

The two stories read by participants (between-conditions) in Experiment 3

a. L16's context   ('Inductive')   *suggesting a conclusion supported*

                                                     by a probabilistic computation

Yesterday, Bill bought a single ticket in a raffle with 1000 total tickets. There were also 999 other people who bought one ticket each. That is, the tickets were distributed like this: People holding one ticket: Bill, Mary, Jane, … [997 more]. The drawing was held last night, and the winner will be announced this evening.

b. Modified version   ('Explanatory')   *suggesting a simple, coherent, and*

                                                     consistent with prior biases

                                                     explanatory conclusion

Yesterday, Bill bought a single ticket in a raffle with 1000 total tickets. There were also 999 other people who bought one ticket each. That is, the tickets were distributed like this: People holding one ticket: Bill, Mary, Jane, … [997 more]. The drawing was held last night. **Today, you meet Bill, and he looks a little bit disappointed**.

(106)

        <u>Experimental sentences</u>

        a. Bill must not have won the raffle.   (Must)

        b. I conclude that Bill did not win the raffle.   (Conclude)

        c. It is certain that Bill did not win the raffle.   (Certain)

        d. It is highly probable that Bill did not win the raffle.   (Probable)

        <u>Clearly true baselines</u>

        e. Bill bought exactly one ticket in the raffle.   (One)

        f. 1000 different people bought one lottery ticket each in the raffle.   (1000)

        <u>Clearly false baselines</u>

        g. Mary bought two tickets in the raffle.   (Two)

        h. The winner will be announced tomorrow.   (Winner)

We recruited 140 participants on Prolific aiming at having at least 120 good participants (we assumed that 10% will make an error on one of the four sanity check sentences). We divided participants in two groups of 70 and assigned them to one of the two context conditions. Again, as in the previous two experiments, the dependent measure was the proportion of 'Agree' choices for each sentence.

Predictions

We expect that the endorsement rate of 'Certain' and 'Probable' will respectively be close to floor and close to ceiling in both contexts (based on our intuition) and that the endorsement rate of 'Conclude' will be much higher in the explanatory than in the inductive condition and much higher than that of 'Certain' in the explanatory condition (assuming the existence of an explanation effect). Assuming these general expectations, the three hypotheses make the following predictions about the agreement rate of 'Must':

- the hyperbolic logical must hypothesis predicts that it will pattern with that of 'Certain' in both the inductive and the explanatory context.

- the probabilistic must hypothesis predicts that it will pattern with that of 'Probable' in both the inductive and the explanatory context.

- the evidential must hypothesis predicts that it will pattern with that of 'Conclude' in both the inductive and the explanatory context.

Results

We excluded 8 participants from analysis because they didn't rate all the baselines as expected, which left us with 132 participants (67 in the Inductive condition and 65 in the Explanatory condition) The mean agreement rates for the critical sentences are reported in Table 7. The mean agreement ratings of the critical sentences are plotted in Figure 3.

155

Inference	Sentence Type	N	Mean	Sd
Inductive	Must	67	0.21	0.41
	Conclude	67	0.37	0.49
	Certain	67	0.09	0.29
	Probable	67	0.97	0.17
Explanatory	Must	65	0.82	0.39
	Conclude	65	0.89	0.31
	Certain	65	0.40	0.49
	Probable	65	0.97	0.17

Table 7. Mean agreement rate, standard deviation for each of the four experimental sentences in Exp 3.
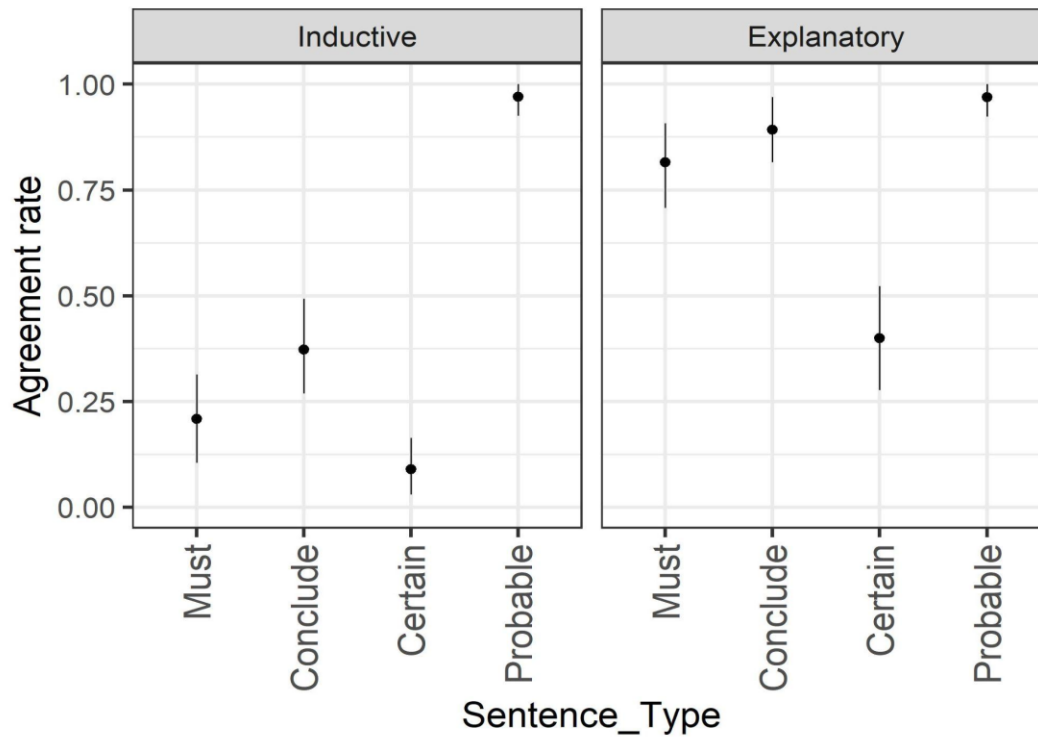

Fig. 3   Mean ratings from participants who rated correctly the baselines in Experiment 3

for 'Must', 'Conclude', 'Certain', and 'Probable'.

We analyzed the agreement responses of the four critical sentences with a logistic mixed-effects regression model (using the glmer function from the lme4 package). The model included the factor 'Sentence' ('Must', 'Conclude', 'Certain', 'Probable',) as a dummy-coded predictor (with 'must' as the reference level), the factor 'Context' as an effects-coded predictor (Inductive = -0.5; Explanatory = 0.5), their interaction, and random intercepts for participants. We followed-up this analysis with pairwise comparisons of the four sentences across the two contexts using the r function emmeans.

First, we asked whether our manipulation of the contextual information was successful. Indeed, we found a main effect of context type such that overall, the agreement rates of the sentences in the explanatory context were significantly higher than in the inductive context ($\beta =$ 6.59, $z = 3.85$, $p < 0.001$), which suggests that the explanatory context induced higher agreement rates overall, supporting the existence of an explanation effect.

Next, we asked whether our expectations about the agreement rates of 'Probable', 'Certain', and 'Conclude' were confirmed. Concerning 'Probable', we observed that its agreement rate was exactly as we expected, namely very close to the ceiling across the two contexts (0.97 in both contexts). Concerning 'Certain', while in the inductive condition its agreement rate was close to floor as expected (0.09), in the explanatory condition it was much higher than expected (0.41), which, we think, was due to an exaggerated behavior of participants triggered by the explanatory effect. Concerning 'Conclude' on its own, we observed that indeed its agreement rate was much higher in the explanatory (0.89) than in the inductive (0.37) condition (pairwise comparison 'Conclude inductive' vs 'Conclude explanatory': $\beta = -6.20$), which suggests that the "best explanation nature" of the event of Bill not winning the lottery prompted more participants to conclude that such an event happened than its objective high probability alone. Concerning

'Conclude' in relation to 'Certain' and 'Probable', we observed that in the inductive condition the agreement rate of 'Conclude' was much closer to that of 'Certain' (pairwise comparison 'Certain inductive' vs 'Conclude inductive': β = -3.59) than to that of 'Probable' (pairwise comparison 'Conclude inductive' vs 'Probable inductive': β = -7.19), whereas in the explanatory condition this patterned was reversed with the agreement rate of 'Conclude' much closer to that of 'Probable' (pairwise comparison 'Conclude explanatory' vs 'Probable explanatory': β = -2.62) than to that of 'Certain' (pairwise comparison 'Certain explanatory' vs 'Conclude explanatory': β = -5.79). In our opinion, this asymmetry of the relation among 'Conclude', 'Certain', and 'Probable' across the two contexts suggests that whether participants are willing to conclude that an event happened is not strictly linked to their assessment of the objective probability of such an event, which support the findings in the psychology of decision making that people form a subjective probability of an event beside its objective probability.

Lastly, we investigated the agreement rates of 'Must' to assess the predictions of the three hypotheses. Concerning 'Must' on its own, we observed that its agreement rate was much higher in the explanatory (0.82) than in the inductive (0.21) condition (pairwise comparison 'Must inductive' vs 'Must explanatory': β = -6.59), which suggests that people's endorsement of 'must p' - like that of 'Certain p' and 'I conclude that p' - is sensitive to the explanation effect. Concerning 'Must' in relation to the other three sentences, first we found that, in terms of statistical significance, the agreement rate of 'Must' was different from that of each of the other three sentences on average across the two contexts: higher than that of 'Certain' (β = -3.23, z = -3.954, p < 0.001), lower than that of 'Probable' (β = 6.36, z = 5.10, p < 0.001), and lower than that of 'Conclude' (β = 1.46, z = 2.81, p < 0.01). However, in terms of effect size, the agreement rate of 'Must' was much less different from that of 'Conclude' than from that of 'Certain' and 'Probable',

as it can be noticed by a comparison of the beta coefficients (Conclude' vs 'Must': $\beta = 1.46$; 'Certain' vs 'Must': $\beta = -3.23$; 'Probable' vs 'Must': $\beta = 6.36$). Next, we found: an interaction effect between the agreement rates of 'Must' and 'Certain' across the the two contexts such that participants agreed with 'Certain' less than with 'Must' at a statistically larger rate in the explanatory than in the inductive context ($\beta = -2.59$, $z = -2.09$, $p < 0.05$); an interaction effect between the agreement rates of 'Must' and 'Probable' across the the two contexts such that participants agreed with 'Probable' more than with 'Must' at a statistically larger rate in the inductive than in the explanatory context ($\beta = -4.97$, $z = -3.01$, $p < 0.01$); no interaction effect between the agreement rates of 'Must' and 'Conclude' across the two contexts such that participants agreed with 'Conclude' more than with 'Must' at a statistically not different rate in the inductive and in the explanatory context ($\beta = -0.39$, $z = -0.42$, $p = 0.68$).

By inspecting the results of the pairwise comparisons, we observed that: the estimated effect size of how much participants agreed with 'Must' less than with 'Conclude' were very similar in the Inductive condition ($\beta = -1.66$) and in the explanatory condition ($\beta = -1.26$); like in the case of 'Conclude', in the inductive condition the agreement rate of 'Must' was much closer to that of 'Certain' (pairwise comparison 'Certain inductive' vs 'Must inductive': $\beta = 1.94$) than to that of 'Probable' (pairwise comparison 'Must inductive' vs 'Probable inductive': $\beta = -8.84$), whereas in the explanatory condition this patterned was reversed with the agreement rate of 'Must' much closer to that of 'Probable' (pairwise comparison 'Must explanatory' vs 'Probable explanatory': $\beta = -3.88$) than to that of 'Certain' (pairwise comparison 'Certain explanatory' vs 'Must explanatory': $\beta = 4.53$). Overall, these findings suggest that participants' agreement rate with 'Must' aligned to that of 'Conclude' in the way both rates differed from those of 'Certain' and 'Probable' across the two contexts. Overall, these findings suggest that, although the

agreement rate of 'Must' is statistically lower than that of 'Conclude', yet the two agreement rates

do form a pattern in both contexts and differ similarly from those of 'Certain' and 'Probable' across

the two contexts.

Discussion

In Experiment 3, participants were asked to decide whether they agree with each of the

four sentences 'Bill must not have won the lottery' ('Must'), 'I conclude that Bill did not win the

lottery ('Conclude'), 'It is certain that Bill did not win the lottery' ('Certain'), and 'It is highly

probable that Bill did not win the lottery' ('Probable') plus 2 clearly true baselines and 2 clearly

false baselines given one of these two contexts: either Lassiter's original context defining the

proposition 'Bill did not win the lottery' as a non-logical conclusion derived through a

probabilistic computation (inductive context) or a context defining the same proposition as a

non-logical conclusion but a very plausible explanation for some of the described facts

(explanatory context).

We found that participants agreed with 'Conclude' at a much higher rate in the

explanatory than in the inductive context, suggesting that a highly probable event which also

explains well some other event is believed more than a highly probable event just based on a

statistical computation. The endorsement rate of 'Must' was: (i) much lower than that of

'Probable' and close to that of 'Certain' in the inductive context; (ii) much higher than that of

'Certain' and closer to that of 'Probable' in the explanatory context; (iii) similar to (although

slightly lower than) that of 'Conclude' in both contexts. We take these findings as suggesting

that speakers' endorsement of 'must p' in a context is more affected by whether they would say

that they conclude that 'p' in that context than whether they would say that 'p' is certain or

160

probable in that context. More specifically, these findings suggest that speakers say 'must p' as well as 'I conclude that p' not whenever they judge 'p' as objectively highly probable and not only when they judge 'p' as objectively certain but whenever their subjective probability of 'p' passes the threshold above which they would conclude that 'p', i.e., adopt 'p' as one of their beliefs. Our findings show that one of the factors that play a role in the computation of such a subjective probability threshold of an event is whether, in addition to being objectively highly probable, this event is also the most plausible explanation for another event whose occurrence would be hard to motivate otherwise. Overall, these findings support the evidential hypothesis over the hyperbolic logical hypothesis and the probabilistic hypothesis.

Interestingly, we found that the endorsement rate of 'Certain' also was higher in the explanatory than in the inductive condition, suggesting that even the computation of the objective certainty of an event is sensitive to an explanation effect: for some people, a conclusion that represents a highly probable, good, and simple, explanation of some other facts is certain.

So far, we have only considered the English 'must', which raises the question as to how generalizable are our findings to other languages: is a general property of necessity auxiliary verbs that they are used as inferential evidentials? To start answering this question, we attempted to reproduce the findings of experiment 3 in an experiment with Italian speakers, which we describe in the following section.

4.1.5 Experiment 4: Best explanations vs inductive conclusions in Italian.

In Experiment 4 we attempted to reproduce the findings of Experiment 3 in Italian. Notice that in Experiment 3, we could not compare epistemic 'must' to any inferential evidential because there is no such type of morpheme independently identified in English; so, we compared

the endorsement rate of 'must p' to that of 'I conclude that p', which closely approximates the meaning of 'must p' under the evidential hypothesis. In Italian, instead, as I have argued in chapter 2, there exists one morpheme that behaves like an inferential evidential, i.e., the future morphology in its non-temporal uses, shortly 'INFER-p'. So, Italian offers the chance to directly compare the endorsement rate of a sentence with a necessity auxiliary verb like 'deve p' to a sentence with an inferential morpheme like 'p-INFER'.

Participants and methods

We adopted the contexts from Experiment 3 and translated them into Italian. We changed the name of the protagonist of the story from Bill to Gianni.

(107)

     The two stories read by participants (between-conditions) in Experiment 4

        L16's context   ('Inductive')   *suggesting a conclusion supported*

                                                                        by a probabilistic computation

        Ieri Gianni ha comprato un biglietto di una lotteria comprendente 1000 biglietti in tutto. Altre 999 persone hanno comprato un biglietto ciascuna. Quindi, i biglietti sono distribuiti come segue: le persone con un biglietto sono: Gianni, Maria, Sandra, ... [altri 997]. L'estrazione dei biglietti è stata effettuata ieri e il vincitore verrà annunciato stasera.

| Modified version | ('Explanatory') | *suggesting a simple, coherent, and consistent with prior biases explanatory conclusion* |

Ieri Gianni ha comprato un biglietto di una lotteria comprendente 1000 biglietti in tutto. Altre 999 persone hanno comprato un biglietto ciascuna. Quindi, i biglietti sono distribuiti come segue: le persone con un biglietto sono: Gianni, Maria, Sandra, ... [altri 997]. Il vincitore è stato annunciato ieri. Oggi, ti capita di incontrare Gianni che sembra deluso.

We also translated the eight sentences seen by participants in experiment 3 and added the critical condition containing the inferential future 'Gianni non avra' vinto alla lotteria' ('John will not have won the lottery'; labeled as 'Future'). The full list of sentences (nine) that were seen by each participant are listed below.

(108)

Critical sentences

 a. Gianni non deve aver vinto la lotteria   ('Must')

   'Bill must not have won the raffle'

 b. Gianni non avrà vinto la lotteria   ('Future')

   'John will have not won the lottery'

c.   Deduco che Gianni non ha vinto la lotteria   ('Conclude')

         'I conclude that Bill did not win the raffle'

d.   È certo che Gianni non ha vinto la lotteria   ('Certain')

         'It is certain that Bill did not win the raffle'

e.   È altamente probabile che Gianni non ha vinto la lotteria   ('Probable')

         It is highly probable that Bill did not win the rafffle

Clearly true baselines

f.   Gianni ha comprato esattamente un biglietto della lotteria   ('One')

         'Bill bought exactly one ticket in the raffle'

g.   1000 persone hanno comprato ciascuna un biglietto della lotteria   (X1000)

         '1000 different people bought one lottery ticket each in the raffle'

Clearly false baselines

h.   Maria ha comprato due biglietti della lotteria   ('Two')

         'Mary bought two tickets in the raffle'

i.   Il vincitore verra' annunciato la prossima settimana   ('Winner')

         'The winner will be announced next week'

Like in Experiment 3, we adopted a 2x4 mixed-design by crossing the factor 'Context' (2 levels: 'Inductive', 'Explanatory') manipulated between-subjects and the factor 'Sentence' (9 levels: 'Must', 'Future', 'Conclude', 'Certain', 'Probable', 'One', 'X1000', 'Two', 'Winner') manipulated within-subjects. The procedure is as in Experiment 3. First, each participant saw either the inductive or the explanatory context. Next, each participant saw all nine sentences simultaneously on the screen in a random order and chose between the response options 'Agree' and 'Disagree' for each of the sentences. Responses to 'Must', 'Future', 'Conclude', 'Certain', and 'Probable' are critical and the clearly true and false baselines are there to provide sanity checks: we analyzed only the data from participants who rated as expected all four clearly true and clearly false baselines. We recruited 140 participants on Prolific aiming at having at least 120 good participants (we assumed that 10% will make an error on one of the four sanity check sentences). We divided participants in two groups of 70 and assigned them to one of the two context conditions. Again, the dependent measure was the proportion of 'Agree' choices for each sentence.

Predictions

We expect that the endorsement rate of 'Certain' and 'Probable' will respectively be close to floor and close to ceiling in both contexts (based on our intuition) and that the endorsement rate of 'Future' and 'Conclude' will be much higher in the explanatory than in the inductive condition and much higher than that of 'Certain' in the explanatory condition (assuming the existence of an explanation effect). Assuming these general expectations, the three hypotheses make the following predictions about the agreement rate of 'Must':

165

-       the hyperbolic logical must hypothesis predicts that it will pattern with that of 'Certain' in both the inductive and the explanatory context.

-       the probabilistic must hypothesis predicts that it will pattern with that of 'Probable' in both the inductive and the explanatory context.

-       the evidential must hypothesis predicts that it will pattern with that of 'Future' and 'Conclude' in both the inductive and the explanatory context.

Results

We excluded 27 participants from analysis because they didn't rate all the baselines as expected, which left us with 113 participants (54 for the Inductive condition and 59 for the Explanatory condition). The mean agreement rates and standard deviation for the five experimental sentences are reported in Table 8. The agreement ratings of the five experimental sentences in the two context conditions are plotted in Figure 4.

Inference	Sentence Type	N	Mean	Sd
Inductive	Must	54	0.20	0.41
	Future	54	0.22	0.42
	Conclude	54	0.17	0.38
	Certain	54	0.00	0.00
	Probable	54	0.83	0.38
Explanatory	Must	59	0.83	0.38
	Future	59	0.90	0.30
	Conclude	59	0.86	0.35
	Certain	59	0.29	0.46
	Probable	59	0.98	0.13

Table 8. Mean agreement rate and standard deviation for each of the five experimental sentences in Exp 4.
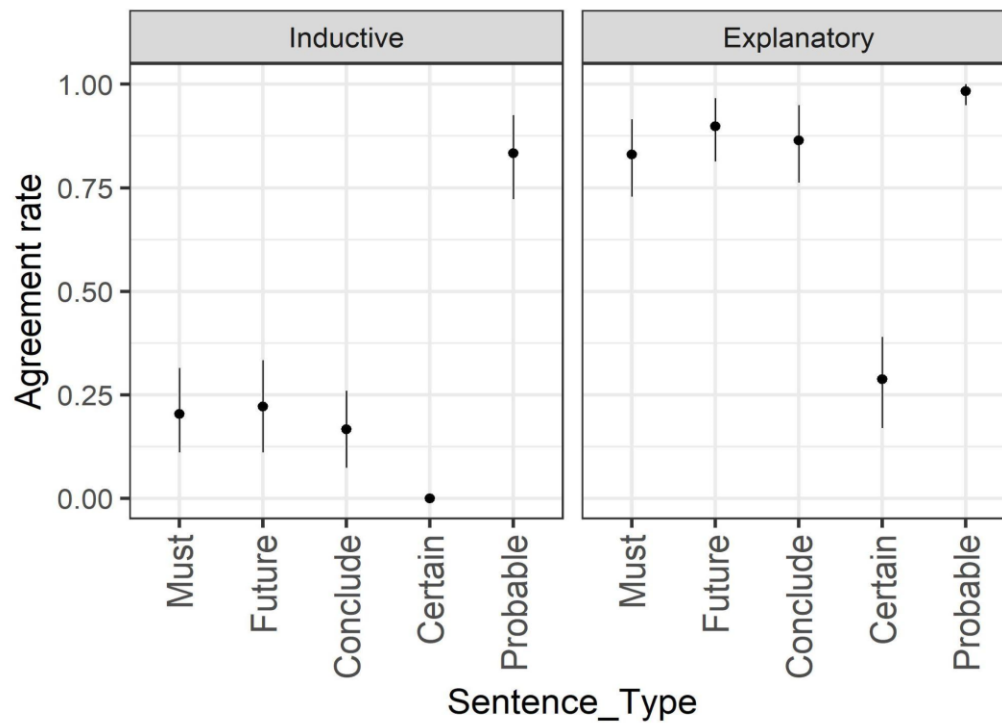

Fig. 4   'Agree' proportions from participants who rated correctly the baselines in Experiment 4 for 'Must', 'Future', 'Conclude', 'Certain', and 'Probable'.

First, we asked whether our expectations about the agreement rates of 'Probable', 'Certain', 'Future, and 'Conclude' were confirmed.

Concerning 'Probable', we observed that its agreement rate was as we expected (very close to the ceiling), in the Explanatory condition (0.98) but not as close to ceiling in the inductive condition (0.83). We don't know why this was the case.

Concerning 'Certain', we observed a similar behavior as in the 'Certain' condition of Experiment 3: while in the inductive condition its agreement rate was at floor as expected (0), in the explanatory condition it was higher than expected (0.29), which, we think, was due to an exaggerated behavior of participants triggered by the explanatory effect, as in the English case.

Concerning 'Future' and 'Conclude', we observed that their agreement rates were numerically very similar in both contexts and patterned in being much higher in the explanatory (Future: 0.90; Conclude: 0.86) than in the inductive condition (Future: 0.22; Conclude: 0.17), which suggests that the "best explanation nature" of the event of Gianni not winning the lottery prompted more participants to conclude that such an event happened than its objective high probability alone. Moreover, concerning 'Future' and 'Conclude' in relation to 'Certain' and 'Probable', we observed that in the inductive condition the agreement rates of 'Future' and 'Conclude' were much closer to that of 'Certain' than to that of 'Probable', whereas in the explanatory condition this patterned was reversed with the agreement rates of 'Future' and 'Conclude' much closer to that of 'Probable' than to that of 'Certain'. So, again like in English, there is an asymmetry between 'Future' and 'Conclude' on one side and 'Certain' and 'Probable' on the other side across the two contexts, which suggests that whether participants are willing to conclude that an event happened is not strictly linked to their assessment of the objective probability of such an event: this supports for Italian speakers as well the claim in the

168

psychology of decision making that people form a subjective probability of an event beside its objective probability.

Lastly, we investigated the agreement rates of 'Must' to assess the predictions of the three hypotheses. We observed that the agreement rate of 'Must' was numerically very similar to that of 'Future' and 'Conclude' across the two contexts, supporting the prediction of the evidential hypothesis: the endorsement rate of 'deve p' is like the endorsement of a statement with an inferential evidential and depends not on the participants assessment of the objective likelihood of 'p' but on the participants willingness to conclude that 'p' based on the provided contextual information

Discussion

In Experiment 4, participants who were Italian native speakers were asked to decide whether they agree with each of the four sentences 'Gianni non deve aver vinto la lotteria' ('John must not have won the lottery'; 'Must') , 'Gianni non avrà vinto la lotteria' ('John will have not won the lottery'; 'Future'), 'Deduco che Gianni non ha vinto la lotteria' ('I deduce that John did not win the lottery'; 'Conclude'), E' certo che Gianni non ha vinto la lotteria' ('It is certain that John did not win the lottery'; 'Certain') and 'E' altamente probabile che Gianni non ha vinto la lotteria' ('It is highly probable that John did not win the lottery'; 'Probable') plus 2 clearly true baselines and 2 clearly false baselines given one of these two contexts: either Lassiter's original context defining the proposition 'Bill did not win the lottery' as a non-logical conclusion derived through a probabilistic computation (inductive context) or a context defining the same proposition as a non-logical conclusion but a very plausible explanation for some of the described facts (explanatory context).

169

We found that participants agreed with 'Must' at a numerically very similar rate as 'Future' and 'Conclude' across the two contexts. Specifically, their agreement rates of the three sentences were: (i) much higher in the explanatory than in the inductive context; (ii) much lower than that of 'Probable' and close to that of 'Certain' in the inductive context; (ii) much higher than that of 'Certain' and closer to that of 'Probable' in the explanatory context. We take these findings as suggesting that Italian speakers' endorsement of 'must p' like that of 'INFER-p' in a context is more affected by whether they would say that they conclude that 'p' in that context than whether they would say that 'p' is certain or probable in that context. More specifically, these findings suggest that Italian speakers tend to say 'must p' as well as 'INFER-p' not whenever they judge 'p' as objectively highly probable and not only when they judge 'p' as objectively certain but when their subjective probability of 'p' passes the threshold above which they would conclude that 'p', i.e., adopt 'p' as one of their beliefs. Overall, these findings support the evidential hypothesis over the hyperbolic logical hypothesis and the probabilistic hypothesis.

4.1.6 General discussion of experimental data

In this section, we have assessed the hyperbolic logical must hypothesis, the probabilistic / weak logical must hypothesis, and the evidential hypothesis in comprehension tasks requiring participants to decide whether they endorse a statement based on the given contextual information. Lassiter (2016) was the first to assess the agreement rate with the English 'must p' in such a type of task across many participants. He found that, given a context where the probability of 'p' is known and very high (99.9%), in English 'must p' is agreed with at a higher rate than 'it is certain that p' and 'we know that p' in a task where each participant evaluates only one of these three sentences. Lassiter takes this finding as suggesting that the probabilistic must hypothesis is supported over the standard modal logical view. We assessed Lassiter's interpretation of his finding through a series of follow-up studies where we compared the probabilistic hypothesis to the other two hypotheses mentioned above. To this end, after having successfully replicated Lassiter (2016)'s findings (Experiment 1) - although with a lower agreement overall - we designed two follow-up studies with English speakers.

First (Experiment 2), we manipulated the factor 'Task design' by comparing the original one-sentence task to a multiple sentence task where each participant evaluated 'must p' together with 'certain p', 'know p', and some clearly true and clearly false sentences. This manipulation was meant to prompt participants to focus on the literal meaning of the statements by implicitly inducing them to compare the target sentences to the clearly true type and the clearly false type. This task would allow us to discriminate to discriminate between the probabilistic hypothesis on one side and the hyperbolical logical must on the other side: the probabilistic hypothesis predicts that the agreement rate of 'must p' would be like that of the clearly true sentences while the hyperbolic logical must hypothesis predicts that it would be like that of the clearly false sentences.

171

We found that the agreement rate for 'must p' did not significantly differ from the agreement rate for 'certain p' and 'know p' and patterned with the agreement rate of clearly false sentences, which support the hyperbolic logical hypothesis over the probabilistic hypothesis. However, these findings are compatible with the evidential hypothesis as well.

So next (Experiment 3), we aimed at discriminating between the hyperbolic logical must hypothesis and the evidential must hypothesis. To this end, we manipulated between-subjects the context by comparing the original scenario where 'p' is an inductive conclusion (i.e., a non-logical conclusion based on pure statistical reasons) to a scenario where 'p' is an abductive conclusion (i.e., a non-logical conclusion that offers the best explanation for some other evidence mentioned in the context). Relying on findings from psychology supporting the claim that people tend to use an explanation's quality as a guide to the probability of that explanation, we expected that this manipulation would prompt participants reading an explanatory contest to conclude that 'p' more than participants reading an inductive context, without affecting much their assessment of the certainty of 'p'. Under these expectations, this task would allow us to discriminate between the hyperbolic logical hypothesis and the evidential hypothesis: the hyperbolic logical hypothesis predicts that the agreement rate with 'must p' will stay close to that of 'certain p' and close to floor across the two contexts, whereas the evidential hypothesis predicts that the agreement rate of 'must p' would be higher in the explanatory than in the inductive condition and not much different than that of a statement like 'I conclude that p'. We found that the agreement rate for 'must p' did not differ much from that of 'I conclude p' in both contexts and was much higher than that of 'certain' in the explanatory condition, which support the evidential hypothesis over the hyperbolic logical hypothesis.

Lastly (Experiment 4), we asked whether the findings of Experiment 3 about the English 'must p' could be replicated in Italian with 'deve p' and 'INFER-p'. To this end, we translated the material of Experiment 3 from English into Italian and added a target sentence exemplifying 'INFER-p'. Exactly as for 'must p' in Experiment 3, we found that the agreement rate for 'deve p' and 'INFER-p' did not differ much from that of 'I conclude p' in both contexts and was much higher than that of 'certain' in the explanatory condition.

Overall, we take these findings as supporting the evidential must hypothesis over the hyperbolic logical hypothesis and the probabilistic hypothesis. The typical epistemic use of 'must' and 'deve' is in line with the use of an inferential like the Italian 'INFER': people's everyday use of 'must p' or 'deve p' as well as of 'INFER-p' depends not on whether they assess the objective likelihood of 'p' as high or certain but on whether they assess the subjective likelihood of 'p' as high enough to prompt them to jump to the conclusion (i.e., form the belief) that 'p' is true.

4.2 Further data corroborating the evidential hypothesis

4.2.1 Epistemic necessity verbs and epistemic adverbs

In chapter 2 section 2.2, I have shown that a grammatical inferential like the Italian inferential future can felicitously co-occur with modal adverbs expressing any degree of epistemic commitment from maximal commitment ('certainly') to minimum commitment ('perhaps').

(109)

   i. Certamente Susanna sarà   malata

     Certainly Susan INFER.be.3pl. sick

     'Certainly, Susan will be sick'

   ii. Probabilmente Susanna sarà   malata

     Probably Susan INFER.be.3pl. sick

     'Probably, Susan will be sick'

   iii. Forse   Susanna sarà   malata

     Perhaps Susan INFER.be.3pl. sick

     'Perhaps, Susan will be sick'

The evidential hypothesis predicts that also epistemic 'must' and epistemic 'dovere' should felicitously co-occur with modal adverbs expressing any degree of epistemic commitment. This prediction is borne out as showed by the following series of naturalistically occurring utterances.

(110)   <u>Must co-occurs with certainly</u>

a. As you **must certainly** be aware, we know the location of your current base[8]

b. They **must certainly** have noticed this because it is so common[9]

c. The Founding Fathers **must certainly** be turning in their graves[10]

(111)   <u>Must co-occurs with probably</u>

a.   After five decades, engineering management is a still evolving discipline. But then, it **probably must** always be that way[11]

b.   What about her students? I mean, they **must probably** think she's so smart and love her and look up to her[12].

c.   Obama exclaimed to cheers from the crowd. "If some of these folks were around when Columbus set sail, they **probably must** have been founding members of the flat earth society. They would not believe that the world was round![13]

(112)   <u>Must co-occurs with maybe/perhaps</u>

---

[8] Source: Series Transformers Prime (IMDB) (Years: 2010–2013: 59 episodes) Episode: Persuasion (2013)

[9] Source: Through the Wormhole (IMDB) (Years: 2010–2017: 57 episodes) Episode: Will We Survive First Contact? (2012)

[10] Source: BLOG   http://spectator.org/archives/2010/07/16/americas-ruling-class-and-the

[11] Source: ACAD: Mechanical Engineering; Date 2005 (Mar); Publication information Mar2005 Management Supplement, p4-6, 3p, 3c; Title: Making Sense of Change.

[12] Source: TV Series Party of Five; Episode: Short Cuts (1996).

[13] Source: WEB http://hotair.com/archives/2012/03/15/obama-on-his-opponents-they-would-have-been-founding-members-of-the-flat-earth-society/; Date 2012; Title: Obama on his opponents: They would have been founding - Hot Air

a. **Maybe** she **must** probably have had a plan B all along[14]

b. But, you know, at some level don't you think **perhaps he must** have felt that he

brought this on himself?[15]

Italian

(113)   <u>Dovere co-occurs with certamente</u>

a. A mio avviso, infatti, qualcosa su quel fronte **deve** essere **certamente** accaduto[16]

'In my opinion, something must have certainly been happened on the that side'

b. La realizzazione **deve** essere **certamente** stata diretta da un mosaicista assai

capace[17]

'The realization must have certainly been directed by a very talented mosaicist'

c. Tutti intorno a lui scappano via, qualcuno **certamente dev**e urlare degli

avvertimenti, visto che anche i due personaggi di lato alla jeep interrompono

repentinamente il loro assalto[18]

---

[14] Source: SPOK: TODAY SHOW 7:00 AM EST; Date: 2015 (150421).

[15] Source: SPOK: CNN_King. Title: An Overview of the Future of the USSR

[16] Source: http://it.wikisource.org/wiki/Commissione_parlamentare_di_inchiesta_sulla_Federconsorzi/Audizioni/20

[17] Source: http://it.wikipedia.org/wiki/Arte_tardoantica

[18] Source: http://archive.globalproject.info/art-245.html

'Eveyone around him are running away, someone must certainly be shouting some warning, because even the two individuals next to the jeep have interrupted their assault'

(114)   Dovere co-occurs with probabilmente

a. Tutti i grandi mammiferi partoriscono un unico cucciolo, una caratteristica che **molto probabilmente** anche l'Indricotherium **deve** aver avuto[19]

'All big mammals deliver one single kid, a feature that very probably even the Indricotherium must have had.'

b. Questa scarsa incidenza dell'utilizzo del rame sulle culture preistorica si **deve probabilmente** spiegare con le difficoltà e gli scarsi benefici di questa nuova tecnica[20]

'This weak influence of the use of copper on the prehistoric civilizations must be probably due to the fact that this new technique was difficult and with few benefits'

c. La revoca della disposizione ad opera di Antonino Pio **deve probabilmente** essere letta come una riconferma dei limiti previsti anche nelle fonti giuridiche[21]

'The withdrawal of the action by Antonino Pio must probably be interpreted as a reconfirmation of the constraint established in the laws'

---

[19] Source: http://it.wikipedia.org/wiki/Indricotherium

[20] Source: http://it.wikipedia.org/wiki/Et%C3%A0_del_rame

[21] Source: http://it.wikipedia.org/wiki/Publio_Elio_Traiano_Adriano

(115)   Dovere co-occurs with forse

    a. David Gandy come uno dio minore di tradizione classica si offre al mare e ai suoi

        misteri. Questo **deve aver forse** pensato Greg Lotus mentre preparava il suo

        servizio fotografico per GQ Germania[22]

        'David Gandy like a minor god of classical tradition offers himself to the sea.

        Perhaps this must have thought Greg Lotus while he was preparing his photo

        shooting for the German GQ'

    b. Sì, ma siccome **forse** la mamma Polandia non mi **deve** aver sentito abbastanza

        riconoscente, c'è dell'altro[23]

        'Since the mom Polandia maybe must have felt that I was not enough grateful,

        there is something more'

    c. **Forse** allora **deve** aver avvertito la sagoma della carabina che si formava nelle

        mie pupille[24]

        'Maybe they have felt the outline of the gun that took shape in my eyes'

4.2.2 Epistemic necessity verbs under embedding predicates

In chapter 2, section 2.3, I showed that a clause with the Italian inferential future is syntactically embeddable under a doxastic verb or an epistemic impersonal construction but it doesn't semantically embed under the scope of the embedding predicate.

(116)   INFER under a doxastic verb

---

[22] Source: http://www.fashionblog.it/tag/David+Gandy

[23] Source: http://zulawskiego5.splinder.com/archive/2008-06

[24] Source: http://zulawskiego5.splinder.com/tag/tate

Maria pensa che Susanna sarà malata

'Mary thinks that Susan will be sick'

Interpretation:

=   'Mary thinks that Susan is sick, which Mary has inferred

≠   Mary thinks that she has acquired through an inference that Susan is sick

(117)   <u>INFER under an impersonal verb</u>

È probabile/evidente che Susanna sarà malata

'It is probable/evident that Susan will be sick'

Interpretation:

= 'it is probable/evident that Susan is sick, which the speaker has inferred'

≠ It is probable/evident that the speaker has concluded that Susan is sick

Both the epistemic 'must' and 'dovere' can be syntactically embedded under doxastic verbs and epistemic impersonal constructions.

(118)   <u>'must'/'deve' under a doxastic verb</u>

Maria pensa che Susanna deve essere malata

'Mary thinks that Susan must be sick'

(119)   <u>'must' / 'deve' under an impersonal verb</u>

È probabile/evidente che Susanna deve essere malata

'It is probable/evident that Susan must be sick'

But are 'must' and 'dovere' under the semantic scope of the embedding predicates? Westmoreland (1998) makes this relevant remark: "The word 'must' is invisible to the compositional process that constructs the propositional content of a sentence when it is embedded under a propositional attitude predicate; the information it contributes regarding the source of a proposition comes about by an independent process" (Westmoreland 1998: 80). In other words, according to Westmoreland the word 'must' under a verb like 'think' or 'evident' behaves as a not-at-issue expression whose meaning is on an independent secondary dimension with respect to the at-issue compositional meaning dimension. And this extra-meaning is the very meaning of a grammatical inferential. My intuition in Italian aligns to Westmoreland's.

4.2.3 Epistemic necessity verbs and antecedents of conditionals

In chapter 2, section 2.3, I showed that a clause with the Italian inferential future cannot felicitously occur in the antecedent of a conditional clause.

(120)   #Se Susanna sarà   malata , allora la   andiamo a visitare

   If   Susan   INFER.be3sg   sick   then   her we-go   to visit

   'If Susan will be sick, then we'll go to visit her'

I suggested that this unacceptability has to do with the fact that the antecedent of an if-clause does not have a judge. When a conditional clause is produced neither the speaker nor any other contextually defined agent is committed to have an epistemic stance towards the at-issue proposition expressed by the antecedent clause. When then, the inferential future points to the judge of the at-issue proposition of the antecedent clause, there is no such judge, hence yielding a semantic clash.

Now, it is often observed that epistemic 'must' resists being embedded in the antecedent of a conditional clause.

(121)   #If Max must be lonely, his wife will be worried.

(example from Papafragou 2006: 1690)

Similarly in Italian, epistemic 'dovere' doesn't feel acceptable in the antecedents of conditionals.

(122)   #Se Susanna   deve essere   malata , allora la   andiamo a visitare

          If Susan   must be   sick   then her we-go   to visit

          'If Susan must be sick, then we'll go to visit her'

In this respect, 'dovere' differs from 'è certo che' ('it is certain that') which sounds fine in the antecedent of a conditional.

(123)   Se è certo che Susanna   è   malata , allora la   andiamo a visitare

          If   is certain that Susan   is   sick   then her we-go   to visit

          'If it is certain that Susan is sick, then we'll go to visit her'

The unacceptability of 'must' and 'dovere' in the antecedent of conditionals is easily explained if one assumes that they are grammatical inferentials.

4.3 Data supporting the existence of the original meaning of epistemic necessity verbs

4.3.1 Truth-evaluation of statements with epistemic necessity verbs

So far, I have offered evidence suggesting that in daily conversation the epistemic necessity verbs behave more like an inferential marker (e.g., the Italian inferential future) than like a necessity epistemic operator. This means, for example, that in Italian an utterance of epistemic 'deve p' is more similar to an utterance of INFER-p than to an utterance of 'certo che p'.

(124)   Epistemic necessity verb

Susanna deve essere nel suo ufficio. Vedo la luce accesa

'Susan must be in her office. I see the light on'.

(125)   Inferential evidential

Susanna sarà nel suo ufficio. Vedo la luce accesa.

'Susan will be in her office. I see the light on.'

(126)   Certainty expressions

È certo che Susanna è nel suo ufficio. Vedo la luce accesa.

'It is certain that Susan is in her office. I see the light on.'

However, when one tries to reply to a 'must-p' utterance with 'it is not true that must-p' the result is more similar to 'it is not true that it is certain p' than to 'it is not true that INFER-p'. Recall from chapter 2, section 2.5 that one cannot say 'it is not true that INFER-p' without being committed to the contradictory message that they judge 'p' as false and that they have concluded that 'p'.

(127)   #Non e' vero che   Susanna   sarà   malata

Not is   true   that   Susan   INFER.be.3sg sick

'It is not true that Susan will be sick'

Interpretation:

= ?!It is not true that Susan is sick, which the speaker has inferred

Now, assuming this observation about the evidential future, imagine a conversation between Gianni and Claudia where Gianni utters 'Susan will be in her office. I see the light on' and Claudia thinks that Gianni's conclusion is not warranted because it is possible that Susan is not in her office even though the light is on (she may have forgotten to turn it off). Claudia cannot felicitously express her concern with 'it is not true that because the light in Susan's office is on, she will be inside'.

(128)   Gianni: Susanna **sarà** nel suo ufficio. Vedo la luce accesa.

'Susan will be in her office. I see the light on.'

Claudia: #Non è vero che siccome la luce è accesa nell'ufficio di Susanna,

lei sarà dentro. Può averla dimenticata accesa.

'It's not true that because the light is on in Susan's office, she will

be inside. She might have forgotten to turn it off'

Interpretation:

= ?!It is not true that because the light is on in Susan's office, Susan is inside

which (that Susan is inside) its current judge (Claudia) has inferred

However, if we imagine a similar conversation with the difference that Gianni utters 'deve' instead of 'INFER' now Claudia's reply is acceptable ('it is not true that deve p'), showing that the word 'must' can be at-issue and not being anchored to the original speaker. In fact, 'it is not true that must p' behaves like 'it is not true that it is certain that p'.

(129)   Gianni: Susanna **deve** essere nel suo ufficio. Vedo la luce accesa.

    'Susan must be in her office. I see the light on.'

    Claudia: Non è vero che siccome la luce è accesa nell'ufficio di Susanna, lei deve esservi dentro. Può averla dimenticata accesa.

    'It's not true that because the light is on in Susan's office, she must be inside. She might have forgotten to turn it off'

(130)   Gianni: **È cert**o che Susanna è nel suo ufficio. Vedo la luce accesa.

    'It is certain that Susan is in her office. I see the light on.'

    Claudia: Non è vero che siccome la luce è accesa nell'ufficio di Susanna, è certo che lei è dentro. Può averla dimenticata accesa

    'It's not true that because the light is on in Susan's office, it is certain that she is inside. She might have forgotten to turn it off'

What does this behavior of epistemic 'deve' under 'it is not true that' reveal? I argue that this evidence reveals the original meaning of the word 'dovere' which is still actively encoded as one of the senses of the word in the mind of at least one Italian speaker (the author). However, this evidence does not imply that the original speaker Gianni used the word 'dovere' as an epistemic

necessity operator. In fact, it is likely that in those circumstances a speaker is using 'deve' as an inferential operator. If I see that the light is on in Susan's office, I also would say 'Susanna deve essere nel suo ufficio' ('Susan must be in her office') but at the same time I'm ready to admit that once you ask me "Is it true that Susan must be in her office, given that the light is on in her office?" I'm induced to re-interpret the word with its original meaning as an at-issue necessity operator. As soon as I examine the truth of what I just said, the original meaning of 'must' comes first to my mind. But I speculate that in this case my feelings are no different than the feelings I have in any other case where a word is polysemous between an original more specific meaning and an extended more relaxed meaning like in the case of 'exponentially'. Consider this dialogue involving the word 'exponentially'.

(131)   Gianni: Le vendite delle macchine è cresciuta **esponenzialmente**.

'The sales of cars have been growing exponentially'

Claudia: Non è vero che è cresciuta esponenzialmente, non c' è una funzione esponenziale che può descrivere la crescita.

'It's not true that the sales has grown exponentially, there is no exponential function that can describe the growth'

Is Claudia justified to reply like that to Gianni's utterance? Perhaps the most natural reaction is to say that Claudia is being prescriptive here: she is denying that the word 'exponentially' is correctly used in its extended sense. I would say that a similar attitude would be adopted by whoever labels as incorrect a 'must p because A' statement by replying 'it is not true that because A, MUST p': by saying so, one is being prescriptive and denying that 'must' is correctly used when the inference is not logical.

4.3.2 Epistemic necessity verbs and sentential negation

In section 2.3, I showed that the Italian inferential future, when it co-occurs with a sentential negation, can never be under the semantic scope of negation.

(132)   Susanna **non sarà**   malata

Susan   **not INFER**.be.3sg sick

'Susan will not be sick'

= The speaker has acquired through an inference that Susan is not sick

(INFER>not)

≠ The speaker has not acquired through an inference that Susan is sick

(not>INFER)

Recall that this observation is compatible with both the possibility that INFER is not-at-issue and the possibility that INFER is simply syntactically higher than negation. What do we know about the relation between epistemic necessity verbs and the negation particle 'not'? Let's consider the English 'must' and the Italian 'dovere' separately.

In English all modal auxiliaries always precede negation in terms of linear order and 'must' is no exception.

(133)   Susan must not be home

Susan may not be home

Susan should be home

However, when it comes to the interpretation of the semantic scope, auxiliaries differ among each other. Focusing on 'must', it never semantically scopes under 'not', no matter whether the interpretation of 'must' is deontic or epistemic. Thus, the deontic interpretation of 'Susan must not be home' can only be as in 'It is obligatory for Susan to not be home' and the epistemic interpretation can be only 'I conclude/it is certain that Susan is not home'. So, the relation between 'must' and 'not' does not offer any clue in the discussion at stake here.

In Italian, things are quite different. First of all, the negation particle can either follow the modal 'dovere' (119a) or precede it (119b).

(134)

    a. Gianni **deve**   **non** essere   in casa ora

        John   **must.3sg not** be   in home now

    b. Gianni **non**   **deve**   essere in casa ora

        John   **not**   **must.3sg** be   in home now

        'John must not be home now'

When the particle follows the verb 'dovere' the semantic scope is always aligned to the one displayed by the linear order, no matter whether the verb is interpreted as deontic or epistemic: 'dovere' is always interpreted above negation. When the particle precedes the verb 'dovere' (134b), the counterface reading where 'dovere' scopes above 'non' is available as well, both in the deontic and in the epistemic reading. The critical question is: is the reading with 'non' above 'dovere' available in both the deontic and the epistemic reading? In my intuition, yes. Let's consider first the deontic case. Imagine the following dialogue.

(135)   Mary: "Gianni deve essere nel suo ufficio quando il capo arriverà per l'ispezione"

'John must be in his office when the boss will arrive for the inspection'

Susan: "No, **non deve per forza/necessariamente** essere nel suo

ufficio, quando arriva il capo. Può anche essere fuori per una consulenza"

'No, it is not necessarily the case that he must be in his office when the

boss arrives. He may be out for a consultation'

Mary expressed through 'deve' an obligation for John to be in his office when the boss arrives

for the inspection. Susan can negate this obligation (i.e., deny that John is required to be in in his

office) by having the negation particle precede 'dovere' and an adverb like 'per forza' or

'necessariamente' (both translatable as 'necessarily') following the verb. In this case the

interpretation of the reply is with negation scoping above the modal.

Now, can the reply with 'non' and the adverb 'necessarily' be used when 'deve' is epistemic?

Yes.

(136)   Mary: "Gianni deve essere nel suo ufficio. Vedo la luce accesa"

'John must be in his office. I see the light on'

Susan: "No, Gianni **non deve per forza/necessariamente** essere nel suo ufficio

se la luce è accesa. Può averla dimenticata accesa uscendo"

'No, it is not necessarily the case that John must be in his office if the light

is on. He may have forgotten to turn it off while leaving'

In this case, Susan is replying to the original necessity meaning of the word.

189

However, one may object that this evidence doesn't really show the original meaning of 'dovere' because there is a possible interpretation where what is under the scope of negation is not the epistemic meaning of 'dovere' but the meaning of the adverb 'per forza/necessarily': under this interpretation 'dovere' can be still interpreted as an inferential – i.e., as contributing a not-at-issue evidential claim - while the negation particle takes the meaning of the adverb under its semantic scope.

(137)   Susan: "No, Gianni **non deve necessariamente** essere nel suo ufficio

se la luce è accesa. Può averla dimenticata accesa uscendo"

Possible interpretation:

No, **John is not necessarily in his office (which Susan has inferred)** if the light

is on. She might have forgotten the light on

Thus, this piece of evidence would be more compelling if the necessity meaning of 'dovere' could be under the semantic scope of negation without the occurrence of the adverb 'necessarily'. Notice that this is not only a problem for the epistemic reading only but also for the deontic reading as well: can Susan's replies in (135) and (136) be uttered without the adverb 'necessarily' and still express the same semantic interaction between 'non' and the verb 'dovere'? In my intuition, the answer is positive in both the deontic and the epistemic case: Susan doesn't need the adverb to convey the same message although the adverb makes the whole utterance more natural.

(138)   *Deontic reading*

Mary: "Gianni deve essere nel suo ufficio quando il capo arriverà per l'ispezione"

‘John must be in his office when the boss will arrive for the inspection’

Susan: "No, **non deve** essere nel suo ufficio, quando arriva il capo. Può anche

essere fuori per una consulenza"

‘No, he **doesn't have to be** in his office when the boss arrives. He may be

out for a consultation’

Epistemic reading

Mary: "Gianni deve essere nel suo ufficio. Vedo la luce accesa"

‘John must be in his office. I see the light on’

Susan: "No, **non deve** essere nel suo ufficio se la luce è accesa. Può averla

dimenticata accesa uscendo"

‘No, he **is not necessarily** in his office if the light is on. He may have

forgotten to turn it off while leaving’

Crucially, when the same reply is uttered with the inferential future, the only reading available is the one where the speaker has inferred that John is not in his office.

(139)   Mary: "Gianni sarà nel suo ufficio. Vedo la luce accesa"

'John will be in his office. I see the light on'

Susan: "No, #**non sarà** nel suo ufficio se la luce è accesa. Può averla dimenticata

accesa uscendo"

'No, I infer that John is not in his office given that the light is on. He may

have forgotten to turn it off while leaving'

So, to summarize, while the meaning of an inferential like the Italian 'INFER' can never

be under the semantic scope of a sentential negation particle, the meaning of epistemic 'deve'

can be forced to be under the semantic scope of a sentential negation particle, in which case

'deve' is interpreted as 'it is certain that'.

4.4 Problematic interpretations: epistemic necessity or inferential?

In this section, I briefly discuss some cases whose interpretation is not straightforward. More precisely, I discuss cases where it is not clear whether 'must' and 'dovere' are used as epistemic necessity operators or as inferentials.

4.4.1 Embedding under 'know'

In section 2.3, I showed that the Italian inferential future is borderline in situation where the relevant judge knows that 'p'.

(140)

    a. #So che cosa è successo a Susanna. Sarà malata

        'I know what happened to Susan. She will be sick'.

    b. #Maria sa che Susanna sarà malata

        'Mary knows that Susan will be sick'

In my intuition, the verb 'dovere' also is borderline in the same situations.

(141)

    a. #So che cosa è successo a Susanna. Deve essere malata

        'I know what happened to Susan. She must be sick'.

    b. #Maria sa che Susanna deve essere malata

        'Mary knows that Susan must be sick'

Nonetheless, like in the case of the inferential future, it is possible to find naturalistic occurrences of 'must' and 'deve' under 'know'.

(142)

a.   <u>Must and know</u>

    a.   He stopped, knowing what Jackson **must** certainly be thinking now[25]

    b.   We know it **must** have happened, but it's incomprehensible to think how this might have happened[26]

b.   <u>Deve and sapere</u>

    c.   Dalle regole dell'allitterazione sappiamo che il nome della figlia **deve** forzatamente iniziare per vocale.[27]

    'Based on the rules of alliterations, we know that the daughter's name must necessarily start with a vowel'

    d.   Benjamin Peirce, il noto matematico e professore di Harvard del XIX secolo, dopo aver dimostrato l'identità di Eulero in una lezione, disse: "Signori, posso dirlo con certezza, è assolutamente paradossale; non possiamo capirla, e non sappiamo che cosa significa. Ma l'abbiamo provata, e quindi sappiamo che **deve** essere la verità."[28]

    'Benjaming Pierce, the famous mathematician and professor at Harvard in the nineteenth century, after having proved the identity of Euler in a lecture, said: "Gentlemen, I can say with certainty that this claim is paradoxical; we cannot understand it. But we have proved it, therefore **we know that it must be true**'

---

[25] Source: FIC: Virginia Quarterly Review Date 2015
Publication information   Summer2012, Vol. 88 Issue 3, p116-125. 10p. 1 Black and White Photograph.

[26] Source: SPOK: CNN_Burden. Could Evidence in a South Dakota Pasture Solve the Puzzle of Payne Stewart's Death?

[27] Source: "http://it.wikipedia.org/wiki/Yrsa"
[28] Source: http://it.wikipedia.org/wiki/Identit%C3%A0_di_Eulero

Are here the words 'dovere' and 'must' used as inferentials or as epistemic necessity operators? The judgment is not sharp here.

4.4.2 Epistemic necessity verbs and logical inferences

In Chapter2, section 2.3, I showed that it not acceptable to utter INFER-p in a situation where 'p' (e.g., The ball is in the third box) is an inferred and known proposition. But both 'must' and 'dovere' are claimed to be acceptable in this situation (cf. von Fintel & Gillies 2010; Giannakidou & Mari 2018; Ippolito & Farkas 2020).

(143)

Premise 1:	You place a ball in one of three boxes positioned in a specific order.
Premise 2:	The order of the three boxes is changed by another individual while you cannot see.
Premise 3:	The individual doesn't use any trick.
Premise 4:	You open the first two boxes, and they are empty.
Conclusion:	Quindi, la palla deve essere nella terza scatola.
	'So, the ball must be in the third box'

Notice that, however, the pronunciation of the statement has to be different than the usual pronunciation of an epistemic 'must p': in this case the main stress of the utterance goes on 'deve'. Again, are here the words 'dovere' and 'must' used as inferentials or as epistemic necessity operators? In my intuition, they are used in their original necessity meaning.

4.5 Summary

In this chapter, I have offered data supporting the polysemous nature of epistemic 'must' and 'dovere' which I argued for in Chapter 3.

On one hand, I have offered data suggesting that the typical epistemic use of 'must' / 'dovere' corresponds to that of an inferential evidential like the Italian non-temporal future.

- I have offered findings from four comprehension tasks asking participants (English native speakers in Experiments 1-3 and Italian native speakers in Experiment 4) whether they would endorse a 'must p' statement in a given context, which suggest that the typical epistemic interpretation of 'must' and 'deve' is in line with the interpretation of an inferential like the Italian 'INFER': people's everyday use of 'must p' or 'deve p' as well as of 'INFER-p' depends not on whether they assess the objective likelihood of 'p' as high or certain but on whether they assess the subjective likelihood of 'p' as high enough to prompt them to jump to the conclusion (i.e., form the belief) that 'p' is true.

- I have reported several naturalistic occurring utterances where English and Italian speakers combine 'must' and 'dovere' respectively with epistemic adverbs of any degree, a behavior that would be expected if 'must' and 'dovere' are understood as 'INFER' more than as 'it is certain that'.

- I have discussed the instances when 'must p' and 'deve p' occur as embedded under doxastic predicates: in these case 'must' and 'deve' seem to behave like expressions contributing their meaning at the not-at-issue level which is expected if they are understood as inferentials.

On the other hand, I argue based on my introspection that 'dovere', unlike 'INFER', can be interpreted as contributing a regular at-issue meaning which corresponds to that of an epistemic necessity operator.

- In my intuition, an utterance of 'deve p' but not of 'INFER-p' can be judged false in a context where 'p' is not certain.

- The meaning contribution of 'deve' in 'deve p' but not the meaning of 'INFER' in 'INFER-p' can be under the semantic scope of a sentential negation particle.

Chapter 5.   Perspective taking in the truth-evaluation of epistemic statements

5.0 Introduction

Epistemics encode an at-issue semantic operator that take as its complement a proposition and whose semantic contribution is to convey an estimation of the likelihood that the proposition is true based on some other known facts. To put it in other words, the role of an epistemic in a statement is to express the speaker's (or the relevant judge) assessment of the conditional probability of a proposition given some facts available to the speaker which the speaker may or may not make explicit. The issue at stake in this chapter is: if someone utters in front of us an epistemic statement without making explicit their evidence and we are asked to assess the truth of the statement, do we inquiry into the speaker's evidence for making the assessment? In other words, do we evaluate the truth of an epistemic statement assuming the speaker's or our perspective? This question has given rise to a hot debate in the formal semantic literature (e.g., MacFarlane 2011; Egan et al. 2005; Yalcin 2011; von Fintel & Gillies 2011; Bach 2011; Yanovich 2014; Mandelkern 2019; Rudin 2021). Central to this debate are findings on truth-value judgments of 'might p' statements in "eavesdropping" scenarios where the statement is appropriately asserted from the point of view of the speaker but does not correspond to how in reality things are. (Knobe & Yalcin 2014; Beddor & Egan 2018; Phillips & Mandelkern, 2021).

In this chapter, I offer findings from two studies suggesting that in these scenarios English speakers disagree on the truth-value not only of 'might p' but also, surprisingly, of bare 'p'. To interpret this behavior, we argue that a crucial preliminary step is to understand the meaning of the adjectives 'true' and 'false' in everyday speech. Specifically, we compare two potential accounts of their meaning – one ascribing an ambiguity to the terms and the other ascribing them an inherent relative nature – and discuss their implication for the semantics of epistemic 'might p' statements

5.1 Truth of a statement and the speaker's evidence for the statement

Consider the following dialogue.

(144)   *Context*: Susan and Josh are friends. Mark is a common friend of theirs. They are all college students. It's 5pm on Wednesday. Susan and Josh are working out in the university gym. At a certain point, they start a conversation.

Josh: "I'm surprised to not see Mark: usually he is here around this time on Wednesdays"

Susan: "**He is working in the library**"

Imagine that you overheard Susan making her statements to Josh and you happen to know that Mark is not working in the library now, but he is home in his bed with a fever. In this situation, if you were asked to judge whether the statement made by Sarah is true or false, what would you answer? For anyone trained in formal semantics the answer is straightforward: 'false'. This judgment is based on the assumption that what makes a statement 'true' or 'false' is the correspondence between the piece of information (technically called 'proposition') conveyed by the statement and how things are in reality. This view makes the prediction that, in order to assess the truth of Sarah's statement, it doesn't matter which evidence Sarah was relying on for making the statement because that evidence is not part of the conveyed proposition. Let's call this the standard use of 'true statement' in formal semantics.

Now, imagine that you ask Susan why she said that Mark is working in the library, and she tells you that she had met Mark earlier in the morning and he had told her that today from 3 to 8pm he would start his part-time job at the library. Mark didn't lie to her but, unbeknown to

Susan, he got sick over lunch and could not make his shift. After you know the information

based on which Susan made her statement, would you consider that Susan was justified in

making her statement? If yes, would you still judge the statement as 'false'? According to the

standard formal semantic use of 'true statement', yes you should still judge the statement false:

even though Susan was justified in making her statement, yet the proposition conveyed by the

statement doesn't correspond to how things are.

Notice that here I have assumed that, even though the assertion is not true, Susan is

justified in asserting the sentence 'Mark is in the library now' because she had good reasons to

believe that it was true. However, this is not an uncontroversial assumption among philosophers.

In fact, over the past two decades in the epistemology and philosophy of language literature there

has been an intense debate about the conditions under which a speaker is justified in making an

assertion, which goes under the label of 'debate about the norms of assertion' (Williamson 1996,

2002; DeRose 2002; Weiner 2005, Douven 2006; Lackey 2007; Turri 2013, 2015, 2021; Reuter

& Brossels 2019; Marsili & Wiegmann 2021; Kneer 2018, 2021). The critical question is: what

is the necessary condition ("the norm") for a speaker to be entitled to assert a sentence 'p'?

Simplifying, it is possible to identify three types of answers to this question: (i) the *factivist*

hypothesis holding that to assert a sentence the speaker needs to know 'p' which implies that 'p'

needs to be true (e.g., Williamson 2000; Weiner 2005; Turri 2013), (ii) the *justified belief*

hypothesis holding that to assert a sentence the speaker needs to have good evidence supporting

the belief 'p' (e.g., Douven 2006; Reuter & Brossels 2019; Marsili & Wiegmann 2021; Kneer

2021); (iii) the *just belief hypothesis* holding that to assert a sentence the speaker needs to just

believe 'p' (e.g., Hindriks 2007; Bach 2008). Here, I place myself among the supporters of the

justified belief hypothesis: the evidence which Susan relies on to make her statement (Mark told

her in the morning that today from 3 to 8pm he would start his part-time job at the library) is

good enough to make her justified in asserting at 5pm that Mark is in the library now.

Let's imagine, now, that instead of making a bare 'p' statement Sarah said a statement

containing 'might' in its epistemic use.

(145)   *Context*: same as in 1.

Josh: "I'm surprised to not see Mark: usually he is here around this time on

Wednesdays"

Susan: "**He might be working in the library**"

When epistemic 'might' (or any other epistemic expression) occurs in a statement, is the

speaker's evidence supporting the statement part of the conveyed proposition? An influential line

of thought - which can be traced back to Kratzer (1977) - answers positively to this question.

More specifically, according to this view, Sarah's statement, which appears to be 'bare', actually

includes a non-overtly spelled-out component that refers to the body of evidence K relative to

which Sarah is making her modal judgment: "[Relative to the body of evidence K] He might be

working in the library". Within the possible world framework adopted by Kratzer, this

hypothesis consists of defining the proposition expressed by Sarah's statement roughly as: 'there

is at least a possible world compatible with the body of evidence *K* such that John is working in

the library in that world at the time of utterance'. More generally, in this view, any utterance of a

bare epistemic 'MODAL p' statement expresses a proposition of the type 'Relative to (from the

perspective of) the body of knowledge K, MODAL that p'. This view is usually called 'the

standard contextualist approach to the semantics of epistemic modality', where 'contextualist' is meant to indicate that the relevant body of evidence for interpreting the modal is established in the context of utterance (e.g., Kratzer 1977, 1981, 1991, 2012).

This standard contextualist hypothesis makes the prediction that if one were asked, for example, to assess Susan's statement in (2) as 'true' or 'false', one would first try to find out what exactly K is in this case and then check whether *based on K* it is possible that John is working at the library now. Assuming that for a matrix bare epistemic statement K is virtually always some evidence available to the speaker, then this hypothesis makes the prediction that in order to assess the truth of a matrix bare epistemic statement like in (2) it *does matter* which evidence the speaker (e.g., Sarah) was relying on making the statement because that evidence is part of the conveyed proposition. Thus, the kratzerian view predicts an *asymmetry b*etween the truth-evaluation of bare 'p' statements and bare epistemic 'might p' statements: *while in the bare 'p' case the speaker's evidence doesn't matter, in the epistemic case it does matter*.

This prediction has been explicitly challenged by MacFarlane (2003) who made the observation (crediting Hawthorne 2004 for having first pointed it out) - that "people tend to assess epistemic modal claims for truth in light of what they (the assessors) know, even if they realize that they know more than the speaker (or relevant group) did at the time of utterance" (MacFarlane 2011: 20). Although this observation is phrased as applying generally to epistemic statements, MacFarlane exemplifies the claim only with bare epistemic possibility statements and so does the rest of the literature on this issue. We follow this practice of focusing on 'might p'. Thus, MacFarlane's intuition is that epistemic 'might p' statements are assessed as 'true' or 'false' by checking whether 'p' is a possibility from the perspective of the information available to whoever is making the assessment disregarding the information available to the speaker at the

time of utterance. Based on this intuition MacFarlane (2003) advanced an hypothesis which is known as 'relativist approach to the semantics of epistemic modality', where 'relativist' refers to the fact that the truth-value of an epistemic 'might p' statement is relative to the perspective adopted by whoever is in charge of performing the assessment (e.g., MacFarlane, 2003, 2011, 2014; Hawthorne 2004; Egan 2007; Egan et al. 2007; Stephenson 2007). The current consensus among scholars of epistemic modality is that MacFarlane/Hawthorne's empirical observation is right; however, not everyone is of the opinion that this data point compels one to abandon the contextualist hypothesis altogether: some scholars, instead, suggest that what needs to be abandoned is just the standard version of the contextualist view (e.g., Yalcin 2007; von Fintel & Gillies 2008, 2011; Dowell 2011; Yanovich 2014; Stalnaker 2014) or the standard view of assertion of a statement (e.g., Rudin 2021). Thus, it is safe to say that the current consensus is that there is a *symmetry* between the truth-evaluation of bare 'p' statements and bare epistemic 'might p' statements: *in both cases the speaker's evidence doesn't matter*.

Here, we aim at assessing this very empirical generalization by asking the following question: when English speakers are asked to evaluate the truth of a 'p' or a 'might p' statement as uttered by an agent A, do they consider in their evaluation A's body of information at the time of utterance? In order to answer this question, we report here on findings from two experiments which adopted a version of the "eavesdropping task" previously employed in Knobe & Yalcin (2014), Phillips & Khoo (2018), Phillips & Mandelkern (2020), and Reuter & Brun (2021). Specifically, we focus here on the task adopted in Knobe & Yalcin (2014)'s Experiment 4: first participants read a vignette where a character utters a justified statement - either 'p' or 'might p' - but then they are told that in reality 'not-p'; next, after reading this story, they were asked to judge whether the statement is true or false. If participants don't consider the speaker's evidence

supporting a statement - either 'p' or 'might p' - when evaluating its truth-value, then we would expect that in both cases they would uniformly judge the statement 'false' given that 'not-p'. But Knobe & Yalcin (2014) found that while in the 'p' case participants uniformly judged the statement 'false', in the 'might p' case they didn't converge towards a clear response. These findings are at odds with other findings reported in Reuter & Brun (2021) showing that participants are not uniform in judging the truth of a justified but not corresponding to reality 'p' statement.

In this work, we offer two follow-up studies to Knobe & Yalcin (2014) aimed at assessing whether their findings were due to confounding factors by considering the two cases separately: in Experiment 1 we assess the 'might p' case and in Experiment 2 we assess the 'p' case. To preview, our findings suggest that participants' convergence towards 'false' in Knobe & Yalcin's (2014) 'p' condition was due to the fact that participants judged the character not justified to make the statement; instead, once one makes sure that participants judge the character justified in making the statement (as we did in Experiment 2), it turns out that a significant (although still a minority) portion of participants shifts towards 'true', making the behavior in the 'p' and the 'might p' case look much more alike, which provides further support to the picture depicted by the findings in Reuter & Brun (2021). So, before to consider the 'might p' case, what needs to be explained is the surprising (from the point of view of formal semantics) behavior in the 'p' case a significant portion of English speakers is willing to judge as 'true' a 'p' statement not corresponding to how things are in reality but justified from the point of view of the character uttering the statement. We discuss two potential explanations for this behavior which hinge on a revisitation of the meaning of the adjective 'true' (and 'false') as used in everyday speech compared to how it is used by philosophers. Moreover, we discuss how these findings

contribute to the debate on the norms of assertion. Lastly, we discuss the implication of this behavior for the semantic analysis of 'might p' statements.

The paper is organized as follows. In section 2, we review previous experimental studies that have addressed this very question. In section 3 we offer novel experimental findings. In section 4, we offer a general discussion of our findings and of their relevance for the discussion about epistemic possibility statements. In section 5, we offer some concluding remarks.

5.2 Previous studies.

Knobe & Yalcin (2014) designed a battery of experiments using the eavesdropping task. Their Experiment 4 can be considered the most critical in that it used a scenario that has been prominent in the literature and its findings were recently replicated in Phillips & Mandelkern (2020). So, we focus on this experiment. The task prompt participants to read a story where a character has some reasons to utter a statement - either 'might p' ('Modal' condition) or 'p' (Nonmodal condition) - and right after we are told that actually 'not-p' (see 3).

(146)   *Vignette in Knobe & Yalcin's (2014) Experiment 4*

> Sally and George are talking about whether Joe is in Boston. Sally carefully considers all the information she has available and concludes that there is no way to know for sure. Sally says: "Joe might be in Boston." [Modal] / "Joe is in Boston" [Nonmodal]
>
> Just then, George gets an email from Joe. The email says that Joe is in Berkeley. So, George says: "No, he isn't in Boston. He is in Berkeley."

In one of the conditions ("Falsity") participants were asked to perform a truth-value judgment of Sally's utterance. The question was asked in this form: "We want to know whether what Sally said is false. So please tell us whether you agree or disagree with the following statement: 'What Sally said is false'". Participants were asked to indicate their level of agreement on a scale from 1 ("completely disagree") to 7 ("completely agree"). Knobe & Yalcin (2014) found that while whereas the mean agreement rating for the nonmodal statement was around 6, the mean agreement rating for the modal statement was around 3. The authors interpret these two findings as suggesting that, while there is convergence among participants in judging

Sally's statement false when the statement is 'p', there is a lack of agreement among participants as to whether Sally made a false statement when the statement is 'might p'. So, based on these findings, one can be tempted to conclude that, at least for the bare 'p' statements, adult English speakers don't take in consideration the speaker's perspective at the time of utterance when invited to evaluate the truth of the statement. However, this conclusion is too hasty. In fact, Reuter & Brun (2021) report findings suggesting a different picture about the behavior of English speakers when truth-evaluating bare 'p' statements in an eavesdropping task.

In their experiment 1, Reuter & Brun (2021) designed two vignettes by adopting the following rationale: in both cases the story features a character making a bare objective 'p' statement in a situation where the information available to the character at the time of utterance is good enough to license the statement but, at the same time, the event described by 'p' does not hold. Here are the two stories.

(147)   The two stories used in Reuter & Brun's (2021) Experiment 1

[Party] Anne and Robert go to a party late at night. On their way to the party, Anne asks Robert whether any of his friends are at the party. Robert answers that Jill is at the party, because Jill had told Robert a few hours before that she would go. When they arrive at the party, it turns out that Jill had changed her plans, and actually is not at the party"

[Rolex] Maria is a watch collector. She keeps all her watches in a safe and knows her collection really well. One day, her friend John asks her whether she has a 1990 Rolex Submariner in her safe and, if so, could show it to her. Maria answers that she has got a 1990 Rolex Submariner in her safe. After all, she had purchased

that watch a few years ago. When Maria opens the safe a little later, she finds out

that a burglar has stolen several watches, among them the 1990 Rolex Submariner

After reading either of these vignettes, each participant was asked one of the following two

questions depending on the vignette they read: [Party] Was Robert's answer true or false?

[Rolex] Was Maria's answer true or false? Participants were presented with three options: (1)

true; (2) false; and (3) not sure. Surprisingly, in both scenarios most participants selected the

option 'true': 59.6% in the Party case, and 56.8% in the Rolex case (as for the other two options:

in the Party case 38.1% chose 'false' and 2.1% chose 'not sure; in the Rolex case 27.3% chose

'false' and 15.9% chose 'not sure'). In one of the conditions of their experiment 3, Reuter &

Brun (2021) re-run the task focusing only on the Rolex scenario with these modifications: they

gave participants only two response options 'true' and 'false' and had the truth question preceded

by the question 'Did Maria answer the question to the best of her knowledge?''. The latter

manipulation was based on the hypothesis that some participants might have answered 'true' in

their first experiment because instead of the truth-question they answered a question about the

assertability of the statement by the protagonist of the story: "when we first ask participants

whether the protagonist of the scenario answered the question to the best of her knowledge, it is

unlikely that the participants will still substitute the subsequent truth question with the same

question they just answered" (Reuter & Brun 2021: 13-14). Interestingly, they found that, even

though all participants answered 'yes' to the question whether Maria answered to the best of her

knowledge, they still split in their answer to the question 'Was Maria's answer true or false?':

51.1% answered 'true' and 48.9% answered 'false'. At the least, these findings suggest that for

some English speakers a bare objective 'p' statement is true as soon as its speaker made it to the

best of their knowledge (i.e., its speaker was entitled to make it based on the information

available to them) no matter whether the state of affairs described by 'p' holds in reality. What's underlying this behavior? We discuss the answer to this question in the general discussion in section 4.

To summarize, the findings in Knobe & Yalcin's (2014) Experiment 4 - replicated in Phillips and Mandelkern (2020) – suggest that English speakers uniformly disregard the speaker's evidence when truth-evaluating a 'p' statement in an eavesdropping task but are divided when it comes to truth-evaluating a 'might p' statement. However, the findings in Reuter & Brun's (2021) Experiment 1-3 suggest that English speakers are divided in truth-evaluating bare 'p' statements as well. In the next section, we offer two follow-up studies to Knobe & Yalcin's (2014) designed with the following goals: (i) to investigate whether the findings in their Modal condition were due to task effects (Experiment 1); (ii) to investigate why Knobe & Yalcin's (2014) findings in their Nonmodal condition differed from those in Reuter & Brun (2021) (Experiment 2).

5.3 Our studies.

5.3.1. Experiment 1.

In this study, our goal was twofold: first, we wanted to investigate whether the findings in Knobe & Yalcin's (2014) Nonmodal-Falsity condition (i.e., truth-value judgment of 'might p') were due to task effects; second, we wanted to make the task more homogeneous with the task in Reuter & Brun (2021) to facilitate a comparison between the two. To start, we identified two potential sources for the variability in truth-value judgment: one in the interpretation of the task and the other in the interpretation of the story. Concerning the interpretation of the task, notice that the original test question asked participants to express agreement with the target statement 'What Sally said is false'. We reasoned that, while some people may have interpreted the task as a truth-value judgment of the statement uttered by Sally (the task at stake in the theoretical discussion and the one intended by Knobe & Yalcin), some other people may have interpreted the task as an evaluation of the felicity of Sally's utterance. The hypothesis that non-linguistically trained speakers might not distinguish between truth-value judgments and felicity judgments has been advanced in recent literature concerned with methodological issues (e.g., Tonhauser & Matthewson 2015; Jasbi et al. 2019; Waldon & Degen 2020; Scontras & Pearl 2021). Under this interpretation, the variability found by Knobe & Yalcin (2014) could be due to participants adopting either interpretation: those who interpreted the task as a truth-value judgement assessed Sally's statement as false, whereas those who interpreted the task as a felicity judgement assessed Sally's statement as true. To test this hypothesis, we did two things.

First, we wanted to avoid using in the test question the anaphoric phrase 'what Sally said' which, in virtue of explicitly mentioning Sally, may prompt participants to focus on the felicity of

Sally's communicative performance more than on the truth of what she said; so, we adopted from Doran et al. (2012) the strategy of presenting the story with the statement uttered by Sally as underlined and then refer back to it in the test question through the definite description 'the underlined statement'. With this move, the new test question would have looked like: 'We want to know whether the underlined statement is false. So please tell us whether you agree or disagree with the following statement: The underlined statement is false'. However, we judged the original way of posing the test question a little bit too convoluted in that it asked participants to indirectly judge the truth-value of a statement through agreeing with another statement which refers to the statement to be truth-evaluated. So, we decided to simplify the task by eliminating the intermediate step and directly asking 'Is the underlined statement true or false?'. We also offered only two response options ('True' or 'False') because we were interested in detecting which tendency would emerge when forcing participants to take either stance. Moreover, this way of asking the test question makes the task very similar to the one adopted in Reuter & Brun (2021).

Second, we also explicitly tested whether participants judge Sally's communicative act felicitous by asking "Based on what she knows, is Sally justified to say the underlined statement?" with the response options "Justified" and "Not Justified". We had participants see either judgement alone (Between-subjects condition) or both judgements with justification preceding truth (Within-subjects condition). The goal was to test whether there is convergence in judging Sally's communicative act felicitous independently from the truth-value judgement question and whether having participants first answering the felicity question had any effect on their performance in the truth-question. In this respect, we had a similar goal in mind as the one Reuter & Brun (2021) had in mind in designing one of the conditions in their experiment 3 when

212

they asked participants 'Did Maria answer the question to the best of her knowledge?' before the truth question.

Concerning the interpretation of the scenario, we reasoned that in the original scenario designed in Knobe & Yalcin (2014) (see 3) there are three features which may be problematic: a) participants are not told any positive evidence for Sally's claim which makes Sally's utterance not completely natural in the first place; b) participants have to assume George's perspective which may increase the cognitive load; c) participants can decide to assess the statement based on what they know at the utterance time T - when the assessor/participant is in the same epistemic position as the utterer and has no reason to judge the sentence false - or at the time T' - when participants learn that 'p' is false through the email received by the character George. We speculated that part of the variability could be due a split between participants assessing Sally's statement at T or a T'. So, we also manipulated between-subjects the type of conversation such that participants read either the original conversation or a variant of it with these changes: a) Sally's evidence supporting her 'might p' claim is made explicit; b) Sally is talking directly to the participant; c) the participant is told right away at the utterance time that 'p' is false. The full materials are reported below.

(148)   Experiment 1: materials

Original story

Sally and George are talking about whether Joe is in Boston. Sally carefully considers all the information she has available and concludes that there is no way to know for sure. Sally says: 'Joe might be in Boston'. Just then, George gets an

email from Joe. The email says that Joe is in Berkeley. So, George says: 'No, he isn't in Boston. He is in Berkeley.'

New story

Imagine that you are sitting in a café in Berkeley with a friend of yours named Sally, having a conversation about another friend, Joe. Sally says: "Joe might be in Boston, since he had a business meeting scheduled there this week". You know that Joe's meeting in Boston was canceled, and he is in Berkeley right now.

Truth question: Is the underlined statement true or false?

Justification question: Based on what she knows, is Sally justified to say the underlined statement?

Putting together our manipulations, we obtained a 2X2X2 design with the following three factors each consisting of two levels: Story ('Original'; 'New') which was manipulated between-subjects; Test question ('Justification', 'Truth'); Number of test question per participant ('Between'; 'Within). We recruited 300 participants (200 for the four between conditions and 100 for the within conditions) on Mturk. We adopted the following criteria of exclusion: participants had to be native speakers of English, live in the US, correctly answer the question: "What was the name of the woman talking in the story? (Correct answer: "Sally")". Our goal was to assess the following questions: is there convergence among participants in judging Sally's utterance felicitous across stories? Does the behavior in the truth-judgment differ depending on the story (i.e., is there a main effect of the factor Story)? Does the behavior in the truth-judgment differ depending on whether it is preceded by the felicity-judgment (i.e., is there a main effect of the factor Number of test questions per participant)?

214

We present the results by judgment. Concerning the justification judgment, we found that participants converged in judging Sally justified in uttering 'might p' no matter whether the story was the original or the new one or whether the justification judgment was followed by the truth-judgment (the proportion of 'justified' selections was at ceiling in all conditions). Concerning the truth judgment, participants split evenly in their truth-value judgment of the same statement with the proportion of 'True' choices not statistically different from chance no matter the type of story or whether it was preceded by the 'Justification' judgment (see Fig. 7).
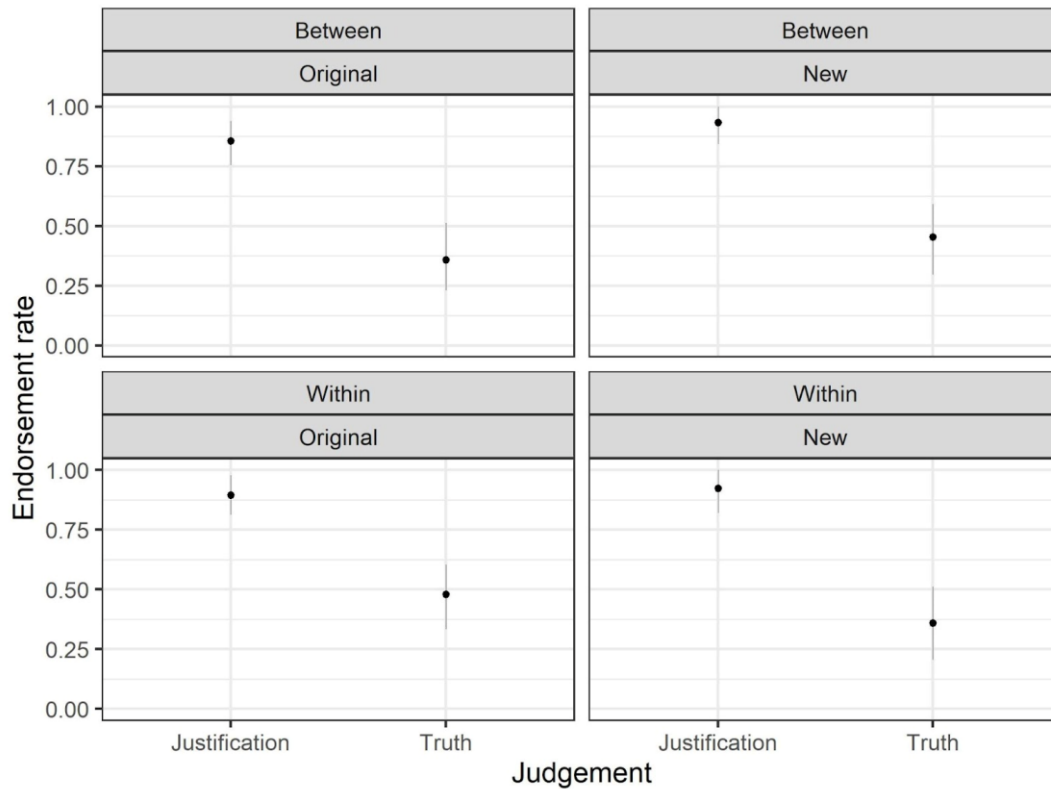

Figure 5. Experiment 1: Proportion of 'Justified' and 'True' selections by type of story and number of test questions per participants

We take our findings as suggesting that English speakers, in a situation where they all agree on judging a speaker justified to make a 'might p' statement yet split on judging whether the statement is true or false, even when they make the justification and the truth judgment one after the other. At the very least, these findings suggest that there is a significant portion of English users who judge a 'might-p' statement true as soon as they judge its speaker justified in making such a statement.

Now, notice that our 'Within' conditions in this experiment are very similar to the condition in Reuter & Brun's (2021) Experiment 3 where participants were asked in this order: "Did Maria answer to the best of her knowledge?' and 'Is Maria's answer true or false?' (See section 2 for a full description of their experiment). This allows a more direct comparison between the two experiments: in both experiments, although participants all agree that the protagonist of the story is entitled to make the statement ('might p' in our case and 'p' in Reuter & Brun's case), they disagree on whether the same statement should be judged as true or false, independently from whether the statement is 'might p' or 'p'. This raises the question as to why we didn't observe a similar behavior in the nonmodal-falsity condition in Knobe & Yalcin's (2014) Experiment 4? Experiment 2 was designed to investigate this issue.

5.3.2. Experiment 2

In this study, we aimed at investigating why in the nonmodal falsity condition in Knobe & Yalcin's (2014) Experiment 4 we didn't observe a disagreement among participants. Recall that the mean agreement rating for the nonmodal statement was around 6, suggesting that participants converged towards judging that what Sally said in the story (Joe is in Boston) is false. This judgement is relevant for the discussion at stake here only if Sally is justified in making the statement based on her information. However, if one looks at the original vignette

(see 3) then it seems that Sally is not justified to utter the bare 'Joe is in Boston' based on the information available to her (we are explicitly told that she doesn't know for sure), therefore there seems to be no conflict of perspective. So, it is not surprising that participants converged towards 'false' since from either the speaker's or their perspective the statement is not assertable. In this experiment we aimed at assessing two questions: (i) is our intuition correct that Sally is not justified in uttering 'Joe is in Boston' in the original vignette in Knobe & Yalcin's (2014) Experiment 4? (ii) If we modify the vignette such that Sally is justified in making the statement would we observe the same behavior observed in Reuter & Brun (2021)?

To answer question (i) we designed a task that adopted the methodology of the Within conditions of Experiment 1 relative to the bare 'p' statement in the original vignette. To answer question (ii) we manipulated the original vignette so that Sally is now justified in making the statement: in the new version Sally is uttering 'Joe is in Boston' based on a trustworthy report by Joe himself that he would have a job interview in Boston at the time of utterance but later we learn that the job interview was cancelled, and he didn't go to Boston (see 149).

(149)   Experiment 2: Materials

Non-justified p (Original story)

Sally and George are talking about whether Joe is in Boston. Sally carefully considers all the information she has available and concludes that there is no way to know for sure.

Sally says: "<u>Joe is in Boston</u>"

Just then, George gets an email from Joe. The email says that Joe is in Berkeley. So

George says: 'No, he isn't in Boston. He is in Berkeley.'

Justified p (Modified version)

It's late afternoon and Sally and George are in a pub in Berkeley talking about a common friend, Joe. Yesterday, Joe told Sally that he had a job interview in Boston at 5pm today and he would fly there early in the morning.

So, Sally says: "Joe is in Boston"

Just then, George gets an email from Joe. The email says that the job interview was cancelled and that he is still in Berkeley. So, George says: "No, he isn't in Boston. He is in Berkeley."

We adopted a 2X2 design with the following two factors: SALLY'S JUSTIFICATION STATUS ('justified_p'; "nonjustified_p') and TEST QUESTION ('Justification', 'Truth'). The first factor was manipulated between-subjects while the second was manipulated within-subjects. We adopted the same procedure as Experiment 1. We recruited 100 participants (50 per story) on Prolific. We predict that in the 'justified_p' condition the proportion of 'justified' selections will be close to 1 and higher than in the 'nonjustified_p' condition, which will be close to 0. Moreover, based on the findings in Reuter & Brun (2021), we predict that the proportion of 'true' selections will be significantly higher in the 'justified_p' condition than in the 'nonjustified_p' condition.

We present the results by judgment. Concerning the justification judgment, we found that the proportion of 'justified' selection is close to ceiling in the 'justified_p' condition and close to

floor in the 'nonjustified_p' condition. Concerning the truth judgment, we found that the proportion of 'true' selection is significantly higher in the 'justified_p' condition than in the 'nonjustified_p' condition as measured through a logistic regression model with "Sally's justification status" as effects-coded fixed effect (β = -1.9650, z = -2.459, p = 0.0139).
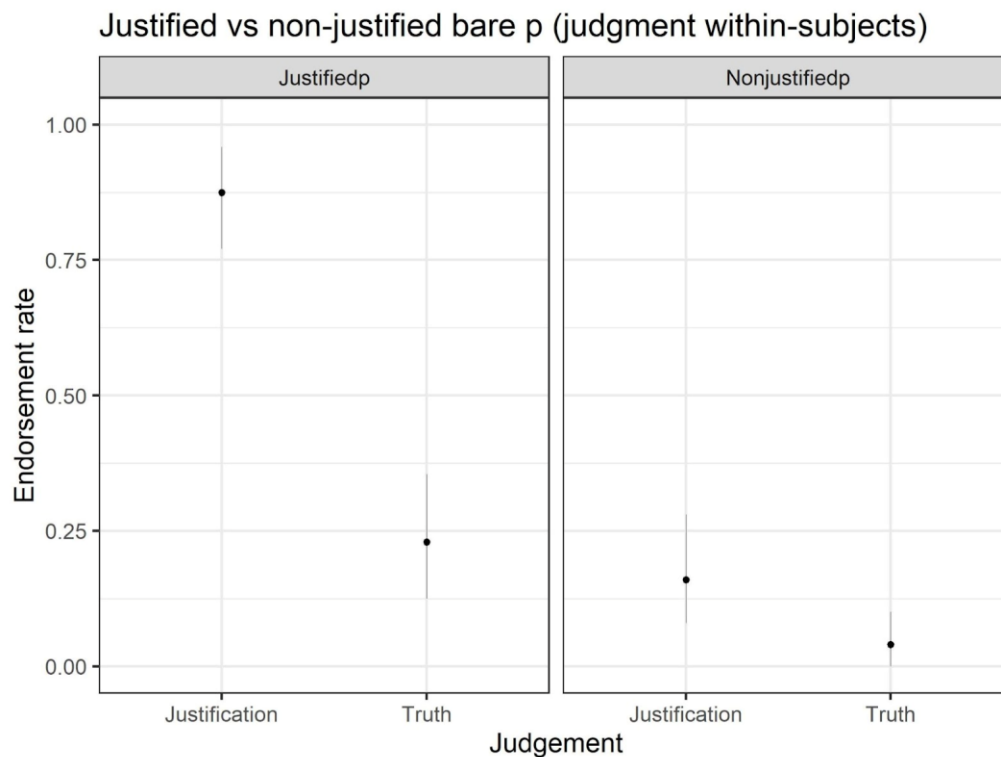

Figure 6. Experiment 2: Proportion of 'Justified' and 'True' selections by
Sally's justification status.

We take these findings as suggesting that in a context where a character produces a bare 'p' statement and it is said that 'not-p', there is a higher number of participants who judge the statement 'true' when the character is justified in making the statement than when the character is not justified.

5.4 Discussion

In this work, we have offered findings from two experimental studies. In Experiment 1, we offered data showing that when participants converge on judging a speaker justified in making a bare 'might p' statement and they know that 'not-p', they tend to evenly split in judging whether the 'might p' statement should be labeled as 'true' or 'false'. In Experiment 2, we offered data showing that in a context where a character produces a bare 'p' statement and it is said that 'not-p', there is a higher number of participants who judge the statement 'true' when they also judge the character justified in making the statement than when they judge the character not justified. These results provide further support for an empirical generalization which was already suggested by findings reported in Knobe & Yalcin (2014) and Reuter & Brun (2021): when an agent A appropriately (based on their evidence) utters a bare 'might p' or 'p' statement but 'not-p' is the case, English speakers are not homogenous in judging whether the statement should be labeled as 'true' or 'false'. We argue that underlying this behavior is not a disagreement on the semantic content of the statements but a disagreement on how to interpret the meaning of the adjectives 'true' and 'false' when applied to a statement uttered by another agent.

Let's first consider the 'p' case. The finding is surprising assuming what can be considered the default assumption in formal semantics: as soon as one learns that 'not-p' is the case, one should judge a bare 'p' statement as false independently from the information state of the original speaker. Underlying this view are these two assumptions: (i) for a bare 'p' statement being 'true or 'false' is a matter of correspondence between the information content (the proposition) expressed by the statement and how things are in the world (in the philosophical

jargon this is known as the "correspondence theory of truth"); (ii) the speaker's evidence for the proposition is not part of the proposition itself. Thus, based on these two assumptions, it is unexpected that some English speakers judge as 'true' a bare 'p' statement when in reality 'not-p' is the case. So, what is underlying this behavior? Reuter & Brun (2021) offer the following hypothesis: the lack of convergence among English speakers in performing these truth-value judgments is due to an inherent ambiguity of the adjective 'true' (and 'false') in everyday English speech. More specifically, they argue that the adjective 'true' is ambiguous between a correspondence sense (as that employed by formal semanticists) and a coherence sense according to which a "statement S of Person P [is] true or false depending on whether S coheres with other relevant beliefs of Person P, and regardless of S's correspondence with reality" (Reuter & Brun 2021: 7). Assuming this ambiguity in the use of 'true', the non-uniform behavior among participants in the relevant task can be explained as follows: those who selected 'true' adopted the coherence sense of 'true' whereas those who selected 'false' adopted its correspondence sense. Notice that using 'true' in the coherence sense is tantamount to make an assertability judgment.

Here, we suggest an alternative account of the same behavior that doesn't postulate an ambiguity for the lexical item 'true'. Recall that by default in formal semantics a statement is true if its proposition describes an actual piece of reality. At a first glance, this notion of 'actual piece of reality' seems an absolute one in the sense that establishing what counts as an actual piece of reality is independent of anyone's perspective: you just go out there in the actual world and check if the described piece of reality exists (existed or will exist) there. However, this absolute sense becomes problematic when, for example, one wants to establish the truth-value of a bare 'p' containing a predicate of personal taste like 'Jill is funny': what should one check in

the actual world to assess whether the described piece of reality exists there? A common answer among semanticists is to assume that the piece of reality described by the statement 'Jill is funny' does not exist independently from some human agent's perspective and that to establish the truth of the statement it is necessary to first assume someone's perspective on the world (e.g., Lasersohn 2005, Stephenson 2007;) In other words, statements containing predicates of personal taste can never be true simpliciter but they can only be *true for someone*, namely true in the actual world as perceived by x. Now, the standard view is to assume that the notion 'true for someone' is at play only when assessing the truth of a statement describing an inherently subjective piece of reality; instead, as soon as the statement describes an inherently objective piece of reality, there is no perspective at stake because whether an objective piece of reality exists does not depend on how someone sees the world. We speculate that the data reported here show that English speakers adopt the notion of 'true for someone' even for objective statements. More specifically, we suggest that those who selected 'true' in the task interpreted the question as 'Is the underlined statement true for the speaker?' and those who selected 'false' interpreted the task as 'Is the statement true for you?', where 'for x' can be interpreted as 'based on what x knows about the world'. Thus, we take the data reported here as supporting a view of the meaning of 'true' as denoting an inherently relativized concept.

To summarize the discussion about the bare 'p' case, findings from our Experiment 2 and from Reuter & Brun (2021) showed that in situations where an agent A appropriately (based on their evidence) utters a bare objective 'p' statement but in reality 'not-p' is the case, English speakers are not homogenous in judging whether the statement should be labeled as 'true' or 'false'. Reuter & Brun (2021) proposed to account for this lack of uniform behavior by postulating an ambiguity in the lexical items 'true' and 'false' between a coherence and

correspondence sense such that the split among participants can be accounted for by assuming that some people interpreted the words in their coherent sense (those who selected 'true') and some other in the correspondence sense (those who selected 'false'). We suggested an alternative account which, instead of postulating an inherent ambiguity in the meaning of the adjective 'true' (when applied to a statement), proposes that 'true' and 'false' only have the correspondence sense but are equipped with an implicit parameter defining a perspective - 'for x' - that needs to be settled before to use the adjective in a truth-value judgment task (this assumption is not *ad hoc* in that it is already needed for explaining the truth-evaluation behavior with a bare 'p' containing predicates of personal taste): the split among participants between selecting 'true' and 'false' can be accounted for by assuming that some people settled the parameter on the speaker's perspective (those who selected 'true') and some other on their own perspective based on the full story (those who selected 'false'). Further empirical work is needed to discriminate between the two accounts.

Let's now go back to the bare 'might p' case. The findings reported here show that in situations where an agent A appropriately (based on their evidence) utters a bare 'might objective p' statement but in reality 'not-p' is the case, English speakers behave like in the bare 'p' case: they are not uniform in judging whether the statement should be labeled as 'true' or 'false'. What does this suggest about the semantics of bare 'might p' cases? We suggest that, as in the bare 'p' case, the disagreement pertains to the truth-evaluation and not to the semantic content. Notice that a similar hypothesis has been advanced with respect to statements containing predicates of personal taste: "sentences containing predicates of personal taste are not completely objective; their truth values vary from person to person. However, this variation in truth value does not involve a variation in semantic content: if you say roller coasters are fun and I say they are not, I

am negating the very content which you assert, and directly contradicting you" (Lasersohn 2005: 684). Lasersohn's idea is that the proposition communicated by uttering 'Roller coasters are fun' doesn't have a covert 'for me' component, i.e. the expressed proposition is not inherently relativized to anyone's personal taste: it is only when one wonders about the truth-value of the statement that one needs to decide first whose personal taste is relevant (or, according to Reuter & Brun, the problem of disambiguating which of the two 'true' one means). Similarly, we speculate that the proposition expressed by epistemic bare 'might p' statements like 'Joe might be in Boston now' is not inherently relativized to an agent's (or group of agents') information state: it is only when one wonders about the truth-value of the statement that one needs to decide whose information state is relevant for making the judgement. To put it more explicitly, we argue that saying 'it is possible that/might p' is not the same as saying 'based on what I know, it is possible that p/might p: only in the second case 'p' is presented as *a possibility relative to a body of information available to the speaker* whereas in the first case 'p' is presented as a *possibility relative to how things are in the world*. What does 'possibility relative to how things are in the world' mean here? To answer this question let's step back and consider the 'p' case again.

When a speaker asserts a bare 'p' like 'Joe is in Boston now' what are they communicating? Intuitively, one is expressing a piece of information (a proposition) about the world that can be roughly paraphrased as: 'the actual world is such that the event of John being home is one of its facts at the time of utterance'. Now, we assume that a reasonable and cooperative speaker would express this piece of information based on some sort of evidence supporting the occurrence of the event of Joe being in Boston now; but that's not required by the linguistic conventions: all a speaker is doing is presenting the event of Joe being in Boston as

one of the facts making up the actual world without anchoring the occurrence of that event to any specific evidentiary basis. More generally, we can say that a cooperative speaker asserting a bare 'p' is describing the world as warranting the truth of the proposition encoded in 'p'; this claim is presumably based on some evidentiary support but is not presented as *depending on that evidentiary support*, unless explicitly stated so through the insertion of phrases like 'based on what I know' or 'based on my evidence'.

Let's go back to bare epistemic 'might p': when one asserts a bare epistemic 'might p' like 'Joe might be in Boston now' what is one communicating? The standard move made in the kratzerian approach is to postulate that in this case the expressed proposition includes an explicit reference to the speaker's evidentiary basis: '*based on some evidence available to me*, the actual world is such that the event of Joe being in Boston at the utterance time is a lively possibility in it' or 'there exists at least a world w' *compatible with some evidence available to me* such that the event of Joe being in Boston occurs in w' at the utterance time'. This hypothesis postulates an asymmetry between bare nonepistemic statements and bare epistemic statements such that the epistemic ones but not the nonepistemic are anchored to the speaker's evidentiary basis. We, instead, suggest that the epistemic statements behave like the nonepistemic in that the expressed proposition is a piece of information about the world without any explicit reference to the speaker's evidentiary basis: '*the actual world at the time of utterance is such that it doesn't contradict* the occurrence in it of the event of John being home at the time of utterance' or 'there is at least one world w' *compatible with how the actual world is at the time of utterance* such that the event of Joe being in Boston is a fact in w' at the utterance time'. More generally, we can say that a cooperative speaker asserting a bare 'might p' is describing the world at the time of utterance as being such that it allows the possibility of 'p' being true; presumably this claim is

225

based on some evidentiary support (something that the speaker knows about the world now) but is not presented as depending on that evidentiary support, unless explicitly stated so through the insertion of phrases like 'based on what I know' or 'based on my evidence'.

To summarize, we argue that bare 'p' and bare epistemic 'might p' are alike in not being semantically anchored to the speaker's evidentiary support.

Lastly, these findings shed light on the debate about the norms of assertion that I have reviewed in section 5.1. There I identified three positions about the necessary condition for a speaker to be justified in asserting a sentence: the *factivist* position (holding that the assertion needs to be true), the *justified belief* position (holding that the speaker's evidence needs to be good enough to make reasonable for the speaker to believe that the sentence is true), and the *just belief* position (holding that the speaker just needs to believe that the sentence is true). I take the findings reported here as supporting a version of the justified belief position over the other two. Specifically, the just belief hypothesis is at odds with the finding from Experiment 2 that in the original story by Knobe & Yalcin participants converged on judging the character not justified in making the statement because they were explicitly told that the character is not sure about the truth of the statement: just believing that a sentence is true is not enough to justify a speaker to assert it. The factivist hypothesis is at odds with the finding from the justified-p condition where the majority of participants evaluated the assertion as 'false' but yet they judged the character justified in making that assertion. I suggest that these data are compatible with a modified version of the justified belief hypothesis which I label "*justified I know hypothesis*": for a speaker to be justified in asserting a sentence 'p' it is necessary that based on their evidence they are justified in *believing that they know p* -- i.e., they are justified in asserting *I know p* – no matter whether it turns out that 'p' is false and that, thus, they actually don't know p.

5.5 Summary

We have showed that in situations where an agent A appropriately (based on their evidence) utters a bare 'p' or a bare 'might p' statement but in reality 'not-p' is the case, English speakers are not homogenous in judging whether the statement should be labeled as 'true' or 'false'. This behavior suggests a disagreement among English speakers on whether A's evidence at the time of utterance is relevant for the assessing the truth of a statement no matter whether a 'p' or a 'might p'. We have argued that underlying this behavior is not a disagreement on the semantic content of the statements but a disagreement on how to interpret the meaning of the adjectives 'true' and 'false' when applied to a statement uttered by another agent. Specifically, we have discussed two potential accounts: the first account ascribes to the adjective 'true' an ambiguity between a 'coherent' sense (a statement uttered by A is true as long as the statement coheres with A's evidence) and a 'correspondentist' sense (a statement as uttered by A is true as long as the statement correspond to reality); the second account treats 'true' as having only the correspondentist sense but ascribes to it an inherent relative nature (to decide if a statement as uttered by A is true one has to first decide a perspective, i.e. A's or the assessor's). Further research is needed to discriminate between these two accounts. Lastly, we have discussed the implications of the findings for the semantics of epistemic 'might p' statements: in our interpretation these findings suggest that the semantic content of bare 'might p' statements is not inherently relativized to an agent's information state but the relativization takes place only when one assesses the truth-value of the statement. This hypothesis treats bare 'p' and bare 'might p' symmetrically: in both cases the expressed proposition is not implicitly relativized in the semantics.

Chapter 6.   Conclusions

6.1 Summary of the dissertation

In this dissertation, I have reported on three projects investigating the meaning of inferentials and epistemics. All three investigations have been conducted under the assumption that to understand how these expressions are interpreted is critical to properly understand how people reason and form beliefs under conditions of uncertainty. Although studies in psychology of decision making and reasoning have already contributed important insights into the cognitive mechanisms underlying how we reason under conditions of uncertainty, these findings have not been much exploited by linguists in defining the meaning of expressions that explicitly refers to people's reasonings. In this work, I aimed at filling this gap by showing with concrete case studies how people's reasoning biases and their production/comprehension of words like 'must' and 'might' are strongly intertwined: one can shed light on the other.

Below I summarize each of the three projects reported in this document.

What are inferentials? (Chapter 2)

In chapter 2, I have investigated the meaning of inferential morphemes through the case study of the Italian non-temporal future ('INFER-p'). In a nutshell, the semantic job of 'INFER' in a clause 'INFER-p' is to introduce a not-at-issue restriction on how the contextually defined judge of the at-issue proposition 'p' acquired 'p' such that 'INFER-p' is interpreted roughly as "p, which its current judge has acquired through a reasoning".

In more details, first I have shown that INFER ascribes an inferential mode of acquisition of 'p' always to the relevant judge of "p" (the speaker in matrix contexts, the attitude holder in embedding contexts, the original source of a report in reportative contexts, the addressee in interrogative contexts).

Next, I have argued that the infelicity of inferential evidentials across languages (including INFER) in situations where the relevant judge knows that 'p' is not due to the presence of an epistemic component in its semantics, as suggested by several scholars. The main arguments against this view are that: (i) 'INFER' can co-occur in a clause with epistemic adverbs of any degree, which suggests that there is no specific epistemic component encoded; (ii) in natural speech, INFER does occur embedded under 'know'. Instead, I argue that the infelicity of 'INFER-p' in situation where the relevant judge knows that 'p' is due to a combination of two factors. First, an extralinguistic factor: in daily life we are biased to consider an inference/reasoning as a source of information whose strength is generally weak and highly dependent on the authority and the competence of the reasoner; typically, in daily life communication, most of our reasonings are far from reliable. Second, an intralinguistic factor: as soon as one believes that a proposition 'p' describes a piece of knowledge there is a communicative pressure on the speaker to utter 'p' as a bare statement; adding words or morphemes that introduce additional pieces of meaning other than just 'p' – like the inferential meaning added by 'INFER' – invites the listeners to infer that the speaker is not in the position of presenting 'p' as piece of 'knowledge'. To put it in other words, in daily communication, a speaker who is fully convinced that 'p' describes a piece of knowledge is pushed, for economy of communication, to just say 'p' even if the source of information is inferential.

Then, I have shown that the inferential proposition communicated by INFER does not semantically embed under other operators. Critically, when INFER-p occurs as embedded under a doxastic verb like 'x thinks that' the inferential claim contributed by INFER is not interpreted in the semantic scope of 'think' but remains among the commitments of the speaker.

Lastly, I have argued that an assertion of 'INFER-p' escapes the possibility of an assessment with the adjectives 'true' and 'false' because it cannot be felicitously embedded under 'it is (not) true that': as soon as a speaker utters (or says in their mind) 'it is not true that INFER-p', the speaker would be committed to be the relevant judge of 'p' whom INFER points to.

What are epistemic necessity verbs? (Chapters 3-4)

In Chapter 3, I have reviewed the debate about the meaning of epistemic necessity verbs triggered by their standard analysis adopted in logical semantics: a speaker asserting 'must p' is communicating that they know some facts which make 'p' a necessarily true proposition. This analysis predicts that speakers asserting 'must p' are in such a strong epistemic position towards 'p' that they can also assert 'it is certain that p', 'I know that p', or simply 'p'. However, the prediction of this hypothesis is at odds with the intuition of semanticists and with naturalistic occurrences of 'must p' which suggest that speakers overwhelmingly use 'must p' without being in such a strong epistemic position towards 'p', i.e., when they would not say that they are certain that 'p' or they know that 'p'.

This conflict triggered a debate with these main positions: (i) people utttering 'must p' do assert that they know some facts which make 'p' a necessarily true proposition as predicted by the standard hypothesis but tend to overuse these assertions in an exaggerated fashion, which I refer to as the 'hyperbolic logical must hypothesis'; (ii) people uttering 'must p' assert that they know some facts and hold some believes which together make 'p' a necessarily true proposition, which I refer to as the 'weak logical must hypothesis'; (iii) people uttering 'must p' assert that they know some facts which make highly probable that 'p' is true, which I refer to as the 'probabilistic must hypothesis'; (iv) people in uttering 'must p' use 'must' as an inferential

evidential, namely they assert that they know some facts which make them conclude (i.e., form the belief) that 'p' is true, which I refer to as the 'evidential must hypothesis'.

Following a suggestion in Lyons (1977), I have argued that currently epistemic 'must p' is polysemous between two senses: the logical sense (p is necessarily true given some salient facts) and the evidential sense (the speaker concludes that p given some salient facts). Furthermore, I have speculated that this evidential sense of the words 'must' and 'dovere' is a derivative meaning of the words stemming from their overuse as a marker of epistemic necessity: the original meaning indicating logical conclusions (i.e., certainties) is weakened/bleached to indicate just conclusions (i.e., opinions), after reiterated and implausible exaggerated uses of the words in their logical sense.

In Chapter 4, I have offered data supporting the polysemous nature of epistemic 'must' and 'dovere' which I argued for in Chapter 3.

On one hand, I have offered data suggesting that the typical epistemic use of 'must' / 'dovere' corresponds to that of an inferential evidential like the Italian non-temporal future. I have offered findings from four comprehension tasks asking participants (English native speakers in Experiments 1-3 and Italian native speakers in Experiment 4) whether they would endorse a 'must p' statement in a given context, which suggest that the typical epistemic interpretation of 'must' and 'deve' is in line with the interpretation of an inferential like the Italian 'INFER': people's everyday use of 'must p' or 'deve p' as well as of 'INFER-p' depends not on whether they assess the objective likelihood of 'p' as high or certain but on whether they assess the subjective likelihood of 'p' as high enough to prompt them to jump to the conclusion (i.e., form the belief) that 'p' is true. Next, I have reported several naturalistic occurring utterances where English and Italian speakers combine 'must' and 'dovere' respectively with

epistemic adverbs of any degree, a behavior that would be expected if 'must' and 'dovere' are understood as 'INFER' more than as 'it is certain that'. Lastly, I have discussed the instances when 'must p' and 'deve p' occur as embedded under doxastic predicates: in these case 'must' and 'deve' seem to behave like expressions contributing their meaning at the not-at-issue level which is expected if they are understood as inferentials.

On the other hand, I argue based on my introspection that 'dovere', unlike 'INFER', can be interpreted as contributing a regular at-issue meaning which corresponds to that of an epistemic necessity operator. In my intuition, an utterance of 'deve p' but not of 'INFER-p' can be judged false in a context where 'p' is not certain; the meaning contribution of 'deve' in 'deve p' but not the meaning of 'INFER' in 'INFER-p' can be under the semantic scope of a sentential negation particle.

From which perspective are epistemic statements truth-evaluated? (Chapter 5)

In Chapter 5, we have showed that in situations where an agent A appropriately (based on their evidence) utters a bare 'p' or a bare 'might p' statement but in reality 'not-p' is the case, English speakers are not homogenous in judging whether the statement should be labeled as 'true' or 'false'. This behavior suggests a disagreement among English speakers on whether A's evidence at the time of utterance is relevant for the assessing the truth of a statement no matter whether a 'p' or a 'might p'. We have argued that underlying this behavior is not a disagreement on the semantic content of the statements but a disagreement on how to interpret the meaning of the adjectives 'true' and 'false' when applied to a statement uttered by another agent. Specifically, we have discussed two potential accounts: the first account ascribes to the adjective

'true' an ambiguity between a 'coherent' sense (a statement uttered by A is true as long as the statement coheres with A's evidence) and a 'correspondentist' sense (a statement as uttered by A is true as long as the statement correspond to reality); the second account treats 'true' as having only the correspondentist sense but ascribes to it an inherent relative nature (to decide if a statement as uttered by A is true one has to first decide a perspective, i.e. A's or the assessor's). Further research is needed to discriminate between these two accounts. Lastly, we have discussed the implications of the findings for the semantics of epistemic 'might p' statements: in our interpretation these findings suggest that the semantic content of bare 'might p' statements is not inherently relativized to an agent's information state but the relativization takes place only when one assesses the truth-value of the statement. This hypothesis treats bare 'p' and bare 'might p' symmetrically: in both cases the expressed proposition is not implicitly relativized in the semantics.

6.2 Directions for future research

6.2.1 Quantitative data on the truth-evaluation of statements

I have claimed that, in my intuition, for utterances of 'INFER-p' it does not make sense to raise the question of whether they are true or false. When a speaker utters 'INFER-p' and I know that they have seen that 'p' I cannot say 'it is false that INFER-p' to mean 'It is false that you have concluded that p'. If, I know that 'not-p', still I cannot say 'it is false that INFER-p' but only it is false that 'p'. The problem in all these cases is that by saying 'it is true (false) that INFER-p' I'm committed automatically to judge 'p' true or false and to the truth of the meaning contribution of INFER paraphrasable as 'which I have inferred'.

However, I'm aware that this is a matter of not very sharp intuitions. It is an area where experiments across many participants are lacking. I started to fill this gap with Experiment 4 in Chapter 4 where I show that the agreement rate with 'INFER-p' like 'deve p' and 'I deduce that p' is sensitive to the inductive vs explanatory nature of 'p'. I plan to run more experiments aiming to target more directly the intuitions about the assessment of these statements with the adjective 'true' and 'false'.

But, based on the investigation in Chapter 5, I'm aware that the daily use of the adjectives 'true' and 'false' is confounded. There is always the possibility that participants will interpret the critical question 'Is this statement true?' as 'Is the speaker of this statement justified in saying it?', which blurs the distinction that matters here: my intuition is that it is exactly with the adjectives 'true' and 'false' that it doesn't make sense to perform the assessment; there is no problem in assessing whether the speaker is justified in saying 'INFER-p'.

More generally, both my intuitions about the truth-evaluation of INFER-p and the experiments about the truth-evaluation of 'might p' show that there is a need of further investigation into the meaning of the adjectives 'true' and 'false'. The formal semantics approach was built on the assumption (or perhaps I should say the hope) that asking whether a declarative statement S is true or false would simplify the task of asking what the meaning of S is: "The notion of truth, whatever problems it may have, is a little bit clearer than the notion of meaning" (Chierchia & McConnell-Ginet 2000: 71). Unfortunately, the notion of truth applied to declarative statements – which is one thing with the meaning of 'true statement' – turned out to be more complicated than expected. All these statements 'Susan will be sick', 'Susan must be sick', 'Susan might be sick' are considered declarative sentences but their assessment with the adjectives 'true' and 'false' is not straightforward. Similar concerns arise with declarative sentences containing other truth-conditional problematic expressions like 'reportedly', 'allegedly', 'plausibly', 'apparently', 'honestly', 'to tell the truth'. The immediate reaction of formal semanticists to these cases has been to deem them as conveying a non-truth-conditional meaning (i.e., a meaning that doesn't enter in the truth-evaluation of the sentence). However, as soon as the number of these non-truth-conditional expressions grows, the question arises as to whether the link between the notion of truth and the meaning of declarative sentences is natural, or it is a theory-internal artifact.

6.2.2 'Must' vs 'Certain' and 'Certainly'

Here I have argued that the auxiliary 'must' is currently used with the meaning of an inferential and that this meaning is derived through a process of semantic bleaching from the original meaning of the word as an epistemic necessity operator, after reiterated exaggerated uses. If this is true, the question arises naturally as to whether other natural language expressions of epistemic necessity like the adjective 'certain' or the adverb 'certainly' and 'necessarily' show the same pattern.

In this respect, it is interesting to notice that in Lassiter's (2016) original experiment the proportion of 'Agree' with 'Bill certainly did not win the lottery' (54%) was similar to the proportion of 'Agree' with 'Bill must not have won the lottery' (58%) and quite higher than the proportion of 'Agree' with 'It is certain that Bill did not win the lottery' (25%). This suggests that the adjective 'certainly' like 'must' seems to be understood as weaker than 'it is certain that'. This impression is further corroborated by findings reported in Yatsushiro et al. (2022).

However, I would hesitate in drawing the conclusion that 'certainly' like 'must' is used as an inferential. As I discussed in section 4.1, the one-sentence task designed by Lassiter (2016) may favor hyperbolic interpretations of statements which are reduced in a multiple-sentence task. Specifically, I predict that in multiple sentence task where participants rate 'certainly p' next to 'it is certain that p' the two statements will receive a very similar agreement rate, even in an explanatory context. This behavior would contrast with the behavior of 'must p' observed in Experiment 3. But this is an empirical claim that is in need of further verification.

In general, I suggest that the weaker uses of 'certain' and 'certainly' are better accounted by appealing to a sort of hyperbolic 'certainty' theory: when people use 'certain p' and 'certainly

p' in contexts where 'p' is just highly probable people are exaggerating, whereas 'must p' is truly used as an inferential. But this speculation needs further empirical validation.

6.2.3 Types of inferentials: the conclude-type and the suppose-type

In the descriptive literature on evidentials, it has been claimed that inferential in some languages subdivide themselves in two categories: those expressing assumptions based on general knowledge and those expressing inferences from physical evidence (cf. Aikenhvald 2018). At the same time, if one consider the English attitude predicates that refer to inferences, it is possible to identify two groups: the 'conclude'-type which may include verbs like 'conclude', 'deduce', 'infer' and the 'suppose'-type which may include verbs like 'suppose', 'assume', 'guess'. This raises the questions: (i) whether the two types of inferentials mentioned in the descriptive literature map onto the two types of inferential attitude predicates; (ii) whether 'INFER' or 'must' can be further defined as belonging to one of these two groups.

First of all, a careful investigation of the differences between the 'conclude-type' and the 'suppose'-type of inferential verbs is missing in the literature. Perhaps one difference between the two types is that the 'conclude-type does require some physical evidence from which a conclusion is drawn, whereas the suppose-type can be used with conclusions derived from no specific factual evidence available to the speaker. Moreover, it seems natural to postulate the existence of an explicit anti-knowledge requirement for the suppose-type but not for the conclude-type. If that's right, then maybe one can argue that 'INFER' maps onto the suppose-type of verbs.

Concerning 'must', I think that in its interpretation as an inferential is not sensitive to this distinction: it is naturally used both when some evidence prompts the conclusion (150) and when someone is making a guess just based on their gut (151).

(150)   My great grandfather John Hartman married a lady by the name of Margaret
          Kessler. John died in Colorado 1 Oct 1896. His wife was not in his will so must
          have died before that time

(151)   He had sent me this photograph of himself as a very attractive man, who I think
          must have been about maybe 60. He turned out to be 82.

6.2.4 The semantic extension from certainty to inference from a typological perspective

In this work, I have offered evidence that the English word 'must' and the Italian word 'dovere' exhibit the meaning of an inferential alongside the original meaning as an epistemic necessity operator. This raises naturally the question as to whether in other languages too expressions denoting an epistemic necessity operator undergo such an extension. I speculate that expressions of this type in every language can potentially undergo this extension if the frequency of use of those expressions increases.

However, a preliminary question is: how do we identify in any given language which expression denotes an epistemic necessity operator in the first place? To solve this initial problem, one can use the "Modal questionnaire for cross-linguistic use" developed by Jozina Vander Klok (https://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaire/cross-linguistic-use.php) which presents a number of contexts that target one cross-section of modal force (e.g.

239

necessity, weak necessity, or possibility) and modal flavour (e.g. plain epistemic, deontic, circumstantial, teleological).

Once one has identified which expressions denote an epistemic necessity operator in a given language, the next step is to check whether they are routinely used as inferentials. The issue is delicate because what we want is to establish that the inferential use is rooted in an established meaning of the expression and not due to a contextually defined pragmatic enrichment. In this respect, a clear indication that the expression is stored in the speakers' mind as an inferential is offered by the diagnostics based on the combination with epistemic adverbs (see section 4.2.1): if the expression is routinely combined with adverbs expressing an epistemic commitment lower than maximal (like 'probably' or 'perhaps') then one can safely conclude that the expression is perceived by speakers as an inferential. Indeed, this evidence would suggest that in the mind of the native speakers of that language the expression is perceived as not encoding any specific degree of epistemic commitment, which can be modulated by the adverbial modification.

Additionally, it would be useful to attempt to replicate in other languages the findings of Experiment 3 and 4: I predict that, across languages, a statement with an inferential will receive a higher endorsement rate in an explanatory context than in an inductive context.

6.3 Conclusion

Humans perform acts of inference daily and they talk about the conclusions of those inferences through several expressions. In this work, I have investigated two types of such expressions: the inferentials and the epistemics.

First, I have argued that people use an inferential in a statement when some evidence available to them prompts them to form the belief that the described event occurred, which is independent from their assessment of the objective probability of the event.

Next, I have argued that expressions originally denoting an epistemic necessity operator are bound to be reanalyzed as inferentials when their frequency of use in daily life increases. Reiterated utterances suggesting the certainty of an event based on the speaker's reasoning are likely to be met with skepticism by listeners, who eventually ascribe a more plausible meaning to the expression, namely that of an inferential. So, at the current stage, 'must p' exhibits, in addition to the original meaning 'it is a necessary conclusion that p', the extended meaning 'it is a conclusion of mine that p', which is the result of a semantic extension of weakening.

Lastly, I have argued that the semantic content of statements containing bare epistemics is not inherently relativized to the speaker's premises of the inference but the relativization takes place only when one assesses the truth-value of the statement.

Overall, I have conducted this investigation with the underlying assumption that a full understanding of the mechanism underlying people's actual usage of inferentials and epistemics requires that linguistic investigations be linked to investigations in psychology of reasoning: how people reason is reflected in how they talk about the conclusions of those reasonings.

References

Aikhenvald, Alexandra Y. 2003. Evidentiality in Tariana', in Alexandra Y. Aikhenvald and R. M. W. Dixon (eds.), 131–64. *Studies in Evidentiality*. Amsterdam: John Benjamins.

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press

Aĭkhenvald, AlexandraY. (ed.), 2018. *The Oxford handbook of evidentiality*. Oxford University Press.

Anderson, C. A., Lepper, M. R., & Ross, L. 1980. Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality & Social Psychology*, 39, 1037–49

Bach, K., 2008. Applying pragmatics to epistemology. *Philosophical issues*, 18, pp.68-88.

Beddor, B. and Egan, A., 2018. Might do better: Flexible relativism and the QUD. Semantics and Pragmatics, 11, p.7.

Boas, Franz. 1938. 'Language', in Franz Boas (ed.), General anthropology, 124–45. Boston, New York.

Boye, K., 2016. The expression of epistemic modality. In The *Oxford handbook of modality and mood*.

Brandom, Robert B. 1983. Asserting. Noûs 17: 637-650.

Brem, S. K., & Rips, L. J. 2000. Explanation and evidence in informal argument. *Cognitive Science*, 24, 573–604

Bybee, Joan L., Revere D. Perkins, and William Pagliuca (1994). *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Carnap, Rudolf 1964. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Chafe, Wallace L. and Johanna Nichols (eds.). 1986. Evidentiality: The linguistic coding of epistemology. Advances in discourse processes, Volume XX. Norwood, NJ: Ablex Publishing.

Chapman, L. J., & Chapman, J. P. 1967. Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 72, 193–204.

Chapman, L. J., & Chapman, J. P. 1969. Illusory correlation as an obstacle to the valid use of psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271–80

Charlow, Nate. 2020. Grading modal judgement. *Mind* 129. 769–807

Charnavel, I., 2019. Perspectives in causal clauses. *Natural Language & Linguistic Theory*, 37(2), pp.389-424.

Cinque, G. 1999. Adverbs and functional heads: a cross-linguistic perspective. Oxford, UK: Oxford University Press

Cinque, G., 2006. Restructuring and functional heads: the cartography of syntactic structures volume 4 (Vol. 4). Oxford University Press on Demand.

Coates, Jennifer 1983. The Semantics of the Modal Auxiliaries. London: Croom He

Cruse, Alan 2000. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford University Press

Dahl Ö. 2001. Inflationary effects in language and elsewhere. In *Frequency and the Emergence of Linguistic Structure*, ed. J Bybee, P Hopper, 2: 471–80. Amsterdam: Benjamin

Davidson, K. 2020. Is "experimental" a gradable predicate? In *Proceedings of NELS* (Vol. 50).

Degen, Judith; Andreas Trotzke; Gregory Scontras; Eva Wittenberg; And Noah D. Goodman. 2019. Definitely, maybe: A new experimental paradigm for investigating the pragmatics of evidential devices across languages. *Journal of Pragmatics*, 140, 33-48.

De Haan, Ferdinand. 1999. 'Evidentiality and epistemic modality: setting boundaries', *Southwest Journal of Linguistics* 18: 83–102.

Del Pinal, G., & Waldon, B. 2019. Modals under epistemic tension. *Natural Language Semantics*, 27(2).

Del Pinal, G., 2021. Probabilistic semantics for epistemic modals: Normality assumptions, conditional epistemic spaces and the strength of must and might. *Linguistics and Philosophy*, pp.1-42.

Deo, A. 2015. Diachronic semantics. Annu. Rev. Linguist., 1(1), 179-197.

DeRose, K., 2002. Knowledge, assertion, and context. *Philosophical Review*, 111(2).

Doran, Ryan, Gregory Ward, Meredith Larson, Yaron McNabb, & Rachel E. Baker. 2012. A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language*. 88(1). 124–154

Douven, I., 2006. Assertion, knowledge, and rational credibility. *The Philosophical Review*, 115(4), pp.449-485.

Douven, Igor. 2017. "Abduction", *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition) <https://plato.stanford.edu/archives/sum2017/entries/abduction/>.

Drubig, H. B. 2001. *On the syntactic form of epistemic modality*. Ms University of Tübingen.

Dowell, J.L. 2011. A flexible contextualist account of epistemic modals. *Philosophers' Imprint* 11(14), 1-21.

Eckardt, Regine & Andrea Beltrama. 2018. Evidentials and questions. In Christopher Piñon (ed.), *Empirical issues in syntax and semantics* 12, CSSP

Egan, Andy. 2007. Epistemic modals, relativism, and assertion. *Philosophical Studies* 113(1). 1–22.

Egan, Andy, John Hawthorne & Brian Weatherson. 2005. Epistemic modals in context. In G. Preyer & G. Peter (eds.), *Contextualism in philosophy*, 131–170. Oxford University Press

Falkum, I.L. and Vicente, A., 2015. Polysemy: Current perspectives and approaches. *Lingua*.

Faller, M. 2002. *Semantics and pragmatics of evidentials in Cuzco Quechua* (Doctoral dissertation, stanford university).

Faller, M. 2019. The discourse commitments of illocutionary reportatives. *Semantics and pragmatics* 12(8), 1–46.

von Fintel, Kai & Anthony Gillies. 2008. CIA leaks. *Philosophical Review*, 117(1). 77–98.

Farkas, D.F. and Bruce, K.B., 2010. On reacting to assertions and polar questions. *Journal of semantics,* 27(1), pp.81-118.

Farkas, D.F. and Roelofsen, F., 2017. Division of labor in the interpretation of declaratives and interrogatives. *Journal of Semantics*, 34(2), pp.237-289.

von Fintel, K., & Gillies, A. S. 2010. Must... stay... strong!. *Natural language semantics*, 18(4), 351-383.

von Fintel, Kai & Anthony Gillies. 2011. 'Might'made right. In Andy Egan & BrianWeatherson (eds.), *Epistemic modality*, 108–130. Oxford University Press.

Von Fintel, K. and Heim, I., 2011. Intensional semantics. Unpublished lecture notes.

von Fintel, K., & Gillies, A. S. 2021. Still going strong. *Natural Language Semantics*, 29(1), 91-113.

Foppolo, F., Guasti, M.T. and Chierchia, G., 2012. Scalar implicatures in child language: Give children a chance. *Language learning and development*, 8(4), pp.365-394.

Frana, I. and Menéndez-Benito, P., 2019, December. Evidence and bias: the case of the evidential future in Italian. In Semantics and Linguistic Theory (Vol. 29, pp. 727-747).

Garrett, Edward J. 2001. 'Evidentiality and assertion in Tibetan'. PhD dissertation, Los Angeles: University of California at Los Angeles.

Geurts, B., 2019. Communication as commitment sharing: speech acts, implicatures, common ground. Theoretical linguistics, 45(1-2), pp.1-30.

Giannakidou, A. & Mari A. 2016. Epistemic future and epistemic MUST: nonveridicality, evidence, and partial knowledge. in *Mood, Aspect and Modality: What is a linguistic Category?*, ed. by Blaszack, J. A. Giannakidou, D. Klimek-Jankowska, K. Mygdalski. University of Chicago Press.

Giannakidou, Anastasia & Alda Mari. 2018. A unified analysis of the future as epistemic modality. *Natural Language and Linguistic Theory* 36(1). 85–129

Gibson, E., & Fedorenko, E. 2013. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1-2), 88-124.

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundation of adaptive behavior*. New York: Oxford University Press

Gonzalez, C., 2016. 13 Decision-Making: A Cognitive Science Perspective. *The Oxford handbook of cognitive science*, p.249.

Goodhue, Daniel. 2017. Must ϕ is felicitous only if ϕ is not known. *Semantics & Pragmatics* 10(14). 1–27

Gunlogson, Christine. 2008. A question of commitment. *Belgian Journal of Linguistics* 22: 101-136.

Hacquard, V., 2010. On the event relativity of modal auxiliaries. *Natural language semantics*, 18(1), pp.79-114.

Hacquard, V. and Wellwood, A., 2012. Embedding epistemic modals in English: A corpus-based study. *Semantics and Pragmatics*, 5, pp.4-1.

Hamblin, C.L. 1973. Questions in Montague English. Foundations of Language 10: 41-53.

Harris, J.A., 2021. Extended Perspective Shift and Discourse Economy in Language Processing. Frontiers in psychology, 12, p.613357.

Haspelmath M. 1999. Why is grammaticalization irreversible? *Linguistics* 37:1043–68

Hindriks, F., 2007. The status of the knowledge account of assertion. *Linguistics and Philosophy*, 30(3), pp.393-406.

Kratzer, A. and Heim, I., 1998. Semantics in generative grammar (Vol. 1185). Oxford: Blackwell.

Hintikka, J. 1962. *Knowledge and belief: An introduction to the logic of the two notions*. Studia Logica, 16.

Ippolito, M. and Farkas, D.F., 2019, December. Epistemic stance without epistemic modals: the case of the presumptive future. In Semantics and Linguistic Theory (Vol. 29, pp. 459-476).

Izvorski, R. 1997. The present perfect as an epistemic modal. In A. Lawson (Ed.), *Semantics and Linguistic Theory* (SALT) 7, Ithaca, NY, pp. 222–239. LSA and CLC Publications.

Jackendoff, R. S. 1972. *Semantic interpretation in generative grammar*.

Jasbi, Masoud, Brandon Waldon, & Judith Degen. 2019. Linking hypothesis and number of response options modulate inferred scalar implicature rate. *Frontiers in psychology* 10(189), 1-14.

Karttunen, L. 1972. Possible and must. In J. Kimball, *Syntax and semantics*, vol 1, 1–20. New York: Academic Press

Katsos, N., & Bishop, D. V. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, 120(1), 67-81.

Keller R. 1989. Invisible-hand theory and language evolution. *Lingua* 77:113–27

Khoo, Justin & Jonathan Phillips. 2019. New Horizons for a Theory of Epistemic Modals, *Australasian Journal of Philosophy*, 97(2), 309-324

Kneer, M., 2018. The norm of assertion: Empirical data. *Cognition*, 177, pp.165-171.

Kneer, M., 2021. Norms of assertion in the United States, Germany, and Japan. *Proceedings of the National Academy of Sciences*, 118(37), p.e2105365118.

Knobe, Joshua & Seth Yalcin. 2014. Epistemic modals and context: Experimental data. *Semantics and Pragmatics* 7(10), 1-21

Koehler, D. J., 1991. "Explanation, Imagination, and Confidence in Judgment," *Psychological Bulletin*, 110: 499–519.

Koev T. (2018) Notions of at-issueness. Lang Linguist Compass. 12:e12306

Korotkova, N. 2016. *Heterogeneity and universality in the evidential domain*. Ph. D. thesis, University of California, Los Angeles.

Korotkova, N., 2020. Evidential meaning and (not-) at-issueness. Semantics and Pragmatics, 13, p.4.

Korotkova, N., 2021. The embedding puzzle: Constraints on evidentials in complement clauses. Linguistic Inquiry, 52(1), pp.210-226.

Kratzer, Angelika. 1977. What 'must' and 'can' must and can mean. *Linguistics and philosophy* 1(3). 337-355.

Kratzer, Angelika. 1981. The notional category of modality. In Hans-Jürgen Eikmeyer & Hannes Riese (eds.) *Words, worlds, and contexts*, 38-74. Berlin: de Gruyter.

Kratzer, Angelika. 1991. Modality. In Armin Von Stechow & Dieter Wunderlich (Eds.) *Semantics: An International Handbook of Contemporary Research*. 639–650. Berlin: de Gruyter

Kratzer, Angelika. 2012. *Modals and conditionals: New and revised perspectives*. Oxford: Oxford University Press.

Krifka, Manfred. 2013. Response particles as propositional anaphors. Semantics and Linguistic Theory (SALT). 23. 1-18.

Krifka, Manfred. 2021. Layers of Assertive Clauses: Propositions, Judgements, Commitments, Acts. MS

Lackey, J., 2007. Norms of assertion. *Noûs*, 41(4), pp.594-626.

Lasersohn, Peter. 2005. Context dependence, disagreement, and predicates of personal taste. *Linguistics and philosophy* 28(6). 643-686.

Lassiter, D. 2016. Must, knowledge, and (in) directness. *Natural Language Semantics*, 24(2), 117-163.

Lassiter, D. 2017. *Graded modality: Qualitative and quantitative perspectives*. Oxford University Press

Lewis, D. 1968. *On the plurality of worlds*. Oxford: Basil Blackwell.

Lombrozo, T. 2007. Simplicity and Probability in Causal Explanation. *Cognitive Psychology*, 55: 232–257.

Lombrozo, T. 2012. Explanation and Abductive Inference. in K. Holyoak and R. Morrison (eds.), *Oxford Handbook of Thinking and Reasoning*, Oxford: Oxford University Press, pp. 260–276.

Lyons, J. 1977. *Semantics*, 2 vols. Cambridge University Press, Cambridge.

MacFarlane, John. 2011. Epistemic modals are assessment-sensitive. In Andy Egan & Brian Weatherson (eds.), Epistemic modality, 144-178. Oxford: Oxford University Press.

MacFarlane, John., 2011. What is assertion? (pp. 79-96). na.

MacFarlane, John. 2014. *Assessment sensitivity*. Oxford: Oxford University Press

Mandelkern, M. 2019. What 'must' adds. *Linguistics and Philosophy*, 1-42.

Mari, A., 2010. On the evidential nature of the Italian future. MS

Marty, P., Chemla, E., & Sprouse, J. 2020. The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: a journal of general linguistics*, 5(1), 72.

Matthewson, L. 2004. On the methodology of semantic fieldwork. *International journal of American linguistics*, 70(4), 369-415.

Matthewson, L., Davis, H., & Rullmann, H. 2007. Evidentials as epistemic modals: Evidence from St'át'imcets. *Linguistic variation yearbook*, 7(1), 201-254.

Matthewson, Lisa. 2015. Evidential Restrictions on Epistemic Modals. In Luis Alonso-Ovalle and Paula Menendez-Benito (eds.), *Epistemic Indefinites*. Oxford: Oxford University Press, 141-160.

McCready, E. S. 2010. Varieties of conventional implicature. *Semantics and Pragmatics*, 3, 8-1.

McCready, E. and Ogata, N., 2007. Evidentiality, modality and probability. Linguistics and philosophy, 30(2), pp.147-206.

Moss, Sarah. 2015. On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics* 8. 1–81.

Moss, Sarah. 2018. *Probabilistic knowledge*. Oxford University Press.

Murray, Sarah E. 2010. *Evidentiality and the structure of speech acts*. Rutgers University dissertation. http://www.semanticsarchive.net/Archive/

Murray, S. E. 2017. *The Semantics of Evidentials*. Oxford: Oxford University Press.

Murray, S.E. and Starr, W.B., 2021. The structure of communicative acts. *Linguistics and Philosophy*, 44(2), pp.425-474.

Narrog, H., 2005. On defining modality again. *Language sciences*, 27(2), pp.165-192.

Ninan, Dilip. 2018. Relational Semantics and Domain Semantics for Epistemic Modals. *Journal of philosophical logic* 47 (1). 1-16

Nuyts, Jan. 2016. Analyses of the modal meanings. In *The Oxford Handbook of Modality and Mood*. Edited by Jan Nuyts and Johan Van Der Auwera. DOI: 10.1093/oxfordhb/9780199591435.013.1

Palmer, F. R. 2001. *Mood and modality*. Cambridge University Press.

Pancheva, R. and Rudin, D., 2019. Speaker discourse roles and the discourse profile of reportative evidentials. In *Pre-proceedings of the Amsterdam Colloquium 2019* (pp. 327-336).

Papafragou, A. 2006. *Epistemic modality and truth conditions*. *Lingua*, 116(10), 1688-1702.

Patel-Grosz, P. 2012. (Anti-)locality at the interfaces (Ph.D. thesis). Massachusetts Institute of Technology, Cambridge, MA, United States

Phillips, Jonathan & Matthew Mandelkern. 2020. Eavesdropping: What is it good for? *Semantics and Pragmatics* 13(19).

Portner, Paul. 2009. *Modality*. Oxford University Press.

Potts, C. (2005). The logic of conventional implicatures (No. 7). Oxford University Press on Demand.

Preuschoff, K., Mohr, P.N. and Hsu, M., 2013. Decision making under uncertainty. *Frontiers in neuroscience*, 7, p.218.

Rett, Jessica. 2012. On modal subjectivity. *UCLA Working Papers in Linguistics Volume* v, Article a: 1–20

Reuter, K. and Brössel, P., 2019. No knowledge required. *Episteme*, 16(3), pp.303-321.

Reuter, Kevin & Georg Brun. 2021. Empirical Studies on Truth and the Project of Re-engineering Truth. *Pacific and Philosophical Quarterly.* 1-25

Roberts Craige. 2019. The Character of Epistemic Modality: Evidential indexicals. Ms. https://www.asc.ohio-state.edu/roberts.21/Roberts.EpistemicModality.pdf

Ross, L. D., Lepper, M. R., Strack, F., & Steinmetz, J. 1977. Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality & Social Psychology*, 35, 817–29

Rudin, Deniz. 2018. Uncertainty and persistence: A bayesian update semantics for probabilistic expressions. Journal of Philosophical Logic 47(3). 365–405.

Rudin, Deniz. 2021. Asserting epistemic modals. Manuscript. University of Southern California

Rudin, D., 2022. Intonational commitments. *Journal of Semantics*, 39(2), pp.339-383.

Santorio, Paolo & Jacopo Romoli. 2017. Probability and implicatures: A unified account of the scalar effects of disjunction under modals. *Semantics and Pragmatics*.

Sauerland, U. and Schenner, M., 2007. Embedded evidentials in Bulgarian. In Proceedings of Sinn und Bedeutung (Vol. 11, pp. 525-539).

Schenner, Mathias. 2010. 'Evidentials in complex sentences: foundational issues and data from Turkish and German', in Tyler Peterson and Uli Sauerland (eds.), Evidence from evidentials, 183–220. Vancouver: University of British Columbia Working Papers in Linguistics

Shapiro, Lionel. 2018. Commitment accounts of assertion. In: Sanford Goldberg (ed.), Oxford Handbook of Assertion.

Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2010). What projects and why. Proceedings of Semantics and Linguistic Theory, 20, 309–327

Snider, T. 2018. Anaphoric Reference to Propositions. Ph. D. thesis, Cornell University.

Speas, P. 2004. Evidentiality, logophoricity and the syntactic representation of pragmatic features. Lingua 114.3.255–76

Speas, P. 2008. On the syntax and semantics of evidentials. Language and Linguistics Compass, 2(5), 940-965.

Speas, P. and Tenny, C., 2003. Configurational properties of point of view roles. Asymmetry in grammar, pp.315-344.

Sprouse, J., & Almeida, D. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa*, 2(1), 1.

Scontras, Gregory & Lisa S. Pearl. 2021. When pragmatics matters more for truth-value judgments: An investigation of quantifier scope ambiguity. *Glossa: a journal of general linguistics* 6(1). 1-37

Skordos, D. and Papafragou, A., 2016. Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, 153, pp.6-18.

Squartini, M., 2016. Interactions between modality and other semantic categories. In The *Oxford handbook of modality and mood*.

Stalnaker, Robert., 1978. Assertion. In *Pragmatics* (pp. 315-332). Brill.

Stalnaker, Robert. 2014. Context. Oxford University Press

Stephenson, Tamina. 2007. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy* 30. 487–525.

Stone, M. 1994. The reference argument of epistemic must. In *Proceedings of IWCS 1*, pp. 181–190

Swanson, Eric. 2011. Hownot to theorize about the language of subjective uncertainty. In Andy Egan & Brian Weatherson (eds.), Epistemic modality, Oxford University Press.

Swanson, Eric. 2015. The application of constraint semantics to the language of subjective uncertainty. Journal of Philosophical Logic

Sweetser, Eve. 1990. *From Etymology to Pragmatics*. Cambridge: Cambridge University Press.

Syrett, K., & Koev, T. 2015. Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics*, 32(3), 525-577.

Syrett, K. 2015. Experimental support for inverse scope readings of finite-clause embedded Antecedent-Contained Deletion sentences. *Linguistic Inquiry*, 46, 579-592.

Syrett, K. & Brasoveanu, A. 2019. An experimental investigation of the scope of object comparative quantifiers. *Journal of Semantics*, 36, 285-315.

Tonhauser, J. (2012). Diagnosing (not-) at-issue content. Proceedings of Semantics of Under-represented Languages of the Americas (SULA), 6, 239-254.

Tonhauser, J., Beaver, D., Roberts, C. and Simons, M., 2013. Toward a taxonomy of projective content. *Language*, pp.66-109.

Tonhauser, Judith and Matthewson Lisa. 2015. Empirical evidence in research on meaning. Ms

Traugott, E.C., 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, pp.31-55.

Traugott, E.C., 2011. Modality from a historical perspective. *Language and Linguistics Compass*, 5(6), pp.381-396.

Traugott, E.C., and Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press

Turri, J. 2013. The test of truth: An experimental investigation of the norm of assertion. *Cognition*, 129(2), 279–291.

Turri, J. 2015. Evidence of factive norms of belief and decision. *Synthese*, 192(12), 4009–4030.

Turri, J., 2021. Truth, fallibility, and justification: new studies in the norms of assertion. *Synthese*, 198(9), pp.8073-8084.

Tversky, A. and Kahneman, D., 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), pp.1124-1131.

van der Auwera, Johan and Vladimir A. Plungian 1998. 'Modality's Semantic Map,' *Linguistic Typology* 2: 79–124.

van der Auwera, Johan and Andreas Ammann (2013). 'Overlap between Situational and Epistemic Modal Marking,' in M. S. Dryer and M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library, ch. 76. Available online at http://wals.info/chapter/76

Veltman, F., 1996. Defaults in update semantics. *Journal of philosophical logic*, 25(3), pp.221-261.

Waldon, B. 2020. Linguistic interpretation as inference under argument system uncertainty: the case of epistemic must. In *Proceedings of the Probability and Meaning Conference* (PaM 2020)

Waldon, Brandon & Judith Degen. 2020. Modeling behavior in truth value judgment task experiments. *Proceedings of the Society for Computation in Linguistics* 3(1): 10-19.

Warner, Anthony R. 1993. *English auxiliaries: structure and history*. Cambridge: Cambridge University Press.

Weiner, M., 2005. Must we know what we say? *The Philosophical Review*, 114(2), pp.227-251.

Westmoreland, R. R. 1998. *Information and intonation in natural language modality*. Indiana University.

Willett, Thomas. 1988. 'A cross-linguistic survey of the grammaticalization of evidentiality', Studies in Language 12: 51–97.

Williamson, T., 1996. Knowing and asserting. *The Philosophical Review*, 105(4), pp.489-523.

Williamson, T., 2002. *Knowledge and its Limits*. Oxford University Press on Demand.

Wurmbrand, S., 1999. Modal verbs must be raising verbs. In Proceedings of WCCFL (Vol. 18, No. 1).

Wurmbrand, S., 2004. Two types of restructuring—Lexical vs. functional. *Lingua*, 114(8), pp.991-1014.

Yalcin, Seth. 2007. Epistemic modals. *Mind* 116. 983–1026.

Yalcin, Seth. 2010. Probability operators. *Philosophy Compass* 5(11). 916–937.

Yalcin, Seth. 2011. Nonfactualism about epistemic modals. In Andy Egan & Brian Weatherson (eds.), Epistemic modality, Oxford University Press.

Yalcin, Seth. 2012. Bayesian expressivism. In Proceedings of the aristotelian society, vol. cxii, part 2.

Yanovich, Igor. 2014. Standard contextualism strikes back. *Journal of Semantics* 31. 67–114

Yatsushiro, K., Trinh, T., Żygis, M., Solt, S., Benz, A. and Krifka, M., 2022. Certainly but not certain: The expression of subjective and objective probability. *Glossa: a journal of general linguistics*, 7(1).

Zehr, J. and Schwarz, F., 2018. Returning to non-entailed presuppositions again. *ZAS Papers in Linguistics*, 61, pp.463-480.

Zhang, Y., Ricciardi, G., & Davidson, K. 2021. How many response options in a TVJT? It depends. Poster presented at CUNY 2021

Zimmermann, Malte. 2008. Discourse particle in the left periphery. In P. Cook, W. Frey, C. Maienborn & B. Shaer (eds.), Dislocated Elements in Discourse. Oxford, Routledge. 200-231.

Zimmermann, M., 2011. 76. Discourse particles. In Volume 2 (pp. 2012-2038). De Gruyter Mouton.