Inferential Pragmatics and Epistemic Vigilance

Diana Mazzarella

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

University College London

July 2015

I, Diana Mazzarella, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

…….

Abstract

Current research on linguistic communication is grounded on the well-established assumption that speakers typically communicate more than they linguistically encode (Grice, 1989). This raises the question of what sources of information and types of cognitive operation drive the recovery of the communicated meaning (or 'speaker's meaning').

In this thesis, I argue for the following two claims: (i) pragmatic interpretation is 'inferential' in the sense that it relies on two distinct stages of 'hypothesis formation' and 'hypothesis confirmation'. While hypotheses about the speaker's meaning are constructed on the basis of linguistic evidence and available contextual assumptions, they are assessed against a criterion of pragmatic acceptability based on consideration of the speaker's mental states (i.e. her beliefs and intentions); (ii) This two-stage process is underpinned by the interaction of two distinct systems: a relevance-guided comprehension procedure (Sperber & Wilson, 1995; Wilson & Sperber, 2004) and epistemic vigilance mechanisms, which assess the quality of incoming information and the reliability of the individual who dispenses it (Sperber et al., 2010).

On the basis of this work, I develop a new model of the relationship between comprehension and epistemic assessment, and discuss its implications for the unfolding of pragmatic and epistemic vigilance capacities in the child's cognitive development, as well as the place of these systems in a modular view of the cognitive architecture of the mind. Finally, I uncover some significant methodological issues which arise in the current experimental pragmatics literature when the cognitive distinction between comprehension and epistemic assessment is overlooked.

Table of Contents

Acknowledgments

I am profoundly grateful to my supervisor, Robyn Carston, for her invaluable and unfailing support. She stands as a model of intellectual insight, conceptual clarity and sharp thinking, which represents for me a constant source of inspiration and learning.

I feel very privileged to have met Deirdre Wilson and Dan Sperber. I thank them for formative discussions and for their enthusiastic support. Their helpfulness was exceeded only by their incisive thoughts.

I have been very lucky to find myself at the UCL Linguistics Department, where the lively research environment has greatly contributed to my intellectual and personal development. In different ways, I have learnt a lot from Richard Breheny, Alison Hall, Nausicaa Pouscoulous, Paula Rubio-Fernández and Catherine Wearing. No one could wish for a better start in pragmatics.

The friendship of all my PhD colleagues has made these four years cheerful and enjoyable. A special thanks to Isabelle Needham-Didsbury for sharing all the ups and downs of this long *marathon* we started together, to Jiri Kaspar for being the first person to invite me for an after-work drink, to Kevin Tang for feeding me with halva after my exhausting teaching sessions, to Giulio Dulcinati, Caterina Paolazzi, Simona Di Paola and Irini Symeonidou for brightening up many lunch breaks in Chandler House.

During my PhD I had the opportunity to travel a lot and participate in many international conferences, where I met highly competent and enthusiastic researchers. For insightful discussions and constructive feedback, I would like to thank François Recanati, Marco Mazzone, Mark Jary, Ira Noveck, Mikhail Kissine, Nicholas Allott, Jacques Moeschler, Anne Reboul, Rachel Giora, Steve Oswald and Didier Maillat. The long-lasting friendship of many doctoral students I met at summer schools and conferences has been the most rewarding outcome of all this travelling.

I also gratefully acknowledge the funding offered by the Leverhulme Trust that made

my PhD work possible.

My life in London would not have been the same without the support of my family and friends in Italy, who did not fail to show me their love even from a distance, and the presence of many new friends in London, who made me feel at home in this beautiful city. Thanks to my parents, Patrizia and Michele, who miss me and love me so much, to Francesca, who is always the first to know how I feel, no matter which country she is in, to Eleonora, Martina, Paola, Caterina, Silvia and Alice, who make me feel as I never left them every time I go back to my hometown. Thanks to Giorgio, whose polysemous puns brought me here and to Simona, who took the plane to move to London with me. Thanks to Maria, Andreas and Filippo for all the laughs we had, and to Harris for always being there for me. Thanks to Elia and Niccolò for making our house a cosy place to go back to every day, and to Silvia for reminding me that there is time for working hard and time for being light-hearted.

This thesis is dedicated to them all.

Part I

Chapter 1 Inferential approaches to pragmatics

1.1 Intentional-inferential models

The work of the philosopher Paul Grice pioneered the so-called *intentional-inferential* tradition of theories of meaning and interpretation (Grice, 1957, 1968, 1975, 1989). This tradition relies on the following two claims. The metaphysical claim that non-natural meaning is constitutively determined by the speaker's intentions (Grice, 1957); the epistemological claim that hypotheses about the speaker's intended meaning are arrived at via pragmatic non-demonstrative inference (Grice, 1975). Following Neale (2013) these two claims can be seen as answers to very different questions: a constitutive question, CQ, and a pragmatic question, PQ.

CQ In virtue of what facts is it the case that S means whatever it is [s]he means by uttering something, X, on a given occasion? (Neale, 2013, p. 14)

PQ What sources of information and what types of cognitive operations drive the (typically spontaneous) formation of A's interpretative hypotheses about what S means by uttering X on a given occasion? (Neale, 2013, p. 17)

As far as CQ is concerned, Grice's suggestion is that 'The speaker S $meant_{NN}$ something by uttering X' is roughly equivalent to 'S intended the utterance of X to produce some effect in an audience by means of the recognition of this intention' (Grice, 1957). That is, according to Grice, to mean something by uttering X, the speaker must display the following set of nested intentions (see Strawson (1964a)). S must intend

(a) S's utterance of X to produce a certain response *r* in a certain audience A;

(b) A to recognise S's intention (a);

(c) A's recognition of S's intention (a) to function as at least part of A's reason for A's response *r*.

While this exact definition has been disputed (see, e.g., Chapter 1 of Sperber & Wilson (1986/1995)), the core proposal, that is, that speaker's meaning is an overtly expressed intention which is fulfilled by being recognised, has been endorsed and fruitfully developed by several scholars (e.g., Schiffer, 1972; Sperber & Wilson, 1986/1995; Recanati, 2004).

Interestingly, this suggests that what the speaker means by uttering X (when X is a sentence) is not constitutively determined by the linguistic meaning of X. While the latter is constitutively determined by the conventions of the language to which X belongs[1], the former is speaker-based and purely intentional.

There is an important connection between CQ and PQ. The fact that linguistic meaning does not constitutively determine what S meant by uttering X on a given occasion suggests that A will need to draw on more than knowledge of the linguistic meaning of X in order to identify what S meant by uttering X in that occasion (Neale, 2013, p. 45).

As far as PQ is concerned, Grice's main contribution was to suggest that the process of identifying (or forming an hypothesis about) what S means by uttering X on a given occasion is guided by A's expectations that the utterance should meet certain standards. According to Grice (1975), conversation is a cooperative activity and it is therefore rational to expect participants of a conversation to behave cooperatively. That is, speakers are expected to follow a 'Cooperative Principle' and to produce utterances that are truthful, informative, relevant and formulated in an appropriate manner (in accordance with his maxims of Quality, Quantity, Relation and Manner). The process of identifying the intended interpretation of an utterance is thus seen as a non-demonstrative inferential process during which interpretative hypotheses about the speaker's intended meaning are tested against these expectations. Let us consider, for instance, Grice's analysis of metaphorical uses of language: "You are the cream in my coffee" (Grice, 1975). According to Grice, the speaker utters something so blatantly false, uninformative or irrelevant that the addressee is prompted to

[1] I will not address the issue of whether language conventions can be explained in terms of regularities in speaker meaning (as Grice suggested) or not.

inferentially recover what she really means. The utterance "You are the cream in my coffee" blatantly violates Grice's (first) maxim of Quality (i.e. 'Say only that which you believe to be true'). In order to maintain the assumption that the speaker is obeying the Cooperative Principle, the addressee is led to infer that the speaker may want to communicate a related relevant and informative proposition, e.g. *You are like the cream in my coffee*.[2]

Inferential approaches to pragmatics in the Gricean tradition are characterised by the appeal to general pragmatic principles, which are thought of as broadly rational. They maintain that interpretation is guided by expectations about the speaker's behaviour, but differ as to what these expectations are. On the one hand, Neo-Griceans have developed sets of principles which are based on the Gricean maxims, e.g., Levinson's (2000) *Q-Principle*, *I-Principle*, *M-Principle*, and Horn's (2004) *Q-Principle* and *R-Principle*. On the other hand, relevance theorists have argued that the expectations of relevance raised by every act of ostensive communication account for how interpreters try to identify the speaker's intended meaning. Crucially, while Gricean and Neo-Gricean maxims and principles are rules of behaviour that motivate action and that speakers are expected to (but may not) obey, Relevance Theory suggests that the principles of relevance need not be known: to the extent that they are communicating, speakers cannot produce utterances which do not raise a presumption of their own relevance (while, of course, they may nevertheless fail to be relevant).

1.2 Relevance Theory

Relevance Theory has developed Grice's work into a theory of communication which aims at psychological plausibility. Grice insisted that the implicatures of an utterance must be calculable and provided a general schema for deriving them. This involves several steps of reasoning which counts as a justificatory structure for the selected interpretation:

---

[2] This example represents only one class of utterances discussed by Grice, those which obviously violate a maxim. For a more comprehensive discussion, see Grice (1975).

> He has said that *p*; there is no reason to suppose that he is not observing the maxims, or at least the Cooperative Principle; he could not be doing this unless he thought that *q*; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that *q* is required; he has done nothing to stop me thinking that *q*; he intends me to think, or is at least willing to allow me to think, that *q*; and so he has implicated that *q*. (Grice, 1975/1989, p. 31)

The ability to reconstruct the inferential line of reasoning proposed by Grice has usually been considered as a requirement on the theorist (Grandy, 1989) and has not been taken to correspond to any actual on-line processing in the hearer's mind.[3] That is, Grice's proposal has not been interpreted as a psychological model of utterance interpretation, but rather as a rational reconstruction of whatever processes are actually carried out in on-line interpretation. Indeed, Grice himself recognised this when he mentioned briefly that implicatures "can in fact be intuitively grasped" (Grice, 1975/1989, p. 31). The issue of the kind of processes and mechanisms (or architectural units) involved in on-line pragmatic processing, left unaddressed within Gricean philosophical pragmatics, has received renewed attention in the field of cognitive pragmatics.

Importantly, Relevance Theory proposes a cognitively realistic explanation of the expectations that drive utterance interpretation. These are expectations of *relevance*, a term that has been technically defined as follows. Relevance is a property of inputs to cognitive processes (e.g. utterances) and is a cost-benefit notion: all other things being equal, the greater the positive cognitive effects (benefits), the greater the relevance; the smaller the processing effort (cost) required to derive these effects, the greater the relevance. Positive cognitive effects are changes in an individual's representation of the world that bring about an increase in true representations, more strongly evidenced representations, better organisation of information, that is, improvements of his/her knowledge state. Sperber and Wilson (1986/1995) distinguish among three different kinds of cognitive effects: true contextual

[3] It is worth mentioning that other theorists, such as Recanati (2004, p. 50), interpret the hearer's capacity to provide an off-line (reflective) reconstruction of the inferential link between the premise(s) and the conclusion as constitutively involved in implicature derivation, thus as something that hearers are capable of reconstructing.

implications (i.e. implications that can be derived from the input and the context, but from neither input nor context alone), warranted strengthening of available assumptions and warranted contradiction and elimination of available assumptions.

Relevance Theory is a theory about cognition, in general, and communication, in particular. It claims that human cognition is geared to the maximisation of relevance ('Cognitive Principle of Relevance') and that inferential communication takes place against this cognitive background. According to Relevance Theory, the exercise of pragmatic abilities involves a dedicated inferential mechanism, or module, which takes as input an ostensive stimulus and delivers as output an interpretative hypothesis about the communicator's meaning (Sperber & Wilson, 2002). This special-purpose procedure, which is automatically applied to any attended ostensive stimulus, is motivated by the following regularity in the domain of overt communication:

(1) **Communicative Principle of Relevance**
Every ostensive stimulus conveys a presumption of its own optimal relevance.

That is, every ostensive stimulus raises the expectation that it will be worth the effort required to understand it (more than anything else the addressee could have paid attention to at that time) and as relevant as the communicator can make it given her abilities and preferences:

(2) **Presumption of optimal relevance**
a. The ostensive stimulus is relevant enough to be worth the audience's processing effort.
b. The ostensive stimulus is the most relevant one compatible with the communicator's abilities and preferences.
(Sperber & Wilson, 1995, p. 270; Wilson & Sperber, 2004, p. 612)

The presumption of optimal relevance sets the level of relevance that the audience is entitled to expect, that is, at least relevance enough, and beyond that, the highest level of relevance that the communicator is capable of achieving given her means ('abilities') and goals ('preferences'). It drives and justifies the following comprehension heuristic:

(3) **Relevance-guided comprehension procedure**

a. Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.

b. Stop when your expectations of relevance are satisfied.

(Wilson & Sperber, 2004, p. 613)

This comprehension procedure is driven by occasion-specific expectations of relevance, underpinned by the general presumption of optimal relevance that is carried by all ostensive stimuli.

Consider the following example, discussed by Wilson and Sperber (2004, pp. 615-616). Wilson and Sperber (2004) provide a schematic representation of how the addressee, Peter, might use the relevace-guided comprehension procedure to interpret Mary's utterance.

(4) a. *Peter:* Did John pay back the money he owed you?

b. *Mary:* No. He forgot to go to the bank.

A	Mary has said to Peter, "He_x forgot to go the $BANK_1$/$BANK_2$. [He_x= uninterpreted pronoun] [$BANK_1$ = financial institution] [$BANK_2$ = river bank]	*Embedding of the decoded (incomplete) logical form of Mary's utterance into a description of Mary's ostensive behaviour.*
B	Mary's utterance will be optimally relevant to Peter.	*Expectations raised by recognition of Mary's ostensive behaviour and acceptance of the presumption of relevance it conveys.*
C	Mary's utterance will achieve relevance by explaining why John has not repaid the money he owed her.	*Expectations raised by B together with the fact that such an explanation would be most relevant to Peter at this point.*
D	Forgetting to go to the $BANK_1$ may make one unable to repay the money one owes.	*First assumption to occur to Peter which, together with other appropriate premises, might satisfy expectation C. Accepted as an implicit premise of Mary's utterance.*
E	John forgot to go to the $BANK_1$.	*First enrichment of the logical form of Mary's utterance to occur to Peter which might combine with D to lead to the satisfaction of C. Accepted as an explicature of Mary's utterance.*
F	John was unable to repay Mary the money he owes because he forgot to go to the $BANK_1$.	*Inferred from D and E, satisfying C and accepted as an implicit conclusion of Mary's utterance.*

G	John may repay Mary the money he owes when he next goes to the $BANK_1$.	*From F plus background knowledge. One of several possible weak implicatures of Mary's utterance which, together with F, satisfy expectation B.*

Relevance Theory conceives of utterance interpretation as a process that starts with a metarepresentation of an attributed utterance, A, and ends with a metarepresentation of an attributed thought, e.g., "Peter communicated that John forgot to go to the $BANK_1$, …". In the scenario described, the hearer, Peter, brings to the comprehension process not only the general presumption of optimal relevance (B) but also fairly precise expectations about the way in which Mary's utterance is intended to be relevant, that is, by explaining why John has not repaid the money (C). One of the decoded logical forms in (A), He_x forgot to go the $BANK_1$, provides access to the contextual assumption in (D). This can be used as a contextual assumption to derive the intended implication (F), as long as the utterance is interpreted as explicitly communicating that John forgot to go to the $BANK_1$ (E). The three interpretative sub-tasks of constructing interpretative hypotheses about explicit content, implicated premises and implicated conclusions are thus run in parallel (via a process of 'mutual parallel adjustment') by assessing hypotheses in order of accessibility until the hearer's expectation of relevance are satisfied.

1.3 Inference and communication

The intentional-inferential model of communication proposed by Relevance Theory has been recently challenged by philosophers and cognitive scientists who advocate that the pragmatic component of communication is, at least partly, non inferential (Recanati, 2002, 2004; Mazzone, 2009, 2011, 2013). For instance, Recanati has suggested that "only when the unreflective, normal process of interpretation yields weird results does a *genuine inference process* take place whereby we use evidence concerning the speaker's beliefs and intentions to work out what he means" (Recanati, 2004, p. 38 , *my emphasis* (DM)). I will present and assess these criticisms in Chapter 2 and Chapter 3. For the time being, it is worth focusing on two crucial, but arguably overlooked, preliminary issues. The first concerns the question of what counts as *pragmatic inference* (1.3.1, 1.3.2). The second concerns the relation

between pragmatic inference and the role attributed to a speaker's beliefs and intentions in utterance interpretation (1.3.3).

1.3.1 Levels of analysis

The expressions 'inference' and 'inferential' are often used within the pragmatics literature with various significantly different, even if related, meanings. Most importantly, theorists who place themselves on different sides of the debate between inferential and non-inferential approaches to pragmatics often do not share the same notion of 'inference'. Does this mean that the debate is purely terminological? I believe we cannot even start addressing this issue without engaging in some terminological and conceptual clarifications.

To begin with, it is worth describing the different levels of analysis at which the distinction between inferential and non-inferential approaches to pragmatics arise. By doing this, it becomes possible to clarify the extent to which these approaches come apart from each other. I propose to distinguish three different levels of analysis which are at work in pragmatic theories. Starting from the highest (most abstract) and proceeding to the lowest we find: (i) analysis of communication, (ii) analysis of the requirements on a theory of pragmatic processing, (iii) analysis of pragmatic mechanisms. In what follows I first consider each level of analysis and the respective characterisation of what counts as an inferential explanation at that particular level, and then advance some reasons for believing that the debate at issue should be addressed at level (ii).

The first level of analysis, (i), concerns the explanation of how (verbal) communication is achieved. At this level, inferential models are opposed to code models (see Sperber and Wilson (1986/1995)). According to code models of verbal communication, all communication is a matter of decoding alone. The grammar of a language (perhaps supplemented by code-like pragmatic rules, e.g. a token of "I" refers to the person who utters it) works as a code which pairs phonetic representations with thoughts. As a result, the encoded/decoded message represents and exhausts the communicated content. According to the inferential model, communication is a matter of producing and interpreting evidence of the

communicator's intentions. That is, the linguistically encoded meaning of the sentence uttered functions as a clue or a set of clues for the hearer in recovering (inferring) the communicated message. The inferential approach to verbal communication, as just sketched, does not impose any constraint on the nature of the on-line pragmatic processes which are responsible for the derivation of the speaker meaning, other than that pragmatics cannot be explained in terms of a code. The inferential model is compatible with these processes being deductive, inductive, abductive, associative, or some other kind of (non-coded) process, or a combination of these:

> [This] sense of 'inferential' […] does not imply an opposition to 'associative': this sense refers to the idea that pragmatic processes involve more than simple decoding. It is not this sense, however, that is at issue here. (Mazzone, 2009, p. 337)

As Mazzone (2009) emphasises, this is not the level of analysis at which the debate between inferential and non-inferential approaches to pragmatics has been framed, being rather the inferential model of communication which is endorsed by both parties (e.g., Mazzone, 2009; Recanati, 2002; Sperber & Wilson, 1986/1995). The critical opposition is thus to be found at level (ii) or at level (iii), or both.

The second level of analysis, (ii), concerns the theoretical requirements that an explanation of pragmatic processing should meet. Specifically, I maintain that inferential and non-inferential approaches differ from each other with regard to the kind(s) of stages that are said to be involved in utterance interpretation. The interpretative processes can be described as involving one or both of the following steps (Wilson & Matsui, 1998, pp. 2-3):

(a) Candidate interpretations differ in their accessibility, and are therefore entertained in a certain order.

(b) They are evaluated in terms of some criterion or standard of pragmatic acceptability that the resulting overall interpretation is supposed to meet.

I suggest that what makes a pragmatic account genuinely inferential is that it maintains a distinction between these two stages. Inferential accounts of pragmatic interpretation conceive of pragmatic inferences as a kind of non-demonstrative inference, which involves distinct stages of 'hypothesis formation' and 'hypothesis confirmation'. While the former is driven by consideration of accessibility, the latter adheres to (variously specified) standards of pragmatic acceptability. The interpreter is not justified in simply choosing the first interpretation that comes to mind in virtue of its high accessibility; the selected interpretation needs to satisfy some standards of acceptability in order to be inferentially warranted. Explanatory considerations are a crucial component of the notion of abduction/inference to the best explanation: the interpreter is justified in selecting an interpretation if and only if this is the *best* available explanation of the fact that the speaker has uttered a certain sentence. An acceptability criterion could then be defined as a criterion for selecting the best explanation.

On the other hand, non-inferential (or 'accessibility-based') approaches to pragmatics collapse the distinction between stage (a) and stage (b), and explain utterance interpretation as the result of a single process which simply selects the most accessible interpretation. According to this perspective, there is no need to appeal to any further (specifically inferential) step or 'confirmatory stage' and no pragmatic principle need enter the picture.

Importantly, the relevance-guided comprehension procedure proposed by Wilson and Sperber (2004), (3), is a dedicated *inferential* mechanism in the sense just described: clause (a) suggests that interpretative hypotheses are formed on the basis of considerations of accessibility and clause (b) states that these hypotheses are confirmed when they satisfy the addressee's expectations of relevance.

(3) **Relevance-guided comprehension procedure**
   a. Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.
   b. Stop when your expectations of relevance are satisfied.

The expected level of relevance, thus, determines the stopping point of the comprehension procedure and represents the 'acceptability criterion' that an interpretative hypothesis needs to satisfy in order to be retained and attributed to the communicator.

The third, least abstract, level of analysis, (iii) concerns the actual cognitive mechanisms that implement pragmatic processing. At this level, it is possible to distinguish between (non-inferential) associative and inferential accounts of pragmatic processing. According to the former, pragmatic processes are a type of associative processes, i.e. processes based on associations between representations and the spreading of activation within conceptual networks. The interpretative process is entirely driven by this dynamics of activations-and-associations. Inferential accounts, on the other hand, claim that associative processes are not sufficient. That is, while associative mechanisms might play their role in the initial stages of pragmatic processing, by determining the degree of activation of interpretative hypotheses about the explicit and the implicit communicated contents, these interpretative hypotheses must then be injected into the properly inferential component of the process. Highly accessible interpretative hypotheses about the explicit content, the implicated premises (contextual assumptions) and the implicated conclusions (contextual implications) are mutually adjusted in parallel via forward and backwards inferences. Crucially, the outcome of this process is a set of propositions that forms a sound argument, in the sense that the premises warrant the conclusions (Sperber & Wilson, 1998).

To sum up, the dichotomy between inferential and non-inferential approaches to pragmatics is articulated across different levels of analysis in different ways:

(i) Inferential vs. code models of communication
(ii) Inferential vs. accessibility-based accounts of pragmatic processing
(iii) Inferential vs. associative pragmatic mechanisms

As mentioned above, most theorists adopt an inferential model of verbal communication as the starting point for their pragmatic investigation (this is as much the case for Recanati (2002, 2004) and Mazzone (2009, 2011) as it is for relevance theorists). For this reason, my investigation is not concerned with the dichotomy

described in (i). More contentious are the oppositions related to the second and third levels of analysis. In what follows, I focus on the opposition between inferential and accessibility-based accounts of pragmatics, which arises at level (ii). Two reasons support this choice: first, this distinction captures important aspects of the debate between inferential and non-inferential approaches to pragmatics as conducted so far; second, the level (iii) distinction between inferential and associative pragmatic mechanisms has not yet been sufficiently developed by the different parties involved in this debate. For instance, as Wilson and Carston notice:

> From a cognitive point of view, all inferential relationships are also associations: an inferential mechanism establishes systematic correspondences between (constituents of) premises and (constituents of) conclusions. […] However, as illustrated above, not all associations are inferential. (Wilson & Carston, 2007, p.244)

In contrast, (ii) provides a clear and uncontroversial 'working criterion' to draw a distinction between inferential and non-inferential approaches to pragmatics. While inferential approaches maintain that utterance interpretation involves two distinct stages of hypothesis formation and hypothesis confirmation, driven, respectively, by considerations of accessibility and considerations of acceptability, non-inferential approaches conceive of utterance interpretation as driven by considerations of accessibility alone. Once this distinction is clearly drawn, it is possible to address this debate with a precise question in mind: can accessibility alone account for a range of pragmatic phenomena?

1.3.2 Alternative criteria

In what follows I discuss and rule out a criterion for classifying inferences proposed by Recanati (2002, 1995). As Recanati (2002) emphasises, contemporary cognitive science often relies on a 'broad' notion of inference, which comprises every evidence-based transition from one judgment or representation to another. Recanati proposes to distinguish this cognitive-scientific use of 'inference' from a philosophical use on the basis of the 'availability condition'. The 'availability condition' can be seen as a criterion to isolate, within this broad and liberal notion of inference, what Recanati takes to be the class of genuine or 'narrow' inferences:

(5) **Availability Condition**

i. one judgment (the conclusion) is grounded in another judgment (the premise),

ii. both judgements, as well as the fact that one is grounded in and justified by the other, are available (consciously accessible) to the judging subject.

(Recanati, 2002, p. 117)

Only narrow inferences satisfy this condition. Narrow inferences can be either spontaneous or explicit, depending on whether the inferential steps needed to bridge the gap between premise and conclusion are conscious ('explicit inference') or not ('spontaneous inference'). But they are always *consciously accessible* (in the sense described in (5). In the case of spontaneous inferences

> […] the interpreter must only be *capable* of making the inference (dispositional availability), but the fact *that* an inference is involved must be occurrently grasped. (Recanati, 2007b, p. 53)

On the basis of (5), Recanati suggests that primary pragmatic processes (i.e. processes which contribute to the recovery of the explicit content of the utterance) are not genuinely inferential, that is, they are inferential only in a 'broad' sense. On the contrary, secondary pragmatic processes (i.e. processes of implicature derivation) are genuine (and spontaneous) inferences.[4] Interpreters are usually not aware of the processes through which the primary meaning, 'what is said', is derived. Even if, as Carston (2007) notes, interpreters may occasionally be aware of the relationship between the linguistically encoded meaning and the primary meaning, as in (6) for instance, this awareness is only 'accidental', a contingent matter, and not 'constitutively' involved in the derivation of 'what is said'.

(6) Mother to young child before bedtime: Have you brushed your teeth?
Child (grinning): Yes I have – *[pause]* – last night.

(Carston, 2007, p. 36)

In contrast, in deriving implicatures, the interpreter must be aware that 'what is implicated' is grounded in, or justified by, 'what is said' (even if he may not be

[4] See section 2.1 for a more detailed discussion of the distinction between primary and secondary pragmatic processes.

aware of the specific inferential steps involved). In other terms, he must be aware of the existence of an inferential connection between premises and conclusion, "whatever methods it [the pragmatic system] spontaneously uses in arriving at the 'conclusion'" (Recanati, 2004, p. 50).

As García-Carpintero (2006) underlines, Recanati's 'availability condition' relies on *phenomenological* considerations. That is, the distinction between narrow and broad inferences can be drawn only by appealing to phenomenological evidence concerning the interpretative process as experienced from the addressee's point of view. A psychological process counts as genuinely inferential (i.e. inferential in the narrow sense) if and only if the agent 'consciously grasps' the presence of a justificatory link between the input and the output of the process itself. The 'availability condition' can thus play a significant role in the debate between inferential and non-inferential approaches to pragmatics only if there are reliable intuitions and/or experimental evidence that can be called on in assessing the nature of specific pragmatic processes.

It may be argued that such pieces of (authoritative) phenomenological evidence are missing. As a matter of fact, armchair intuitions about the conscious availability of primary and secondary pragmatic processes seem to diverge (e.g. Carston, 2007; García-Carpintero, 2006), and, as Recanati himself recognises, "Carston says that the constitutiveness claim is hard to test experimentally, and she may be right" (Recanati, 2007b, p. 53). In the absence of the required phenomenological evidence, the 'availability condition' leaves us with no practical method to distinguish inferential from non-inferential pragmatic processes.

Furthermore, it may be questioned whether this is even *in principle* the right way to draw such a distinction. It is far from clear why 'conscious availability' should be the hallmark of genuine inferential processes. The availability condition appears to be a legacy of Grice's 'calculability' requirement on 'conversational implicature':

> The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all)

> will not count as a conversational implicature; it will be a conventional implicature. (Grice, 1989, p. 31)

While 'calculability' applies at the level of rational reconstruction, its counterpart at the level of actual pragmatic processes is, for Recanati, the 'conscious graspability' of the justificatory link between 'what is said' and 'what is implicated'.[5] According to Recanati, implicature derivation is characterised by the fact that the interpreter is aware that the speaker has said that *p*, thereby implicating that *q*. He insists that implicature derivation displays the specific property of 'occurrent conscious availability', which is taken to be constitutive of its inferential nature, and thus of inferential processes quite generally.[6] What this amounts to, then, is nothing more than a stipulation about the character of genuine on-line inferential processes. Recanati does not provide us with an independently motivated criterion to judge whether some pragmatic process is genuinely inferential or not. Rather, he takes a property of a specific pragmatic process, that is, implicature derivation, and he grants it the status of a distinguishing criterion between inferential and non-inferential (or broadly inferential) processes. However, not only is 'conscious availability' a debatable property of implicature derivation (see, for instance, Carston (2007)), but, even under the assumption that implicature derivation is indeed consciously available, Recanati's strategy is controversial. First, the adoption of such a criterion presupposes (rather than explains) the inferential nature of implicature derivation. Second, from the fact that implicature derivation is inferential, and that implicature derivation is a consciously available process, it does not follow that inferential processes must be consciously available quite generally. Because of these methodological and theoretical concerns, I dismiss the 'availability condition' in addressing the debate at issue.

---

[5] Note that Recanati maintains that the addressee must also be able to make explicit to himself (i.e. to rationally reconstruct) what the speaker said and what the speaker implied and that there is an inferential link between them.

[6] Among others, Sperber and Wilson (2002) characterise cognitive (on-line) inference in a different sort of way which does not require 'conscious occurrent graspability': "inferential comprehension is in general an intuitive, unreflective process which takes place *below the level of consciousness*" (Sperber & Wilson, 2002, p. 9, *my emphasis* (DM)). Ultimately this must be an empirical issue.

1.3.3 Intention-reading

As noticed above, the debate concerning the inferential status of pragmatic processes has been tightly linked to the discussion of the role of the speaker's mental states in utterance interpretation: "only when the unreflective, normal process of interpretation yields weird results does a *genuine inference process* take place whereby we use *evidence concerning the speaker's beliefs and intentions* to work out what he means" (Recanati, 2004, p. 38 , *my emphasis* (DM)).

Crucially, it is not the speaker's communicative and informative intentions that can be taken as sources of information in the process of constructing an interpretative hypothesis about the speaker's meaning, because the speaker's meaning precisely consists of this nested set of intentions. As suggested by Neale (2013, p. 44), an answer to the pragmatic question of how the audience A epistemically determines what the speaker S meant by uttering X on a given occasion could involve the following: (a) A's grasp of the linguistically encoded meaning of X, (b) the informational contents of certain of A's perceptual and doxastic states, and (c) cognitive principles and operations governing the identification, evaluation and integration of information arriving from various sources. Interestingly, information in (b) can include "A's *takes on* (i.e. beliefs and expectations about, or perceptions, conceptions or estimations of) [..] the extent to which A and S share specialized knowledge of what is being discussed; S's general knowledge; S's general intelligence; the extent to which S can be presumed to be a rational, cooperative speaker". That is, the addressee's beliefs about the speaker's beliefs and intentions may be used as evidence in the process of recovering the speaker's intended meaning.

The question thus arises of whether these pieces of information are indeed needed in order to recover (part of) the speaker's intended meaning or if they enter the interpretative process only once more basic – brute-causal – processes, which are insensitive to considerations about the speaker's mental states, fail to deliver a satisfactory output. This question, however, must be kept separate from the issue of whether pragmatic processes are inferential in nature. The latter, according to the picture sketched in section 1.3.1, revolves around Neale's point (c). As discussed

above, it is the role played by pragmatic criteria of acceptability, against which interpretative hypotheses are *evaluated*, which justifies classifying the interpretative process as inferential (rather than accessibility-based).

Chapter 2 Accessibility-based approaches to pragmatics

In this chapter, I present and evaluate Recanati's accessibility-based account of primary pragmatic processes (i.e. disambiguation, reference assignment and other cases of saturation, lexical adjustment and other kinds of free enrichment) and discuss some recent experimental studies on the nature of these processes.

2.1 Recanati's accessibility-based account of primary pragmatic processes

Recanati (1995, 2002, 2004) introduces a distinction between primary and secondary pragmatic processes: the former contribute to the determination of 'what is said' (or explicit utterance content) whereas the latter are responsible for the derivation of 'what is implicated'. Not only do they play distinct roles in utterance interpretation, but they are thought of as processes of different kinds. In particular, primary pragmatic processes are conceived of as a type of associative process, whereas secondary pragmatic processes are thought of as essentially inferential, in the strong sense of involving premises and conclusions that are warranted by those premises.

Recanati (2004) sketches a unified account of how primary processes work which is based on the notion of accessibility. The accessibility of a mental representation corresponds to its degree of activation. The more a representation is activated, the more it is consequently accessible. Activation results from two main sources: the current (or recent) processing of the representation and the spread from associatively related representations. Thus, it is a constrained and intrinsically dynamic measure. Recanati's accessibility based framework claims that, in processing a linguistic expression, different candidates for the status of its 'semantic' value are activated in parallel and compete with each other. The candidate which is retained and undergoes semantic composition is the most accessible one. It corresponds to the most highly activated representation when the interpretation process stabilises. Interestingly, the degree of accessibility associated with competing representations is not fixed once and for all. In fact, the occurrence of an accessibility shift may reverse earlier results of the competition among different conceptual representations.

Let us begin by describing how alternative candidates enter the competition for the status of semantic value, i.e. what the process of semantic composition operates on. According to Recanati, a natural candidate for this status is the literal (encoded) meaning of the expression. This claim needs further elaboration and allows us to distinguish three different cases on the basis of the availability of such a literal meaning: (a) there is a single encoded meaning (b) there are two (or more) encoded meanings, (c) the encoded meaning does not correspond to a concept which could contribute to the explicit content of the utterance. The latter circumstance (c) occurs in cases of reference assignment. Consider the following example:

(7) John was arrested by a policeman yesterday. He had just stolen a wallet.
(Recanati, 1995, p. 212)

The obvious candidates for the pronoun "He" in the second sentence are John and the policeman since they are both mentioned in the previous sentence. Assuming that they are the only sufficiently accessible candidates, they are equally compatible with the encoded meaning of the pronoun and they are both provided by the linguistic context. In other terms, the linguistic context raises their accessibility and, as a result, they enter into the competition.

Ambiguity resolution exemplifies circumstance (b), in which all the competing candidates are literal meanings associated with the expression. Finally, in cases of sense modulation (a), the literal (encoded) meaning is accessed first and it triggers the activation of further meanings. These are associatively derived from the literal meaning through which they gain part of their activation. In this respect, Recanati recognises the priority of literal over non-literal (or derived) meanings. However, this priority is in force only at a word level and does not entail that a literal interpretation of the utterance needs to be computed prior to a non-literal one. The priority of the literal meaning is not a warranty of its success during the interpretation process. In fact, derived meanings can gain further activation from the linguistic and extra-linguistic context and eventually win the competition.

To sum up, competing candidates for the status of semantic value of an expression may come from different sources: the decoding of the expression (literal meaning),

the conceptual associations with the linguistic meaning, and the linguistic or the extra-linguistic context. In Recanati's terms, activation can be "*over*-determined" (Recanati, 1995, p. 218).

Let us now turn to the issue of the dynamics of accessibility. The different degrees of accessibility of competing representations can change throughout the course of the utterance. In fact, the occurrence of an accessibility shift may reverse earlier results of the competition among different conceptual representations. In order to offer an account of accessibility shifts, Recanati introduces the notion of a schema, which affects the dynamics of accessibility along the following line:

> One interpretation is the most accessible one, hence wins, at ***s***, but the interpretation fails to fit some schema, hence loses at some later stage ***s'***. (Recanati, 2004, p. 32)

Schemata, or world knowledge structures, are abstract representations of a situation type. Let us clarify this point with the stolen-wallet example:

(7) John was arrested by a policeman yesterday. He had just stolen a wallet.

Let us suppose, for the sake of argument, that the concept of the policeman is the most accessible candidate for "He" at the initial stage of the interpretation of the pronoun. We may further suppose that processing both the predicates "was arrested" and "had stolen" jointly activates a schema in which the two roles (stealing and being arrested) are linked in such a way that the same person steals and is arrested. In virtue of fitting the activated schema, the mental representation of "John" receives extra activation and, as a result, it is more highly activated than its competing candidate. Thus, an accessibility shift occurs and the schema-instantiating interpretation is selected.

This example should provide some insight into the notion of schemata. First, it is worth noticing that the representation given by Recanati of the activated schema is <STEAL (X) – IS ARRESTED (X)>. This clearly suggests that schemata are *abstract* representations of world-knowledge, characterised by a certain degree of generalisation. It seems that what distinguishes a schema from other world

knowledge is that it is not a fully contentful representation. Schemata are abstracted or induced from representations of specific situations which have occurred sufficiently frequently and have sufficient features in common to justify the setting up of a situation template which is potentially useful from a cognitive point of view by allowing the detection and storage of regularities of a particular kind. It is just when the schema is instantiated by relevant semantic values, i.e. when its slots are saturated, that it can be truth-evaluated:

> An expression E activates an *abstract* schema in which *there is a slot for a value of a certain type*; as a result, the semantic value of E will preferably enter into the composition with a semantic value of the relevant *type*. (Recanati, 2004, p. 36, *my emphasis* (DM))

Second, schemata seem to be activated by linguistic material. In the example above, the predicates "was arrested" and "had stolen" jointly activate the schema <STEAL (X) – IS ARRESTED (X)>. This leaves open the issue of whether schemata can be activated by extra-linguistic context alone. Finally, schemata enrich the utterance interpretation in a top-down manner adding material which is not linguistically conveyed. In example (7), for instance, the causal interpretation of the two events mentioned (stealing, and being arrested) is said to be determined by the schema itself. For this reason, <STEAL (X) – IS ARRESTED (X)> is better thought of as an incomplete representation of the schema <STEAL (X) – AND AS A RESULT – IS ARRESTED (X)> which explicitly represents the causal relation between stealing and being arrested. In line with this, interpretation is thought of as a process driven by world-knowledge.

The role played by schemata in utterance comprehension is to promote *coherent* interpretations, where a coherent interpretation is one whose subparts jointly succeed as instances of a certain schema. This role is perfectly compatible with Recanati's accessibility-based account of primary pragmatic processes. Schemata can be evoked by the word used or by other salient features of the speech situation. Furthermore, they drive the interpretation process by making an important contribution to the spread of activation within conceptual networks: an activated schema raises the accessibility of those conceptual representations which fit the schema itself. As a

result, it too gains extra activation due to mutual reinforcement between well fitted representations and the schema they instantiate.

Interestingly, the search for coherence which should guarantee the success of the selected interpretation does not involve the appeal to any kind of pragmatic principle. On the contrary, it is entirely driven by the dynamics of accessibility as it is constrained by schemata. In other terms, coherence is not placed outside of the general dynamics of accessibility, and no further specific (*inferential*) step toward utterance interpretation is required. As suggested in Chapter 1 (Section 1.3.1), the lack of a distinction between the two stages of 'hypothesis formation' and 'hypothesis confirmation' in the interpretive process makes Recanati's account of primary pragmatic processes a non-inferential account in the sense under discussion here.

2.2 Free enrichment: unarticulated constituents and ad hoc concepts

The accessibility-based account proposed by Recanati is supposed to apply to the whole range of 'primary' pragmatic processes, that is, processes which contribute to the recovery of the explicit content of the utterance. These include not only linguistically mandated processes like disambiguation and reference assignment but also optional pragmatic processes, or processes of 'free enrichment', like the provision of unarticulated constituents and lexical adjustment.[7] In the case of unarticulated constituents, pragmatically recovered conceptual constituents are composed into the interpretation of the explicit content of the utterance at a local level; with lexical adjustment or lexical modulation, lexically encoded meanings are modulated in context and the resulting concepts contribute to the explicit content of the utterance. In order to clarify this distinction, we can consider the following examples:

[7] Note that Recanati (2007a) has adopted a relativist position on unarticulated constituents. They are assumed to be part of the situation of evaluation, rather than of the explicitly communicated content. In line with this, his most recent notion of free enrichment is confined to lexical adjustment. In what follows, I address Recanati's (2004) framework, or, more generally, any accessibility-based account of primary pragmatic process which includes the provision of unarticulated constituents.

(8) a. It is raining.
b. IT IS RAINING IN LONDON.
c. I'm not drinking tonight.
d. I'M NOT DRINKING ANY ALCHOL TONIGHT.

Intuitively, sentences (8a) and (8c) can be used, in certain contexts, to express the propositions in (8b) and (8d). The first example is generally treated as a case in which the propositional constituent *[IN LONDON]*, which is linguistically unarticulated, is added to the proposition intuitively expressed by uttering (8a) in some particular contexts (e.g. when the speaker is looking outside the window of her flat in London). By contrast, the second example is often assumed to be a case of lexical adjustment. This process of enrichment takes as input the linguistically encoded concept DRINK, and delivers as output an ad hoc, or occasion-specific, concept DRINK* (whose denotation is narrower than the previous one, e.g. it does not include drinking water).

These optional 'primary' pragmatic processes deserve further attention since, *prima facie*, they do not appear to be reducible to Recanati's standard explanation in terms of activations-and-associations. As far as the provision of unarticulated constituents is concerned, Carston (2007) comments:

> In this regard, there is a very little to go on in Recanati's discussion, which concentrates on cases of reference assignment and disambiguation. The usual talk of competing candidates doesn't seem to make much sense when the constituent is not linguistically articulated – what would they be competing for? Since there is no variable or slot in the structure requiring that a value be provided, there seems to be nothing to motivate the process. [..] it is difficult to see how an interpretation which involves the linguistically unmotivated addition of a constituent of content could be more accessible than one that does not. (Carston, 2007, p. 31)

This passage highlights several distinct issues. First, the provision of unarticulated constituents does not seem to involve any genuine competition among alternative semantic values. Given the local, subpropositional nature of primary pragmatic processes, this competition should take place at a local level, that is, among alternative semantic candidates for a single expression (e.g. competing meanings, either literal or derived, for a single word or phrase). However, the process underpinning the provision of unarticulated constituents does not fit this overall

picture: it is not clear which are the competing candidates, and what they are competing for. In fact, they are not candidates for the semantic value of a linguistic expression belonging to the sentence uttered. At most, it seems to involve a competition among alternative overall interpretations, but this feature is doubtfully coherent with its local character. Second, Carston raises the question of what triggers these optional processes in Recanati's account. If there is no linguistic expression which needs to be either modulated in context or contextually assigned a semantic value, it is not obvious what motivates the process itself. How can top-down pragmatic considerations enter the picture without exceeding the dynamics of accessibility? The challenge is to explain how sheer accessibility, without any further pragmatic motivation such as the search for informative/relevant interpretations, explains the recovery of interpretations characterised by linguistically unmotivated elements.

In an attempt to address Carston's objection, it is worth developing Recanati's accessibility-based account of free enrichment in a way that goes beyond his own explicit considerations. As noted above, Recanati claims that schemata can enrich the interpretation in a top-down manner and add material which is not linguistically conveyed:

> The role played by schemata explains why the process of utterance interpretation is to such a large extent top down and driven by world knowledge. The interpreter unconsciously *enriches the situation described by the utterance with many details which do not correspond to any aspect of the uttered sentence but are contributed in order to fit an evoked schema*. Thus, in the policeman example above, not only is the reference of 'he' determined by the STEAL (X) – IS ARRESTED (X) schema, but (among other things) the causal interpretation between the two events mentioned – the fact that the referent was arrested because he had stolen – is also determined by the schema. (Recanati, 2004, p. 37, *my emphasis* (DM))

Although this passage is not explicitly related to the phenomenon of unarticulated constituents, it is worth considering it here as it may provide a way out of Carston's objections. Specifically, it seems to suggest that it is the activation of schemata during utterance comprehension that could *motivate* the free enrichment of the

explicit meaning of the utterance. To clarify this point, let us consider the following example:

(9) a. He insulted her and she hit him.
b. HE INSULTED HER AND THEN, AS A RESULT, SHE HIT HIM.

Both Recanati and Carston would agree that the sentence (9a) can be used, in certain contexts, to express the proposition in (9b). We could imagine that the predicates "insulted" and "hit" jointly activate a schema of this kind < INSULTING (X,Y) -THEN, AS A RESULT- HITTING (Y,X) >. This schema not only provides an interpretation in which the same person insults and is hit (and the same person is insulted and hits), but enriches the proposition intuitively expressed by adding a temporal and causal interpretation.

The activation of a certain schema would not only explain how free enrichment is motivated (the temporal and causal interpretations are added in order to fit the schema itself) but also how the intended interpretation is derived within an accessibility-based account. In fact, the activation of a schema in turn activates (or adds extra activation to) those components which fit the schema. In the example discussed, the evoked schema < INSULTING (X,Y) -THEN, AS A RESULT- HITTING (Y,X) > both activates those mental representation which do not correspond to any aspect of the sentence uttered (THEN, AS A RESULT) but are part of the schema, and adds further activation to the fitting semantic candidates for X and Y. Thus, an overall schema-instantiating interpretation will tend to be selected in virtue of the mere dynamics of accessibility.

To assess the explanatory power of the analysis provided, it is worth asking whether schemata can motivate every instance of free enrichment. Let us consider the following example.

Context: Bob and Alice have got married recently. Peter and Bob are discussing Bob and Alice's decision to get married after many years of living happily together as a couple. It is (mutually) known to the interlocutors that Bob and Alice did not value the official institution of marriage.

(10) a. *Peter:* Weren't you against formal marriage contracts?
b. *Bob:* We had two children and we got married.
c. WE HAD TWO CHILDREN AND THEN, AS A RESULT, WE GOT MARRIED.
d. WE CHANGED OUR MIND AFTER HAVING TWO CHILDREN.
e. HAVING CHILDREN HAS PUSHED US TO RECONSIDER OUR PREVIOUS CONVICTIONS.
f. BECOMING PARENTS HAS INVOLVED AN IMPORTANT CHANGE IN OUR LIVES AND IN THE WAY OF THINKING ABOUT OUR RELATIONSHIP.

In the context at hand, Bob can utter (10b) with the intention to express the proposition (10c). The latter is freely enriched with a temporal and causal interpretation according to which the event described in the first sentence (Bob and Alice having children) is interpreted as preceding and causing the event described in the second conjunct (Bob and Alice getting married). It is the recovery of (10c), rather than the minimal proposition WE HAD TWO CHILDREN & WE GOT MARRIED, which allows the addressee to derive the implicatures listed above ((10e), (10d) and (10f)). For the sake of the argument, let us suppose that Peter does not know why Bob and Alice decided to get married. Thus, Peter does not have any *particular* world-knowledge related to them and the details of their life together. According to the account of free enrichment which we are trying to assess, the added temporal and causal interpretation should be provided by some schema, conveying either particular or general world-knowledge. Having excluded the first source of enrichment (particular schemata), we have to consider if there is some activated general schema which could provide the intended interpretation. It is really unlikely that there is a schema activated by the predicates "had children" and "got married" which provides the relevant temporal and causal order in the example at issue. On the contrary, it is much more likely that there is a schema, *if any*, priming the opposite interpretation.

The example above should cast some doubt on the explanatory power of schema-instantiation as the source of free enrichment. In fact, it seems that at least some cases of free enrichment tend to resist an explanation of this kind. In this case, the recovery of the explicature (10c) is effect-driven, that is, the addressee has quite specific expectations about the implicatures to be derived (an explanation of why the addressee decided to get married) and he backwards adjusts the content of the explicature to warrant them. This procedure, which Relevance Theory has called

"mutual parallel adjustment", seems to play an important role in motivating free enrichment, especially in those cases in which schemata explanations fail.

Let us now turn to a different issue, which we may call the 'overgeneration' issue, that is, the claim that the schemata-account of free enrichment does not manifest any principled criterion according to which the intended enrichment is selected. As we have previously noticed, an activated schema can embody many details which do not correspond to any aspect of the uttered sentence. Indeed, the fact that a schema contains more conceptual elements than those offered by the processes of decoding the linguistically encoded meaning can justify the enrichment of the latter by means of the former. In principle, a schema can embody many more details than the ones which are actually added to the interpretation. But what could prevent that extra material from entering the intended interpretation? Let us consider the following example:

Context: Peter and Bob are discussing Bob's relationship with Alice. It is (mutually) known to the interlocutors that Bob's family members are rather traditionalist, religious and old fashioned.

(11) a. *Peter:* Did your family approve of your relationship?
b. *Bob:* Yes, we followed the traditional stepwise procedure! We got married and we had two children.
c. WE GOT MARRIED AND, *THEN*, WE HAD TWO CHILDREN.
d. MY FAMILY APPROVED OF OUR RELATIONSHIP SINCE WE FIRST GOT MARRIED AND THEN HAD TWO CHILDREN.

It is likely that the predicates "got married" and "had two children", plus the adjective "traditional", activates a schema, which we could label "traditional married life", of the following kind < GETTING MARRIED IN A RELIGIOUS CEREMONY – AND, THEN – HAVING CHILDREN >. Thus, we could be tempted to conclude that the extra material "IN A RELIGIOUS CEREMONY" enters into the explicature of Bob's utterance: WE GOT MARRIED *IN A RELIGIOUS CEREMONY* AND, *THEN*, WE HAD TWO CHILDREN. But it does not seem that this further enrichment has to be recovered in order to grasp the speaker's meaning. There is simply a very strong intuition that the explicature of the utterance does not include that constituent of content. Furthermore, unlike the unarticulated constituent *THEN*, it does not play any role in the recovery of the

implicature. It is worth noting that the intended implicature is easily recovered on the basis of the expectation raised by Bob's mention of "a traditional stepwise procedure". In fact, this expression is likely to raise the expectation that the family's approval of the relationship is grounded on the temporal order according to which he decided to go through the important steps of a life together.

As the previous examples have shown, expectations of relevance (and, indirectly, mutual parallel adjustment) seem to play a decisive role in governing the process of free enrichment. Not only do they drive the process itself in a top-down manner, but they also provide a reliable criterion to select the kinds of enrichment that should enter the interpretation process. In line with this, an enrichment is needed (and embodied within the interpretation) when it plays a significant role in warranting the intended implicatures. It is significant that Recanati himself resorts to mutual parallel adjustment to capture the relationship between explicature and implicature(s) when the former is characterised by some kind of free enrichment:

> In each case we may suppose that the speaker implies various things by saying what she does. Thus, by saying that she's had breakfast, the speaker implies that she is not hungry and does not want to be fed. By saying that the child is not going to die the mother implies that the cut is not serious; and so forth. Now those implicatures can be worked out only if the speaker is recognized as expressing the (non-minimal) proposition that she's had breakfast *that morning*, or that the child won't die *from that cut*. Clearly, if the speaker had had breakfast twenty years ago (rather than that very morning), nothing would follow concerning the speaker's present state of hunger and her willingness or unwillingness to eat something. The implicature could not be derived, if what the speaker says was not given the richer, temporally restricted interpretation. (Recanati, 2004, pp. 21-22).

To sum up, it appears that Recanati's account of primary pragmatic processes needs to be developed in order to provide a full explanation of how free enrichment is supposed to work. One way of doing this is by extending the role played by schemata in the process of providing unarticulated constituents. While this suggestion is compatible with Recanati's remarks on the nature and function of schemata and follows straightforwardly from them, it falls short in those cases where mutual parallel adjustment seems to be required. Importantly, while the relevance-theoretic inferential account of primary and secondary pragmatic processes can easily

accommodate the idea that such processes operate in parallel and mutually shape their outputs, as Carston (2007, pp. 24-27) emphasises, it is much less clear how a dual pragmatic system, such as Recanati's, can be implemented with the notion of 'mutual parallel adjustment'. This implementation is in need of an explanation with regard to how the two distinct types of processes can interact with each other. The question of how a global inferential process can shape the input of associative processes at a local level is, however, left unaddressed by Recanati.

2.3 Experimental evidence

2.3.1 Free enrichment and emergent properties

In this section, I discuss and evaluate some experimental work that has been recently conducted (Rubio-Fernández, 2007, 2013) in order to contribute to the debate between inferential vs. non-inferential approaches to pragmatics. My main focus is Rubio-Fernández (2013), which represents the first attempt to subject the fine-grained differences between inferential and accessibility-based approaches to pragmatics to rigorous empirical investigation. The focus of the study is the phenomenon of free pragmatic enrichment, which both Recanati and Relevance Theory take to contribute to the level of explicit utterance content, hence a primary pragmatic process.

Rubio-Fernández (2013) investigates the processes that account for instances of pragmatic enrichment "where the reader goes beyond the message linguistically encoded and derives a more specific interpretation as to *why* or *how* a certain action is performed" (Rubio-Fernández, 2013, p. 725). Although it is not explicitly stated, some of the examples that make up the materials of her experiment seem to fall within the category of unarticulated constituents. For instance, let us consider the following examples discussed throughout the paper:

(12) a. John didn't know how to swim, so when he fell into the water, his best friend threw him a lifebuoy *[TO SAVE HIM FROM DROWNING*].
b. John didn't know how to swim, so when he fell into the water, his best friend threw him a basketball *[TO SAVE HIM FROM DROWNING*].

In both cases, it seems difficult to interpret the enrichment *TO SAVE HIM FROM DROWNING* as the result of the modulation of a lexically encoded component of the sentence (which word would it be?). Thus, I assume these cases fall within the 'unarticulated constituents' category. However, as Rubio-Fernández (2013, p. 734) underlines, "the focus of the present study is not the pragmatic process of free enrichment *per se*". Rather, it concerns the accessibility of 'emergent properties' (i.e. context-dependent properties) and, only as a consequence, those instances of free enrichment which are based on the recovery of such properties. Let us then focus on the distinction between context-independent and context-dependent information (Barsalou, 1982). Context-independent information is automatically accessed every time the corresponding word is processed. By contrast context-dependent information is activated only when contextually relevant. The former is said to be accessed via spreading of activation of associates, the latter through local inferential processes (Barsalou, 1982, p. 87).[8] Thus, if we consider the word *basketball*, the property ROUND counts as a context-independent property, which gets automatically activated whenever the word is processed, whereas FLOATS is an emergent property, which is accessed just in relevant contexts, e.g. (12b).

This distinction – and the claimed difference in terms of underpinning cognitive mechanisms – provides the basis for an empirical test of Relevance Theory and Recanati (2004). The two models make different predictions with regard to circumstances in which inferential processes are required at a local, rather than global, level to contribute to the recovery of the explicit content of the utterance. For instance, in order to enrich the utterance of (12b) with the unarticulated constituents *TO SAVE HIM FROM DROWNING*, the interpreter needs to access the emergent property FLOATS of the word 'basketball'. This, according to Barsalou (1982), requires a local inferential process to take place. According to Recanati,

[8] It is worth noticing that Barsalou does not provide any detailed description of (this) alleged inferential process: "The range and nature of these inferential processes are other topics worthy of future discussion" (Barsalou, 1982, p. 87). This raises the issue whether these processes could fall under the notion of 'broad inference' put forth by Recanati (2002), which is still compatible with an accessibility-based account of them.

> the primary pragmatic processes involved in comprehension are not 'inferential'. Only when the unreflective, normal processes of interpretation yield weird results does a genuine inference process take place whereby we use evidence concerning the speaker's beliefs and intention to work out what he means. (Recanati, 2004, p. 38)

This passage suggests that the normal flow of interpretation is disrupted when automatic, purely associative and blind processes fail in delivering a satisfactory interpretation. As a consequence, inferential processes are exploited in order to recover the speaker's meaning on the basis of richer considerations involving the speaker's mental states. In contrast, Relevance Theory advocates a uniformly inferential account of pragmatic processes, which does not only come into play to supplant a more basic system when it goes wrong. Inferential processes affect utterance interpretation from the very beginning, without involving any disruption.

Rubio-Fernández (2013), therefore, suggests that Recanati (2004) would predict emergent properties to be accessed in a second (inferential) stage of processing – after the breakdown of the associative mode of processing. Relevance Theory, by contrast, would account for their accessibility within the normal flow of interpretation.

2.3.2 Experimental setting and results

Rubio-Fernández's (2013) study employs a self-paced reading task, in which participants read segments of sentences, revealed one-at-a-time on the monitor, and decide when to proceed by pressing a button. The material comprised a series of short passages, all of which were presented in three versions, each to a different participant. The passages differed with regard to the familiarity and appropriateness of the target word. A familiar target word is one that is primed by the passage (through semantic associations and world-knowledge priming). An appropriate target word is one that is relevant to the context. The following is an example of a critical passage (in its three versions):

> John went to the public swimming pool every afternoon. He spent most of his time there reading the papers. He didn't know how to swim, so when he fell into the water, his best

friend threw him **a lifebuoy**/ **a basketball**/ **a newspaper** and then just waited by the side of the pool. John was expecting a bit more help.

The three versions correspond to (i) the familiar and appropriate target (e.g. *lifebuoy*), (ii) the unfamiliar but appropriate target (e.g. *basketball*) and (iii) the familiar but inappropriate target (e.g. *newspaper*).

In order to test the predictions sketched in the previous section, two measures of word processing were recorded: an early measure to reveal reading latencies for the critical target phrase (e.g. *a lifebuoy*), and a late measure of the reading times for the segment following the critical target (e.g. *and then just waited*). Reading latencies at this later stage are supposed to reflect any delay in integration processes.[9]

As far as the early processing measure is concerned, both Relevance Theory and Recanati would predict a priming effect in the familiar-and-appropriate condition (e.g. *lifebuoy*) as well as in the familiar-and-inappropriate condition (e.g. *newspaper*). The concept LIFEBUOY would be activated by the current DROWNING scenario, whereas the concept NEWSPAPER would be primed by the earlier part of the discourse (i.e. *He spent most of his time there reading the papers*). Crucially, according to Rubio-Fernández, the two models exhibit different predictions with regard to the unfamiliar-and-appropriate condition (e.g. *basketball*). Recanati (2004) would not predict any priming. The active scenario (i.e. DROWNING scenario) could not drive the enrichment automatically since it does not contain any slot for the concept BASKETBALL. By contrast, Relevance Theory would predict a significant facilitation effect (though not through priming). The latter is consistent with the view that "backwards inferences" can modulate the recovery of the explicit content of the utterance. In other terms, the explicit content would be constructed in such a way that it satisfies *specific* expectations about the relevance of the utterance (e.g. the expectation that the relevance of (12b) relies on the cooperativeness of John's friend). Analogously, with regard to the late processing measure, the only condition

[9] A control list where the targets appeared in a neutral context was used for the early processing measure. Since the targets differed in length and frequency a direct comparison of their levels of activation was not possible.

which Rubio-Fernández considers as critically differentiating Recanati's model and Relevance Theory is the unfamiliar-and-appropriate one (e.g. *basketball*). Recanati, but not Relevance Theory, would predict a disruption of the normal flow of interpretation due to the breakdown of the associative mode of processing and this should be evident in the attempt to integrate the interpretation of the word into the ongoing discourse.

The results supported the relevance-theoretic framework in both cases. In particular, they showed that unfamiliar-but-appropriate targets were significantly facilitated early in processing relative to the control condition (i.e. neutral context), with reading times comparable to the two familiar conditions. Furthermore, their integration in the discourse representation did not disrupt the normal flow of interpretation. The conclusion drawn by Rubio-Fernández (2013, p. 739, *my emphasis* (DM)) is that "inferential processes are *fully integrated* in the processing system, operating not only at the global level of the utterance but also at the *local level*".

In what follows I evaluate the extent to which the experimental study supports this conclusion. In fact, there are two claims being made here which need to be distinguished: first, that inferential processes are fully integrated in the processing system (i.e. they operate in parallel with associative processes); second, that they operate at a local level without causing any disruption. I argue that while the first claim is certainly supported by the present study it is doubtful that it provides evidence for the second claim.

It is worth noting that it is this second claim that mostly differentiates between Recanati (2004) and Relevance Theory. Recanati (2004, p. 47) has indeed argued for a dual system, characterised by a distinction between an associative and an inferential mode of processing, but one which is not to be interpreted as a sequential model. Thus, he would endorse the idea of a full integration of the inferential mechanisms, so far as this is confined to global inferences (e.g. implicature derivation). He does not, however, support the view that (local) inferences are involved in the primary processing system.

2.3.3 Pragmatic enrichments and local inferences

Rubio-Fernández (2013) adopts the distinction between context-independent and context-dependent properties proposed by Barsalou (1982). In line with Barsalou's description, she takes the first to be accessed via spreading of activation (hence an associative process), and the second through local inferential processes. This correspondence between the distinction and the two alternative modes of access is more presupposed (possibly in virtue of its intuitive plausibility) than explained.

In a pre-test of the material used in the study, Rubio-Fernández combined a self-paced reading task with a property verification task. Participants were presented with short passages in which the same target word was embedded either in a relevant or in a neutral context. They were asked to answer a world-knowledge question at the end of each passage. The pre-test was aimed at determining whether the comprehension of the critical (i.e. relevant) context gave access to the intended emergent property. For this purpose, in the critical condition, the question was about a relevant emergent property of the target word, whereas, in the control condition, it was related to the same emergent property of the target word, irrelevant in the context at issue:

> *Critical condition:* John went to the public swimming pool every afternoon. He spent most of his time there reading the papers. He didn't know how to swim, so when he fell into the water, his best friend threw him **a basketball** and then just waited by the side of the pool. John was expecting a bit more help.
>
> *Control condition:* Tom was wondering what to spend his birthday money on. He always ended up getting sports equipment but he decided to get something else this year. He had first thought of getting **a basketball** to join the local team, but maybe he could get a new computer game instead.
>
> *World-knowledge question:* Can a basketball be used to stay afloat?

The results showed that participants were faster in verifying the emergent property in the critical rather than in the control condition. This difference in response times is interpreted by Rubio-Fernández (2013) not only as evidence that the emergent properties were effectively accessed in the critical context, but also that they were not accessed through a chain of automatic associations. The suggestion is that not

only did participants effectively access the emergent property CAN BE USED TO STAY AFLOAT, but also that they accessed it through a local inferential process. This further conclusion is advanced by Rubio-Fernández on the basis of the following consideration: "If participants were equally fast at verifying the emergent properties in both types of context, it could be argued that emergent properties were accessed through an automatic chain of associations *regardless of their contextual relevance*." (Rubio-Fernández, 2013, p. 730, *my emphasis* (DM)). Such an automatic chain of associations could be something along the following line: BASKETBALL → FILLED WITH AIR → FLOATS → CAN BE USED TO STAY AFLOAT. However, since the response times were significantly different, with a strong facilitation in the critical condition, Rubio-Fernández concludes that an associationist explanation is ruled out.

I doubt that the evidence provided is enough to support this stronger claim and I think that Recanati (2004, 2007b) could accommodate the same pattern of results within his own framework. General world-knowledge structures, or 'schemata', and also particular world-knowledge can constrain the dynamics of accessibility and drive the interpretation in the expected direction. The activation of general and particular world-knowledge, however, is not context-independent. Thus, these kinds of information play a significant role when they get activated, i.e. when they are contextually relevant.

In the control condition, the target word is embedded in a neutral context with regard to which the (emergent) property to be verified is irrelevant (incoherent, in terms of Recanati's model). In such a context, Recanati would not predict the same facilitation that occurs in the critical condition. The latter could be characterised by the activation of schemata of different sorts, which would remain inactive in the control condition. Thus, longer response times in the control condition are expected by both an inferential account and an associative account, at least when the latter is supplemented with the notion of schemata.

2.3.4 Pragmatic enrichments or implicatures?

The present study focuses on some particular instances of free enrichment in which the interpretation is enriched with respect to *how* or *why* a certain action is

performed. For instance, the interpretation of (12b) is enriched with respect to why the action of throwing a basketball to John is performed, namely *TO SAVE HIM FROM DROWNING*.

(12) b. John didn't know how to swim, so when he fell into the water, his best friend threw him a basketball *[TO SAVE HIM FROM DROWNING*].

It is worth pointing out that Rubio-Fernández (2013) does not borrow this characterisation from established accounts of free enrichment in the literature (e.g. Carston, 2002; Hall, 2014; Recanati, 2002b, 2004). These accounts barely specify the range of free enrichments that can potentially enrich the interpretation. Rather, they either confine themselves to the analysis of particular instances of free enrichment (e.g. the causal or temporal enrichment of 'and' conjunction) or they provide some general constraints on it (e.g. the 'locality constraint'). This underspecification – I suspect – is not unintentional. Once the context-specific nature of free enrichment is properly appreciated, it is this very nature that prevents us from filling out a complete list of its possible occurrences.

Rubio-Fernández (2013) seems to make a very strong claim about the pragmatic enrichments discussed. In particular, her claim could be interpreted as supporting the idea that whenever the interpretation is enriched with respect to how or why a certain action is performed, this is an instance of free enrichment. This idea, however, is certainly wrong. Several non-controversial instances of implicature can be described as providing an answer to why a certain action is performed. For instance, let us consider this paradigmatic example from Grice (1975). Andy and Barbara are gossiping about their common friend Smith, who lives in Boston.

(13) a. *Andy*: Smith doesn't seem to have a girlfriend these days.
b. *Barbara*: He has been paying a lot of visits to New York lately.

Barbara implicates that Smith has, or may have, a new girlfriend in New York. This implicature enriches the interpretation of Barbara's utterance suggesting a (relevant) reason for Smith travelling so frequently to New York. This seems to fall within the description provided by Rubio-Fernández (2013, p. 725), i.e. "pragmatic enrichments where the hearer goes beyond the message linguistically encoded and derives a more

specific interpretation as to *why* and *how* a certain action is performed", but it is accepted by everyone in pragmatics as a case of implicature rather than an instance of free enrichment.

More importantly, this treatment is further supported by several tests, which have been introduced in the pragmatic literature in order to help with drawing the distinction between 'what is said' (i.e. explicature) and 'what is implicated' (i.e. implicatures) which is not always straightforward to see. Let us consider two varieties of the embedding test: the negation embedding test and the conditional embedding test (both subsumed by the so-called "Scope Principle"). These tests allow us to distinguish between those pragmatically derived meanings which fall under the scope of a logical operator (e.g. negation, conditional) and those which do not.[10] The former are supposed to contribute to the truth-conditional content of the utterance (i.e. its explicature), the latter to its implicatures. Let us apply these tests to (13b):

(13) a. *Andy*: Smith doesn't seem to have a girlfriend these days.
b. *Barbara*: He has been paying a lot of visits to New York lately.
c. He has not been paying a lot of visits to New York lately.
d. If he pays a lot of visits to New York lately, he must spend a lot of money.

The pragmatic inference that Smith has, or may have, a girlfriend in New York does not fall within the scope of the negation in (13c), or in the scope of the conditional in (13d). The consequent of the conditional (i.e. that Smith must spend a lot of money) depends just on the proposition that Smith pays a lot of visits to New York and not on the additional inference that he may have a girlfriend there.

These considerations should suggest that specifying the reason *why* a certain action is performed is not a sufficient condition for a pragmatic enrichment to qualify as an instance of free enrichment. On this basis, I question the idea that all of the pragmatic enrichments discussed in Rubio-Fernández (2013) are genuine cases of free enrichments rather than examples of implicatures (although some of them may

[10] See Carston (2002, pp. 191-197) for a detailed description of these types of test.

well be). Let us apply the two varieties of the embedding test introduced above to the example (12b):

(12) b. John didn't know how to swim, so when he fell into the water, his best friend threw him a basketball *[TO SAVE HIM FROM DROWNING*].
c. His best friend didn't throw him a basketball.
d. If his best friend threw him a basketball, he must have picked it up from the basketball team equipment left in the corner.

From (12b) the addressee can pragmatically infer that John's best friend was probably trying to save him from drowning. But, similarly to the pragmatic inference derived from (13b), this inference does not fall either in the scope of the negation in (12c), or in the scope of the conditional in (12d). Once again, the consequent depends just on the proposition that John's best friend threw him a basketball and not on him having the intention of saving John.

If we further appeal to intuitions and introspective analysis, the pragmatic enrichment characterising (12b) should equally fall on the implicit side. Intuitive judgments have been adopted as a criterion to classify pragmatic enrichments by Recanati under the name of the "Availability principle":

> Availability principle: In deciding whether a pragmatically determined aspect of utterance meaning is part of what is said, that is, in making a decision concerning what is said, we should always try to preserve our pre-theoretic intuitions on the matter. (Recanati, 1989, p. 310, 1993, p. 240)

The notion of explicature is usually conceived of as the intuitive truth-conditional content of the utterance, i.e. the basis on which we would intuitively judge an utterance to be true or false. Thus, an appeal to intuitions, especially in those cases when they are clear and consistent, is grounded within the characterisation of the notion of explicature itself. If we consider (12b), it seems that its intuitive truth conditional content does not include the (alleged) unarticulated constituent *[TO SAVE HIM FROM DROWNING]*. The utterance would be intuitively true even if John's best friend threw him a basketball with the intention of teasing him (rather than saving him). The intention behind the action does not seem to play any role in the assessment of the truth-value of the utterance at issue. These considerations suggest

that the enrichment at issue should fall within the non-truth-conditional communicated content of the utterance, i.e. its implicatures.

The evidence provided so far in favour of an implicature analysis of the pragmatic inference triggered by (12b) seems to find further confirmation in Hall (2014), which is an investigation of the pragmatic constraints on free enrichment, i.e. the ways in which context can or cannot affect explicature. Among them, we find the 'locality constraint', which is motivated by the inferential procedure underpinning utterance interpretation. As a local process, free enrichment modifies subparts of the linguistic logical form. It follows that "any material that can stand alone will not be composed into explicature: this includes independent propositional forms, and semantic arguments or predicates such as NP- or VP-conjuncts which, in context, would likely be derived from fully propositional premises, or if not can be straightforwardly propositionalized" (Hall, 2014, p. 19). The (alleged) unarticulated constituents which enrich the interpretation of (12a) and (12b), *[TO SAVE HIM FROM DROWNING]*, do not fall under the categories of NP- or VP- conjuncts but, as an infinitive phrase, they can be easily propositionalised. Thus, they do not satisfy the 'locality constraint'.

While these arguments support a reanalysis of part of Rubio-Fernández's experimental materials (i.e. the why-cases of pragmatic enrichments) in terms of implicature, it is worth noticing that how-cases seem to be more likely to give rise to lexical adjustment, genuinely affecting the derivation of the *explicit* content of the utterance. When the enrichment involves specifying the mode in which a certain action is performed, this specification can be easily seen as the result of a process of constructing an occasion-specific concept with regard to the context at hand. Let us consider this further example, which is one of the materials used by Rubio-Fernández:

> Lucy's grandmother was giving a big party tonight. She was cooking several dishes and baking bread. Now she was taking a break sitting at the kitchen counter. Lucy wanted to help, so she started working the bread dough **with a rolling pin**/**with an empty bottle**/**with an empty stool** and preheated the oven. She was full of ideas when it came to cooking.

In this case, the enrichment could be seen as involving the concept WORKING. The lexically encoded concept could be modulated (e.g. its denotation could be narrowed) into the concept WORKING*, which specifies a particular mode of using the rolling tool (e.g. *rolling pin* or *empty bottle*). The concept WORKING* would thus denote only those instances of working that involve some stretching out of the item worked.

The difference between why-cases and how-cases is certainly worth further exploration. To begin with, I explore the consequences of a reanalysis of the why-cases used in the study as instances of implicature. In particular, I suggest that this reanalysis fits better with the relevance-theoretic notion of mutual parallel adjustment among explicatures, implicated premises and implicated conclusions.

According to Relevance Theory, the comprehension procedure subsumes three different sub-tasks as it concerns the construction of appropriate hypotheses about explicit content, intended contextual assumptions (in relevance-theoretic terms, implicated premises) and intended contextual implications (or implicated conclusions). These sub-tasks are not sequentially ordered. Thus, the interpreter is not required to *first* recover the explicit content of the utterance, *then* select a useful range of contextual assumptions, and *finally* derive the intended contextual implications. In some circumstances, the comprehension procedure can be effect-driven: the occurrence of highly constrained expectations about the intended implications (i.e. implicated conclusions) can affect the recovery of explicatures or implicated premises in such a way that the premises are constructed with the purpose of warranting the intended effects:

> In particular, the hearer may bring to the comprehension process not only a general presumption of relevance, but *more specific expectations* about how the utterance will be relevant to him (what cognitive effects it is likely to achieve), and these may contribute, *via backwards inference*, to the identification of explicatures and implicated premises. (Wilson & Sperber, 2004, p. 615, *my emphasis* (DM))

The notion of backwards inference is explicitly adopted by Rubio-Fernández (2013) in order to explain the facilitation of unfamiliar-but-appropriate targets (e.g.

basketball). In this case, in fact, Rubio-Fernández argues that facilitation cannot be explained on the basis of global contextual priming (i.e. priming due to world-knowledge structures) since the target word is unfamiliar. In other terms, it is not plausible to assume that the target word *basketball* could be primed by a world-knowledge structure such as the 'DROWNING SCENARIO', which is unlikely to include a slot for BASKETBALL (VS. LIFEBUOY). With regard to example (12b), the idea is that at the time of processing the segment 'a basketball', "the reader has already understood that John's best friend was throwing him a life-preserver" (Rubio-Fernández, 2013, p. 734), on the basis of the expectation that the relevance of the utterance concerns cooperativeness and helpfulness from John's best friend.

While I accept the idea that some kind of backwards inference is at play, I argue that the overall interpretation requires a different characterisation of the three subtasks involved (i.e. recovering the explicature, the implicated premise and the implicated conclusion) in order to fit the relevance-theoretic notion of mutual parallel adjustment. For this purpose, I first report Rubio-Fernández's (2013) proposal concerning what constitutes the implicated premise and conclusion of the interpretation of (12b), and then I sketch an alternative configuration of the overall interpretation.

In discussing example (12b), Rubio-Fernández refers to "the *implicated premise* that John's best friend was throwing him a life preserver" (Rubio-Fernández, 2013, p. 725, *my emphasis*) and to "the *implicated conclusion* that John's best friend threw him a basketball to save him from drowning" (Rubio-Fernández, 2013, p. 729, *my emphasis*). I think two different considerations are in order. First, in the attempt to fit these characterisations into a (valid) argument we would need to adopt as further premises both the proposition based on the linguistically encoded meaning of (12b), 'John's best friend threw him a basketball', and the contextual assumption that a basketball can be used as a life preserver:

(14) *Premise 1:* John's best friend threw him a basketball.
Premise 2 (implicated premise): John's best friend was throwing him a life preserver
Premise 3: A basketball can be used as a life preserver
Conclusion (implicated conclusion): John's best friend threw him a basketball

to save him from drowning.

This argument is assumed to be the result of a mutual parallel adjustment between the proposition based on the linguistically encoded meaning, the contextual assumptions and the implicated conclusion. It is worth noticing, however, that the implicated conclusion represents what Rubio-Fernández (2013) takes to be the explicature of (12b). Thus, on her account, the kind of adjustment at play is not, after all, the mutual parallel adjustment between explicature, implicated premise and implicated conclusion advocated by Relevance Theory. Furthermore, it seems to conflict with Hall's (2014, p. 18) claim that "between logical form and explicature [...] there is no relation of logically valid inference". The reason proposed by Hall (2014) is that free enrichment, *qua* local process (i.e. operating over subpropositional constituents), delivers unarticulated constituents that are recovered on the basis of their high accessibility and confirmed in so far as they contribute to the overall intended interpretation. Second, Rubio-Fernández's terminology seems to suggest that the enrichment *TO SAVE HIM FROM DROWNING* is to be treated as an implicated conclusion. If we take this suggestion seriously, according to Relevance Theory, this 'enrichment' should count as an implicature (rather than as a constituent of the explicature).[11] Most importantly, this line of interpretation provides us with a straightforward inference which via further process of mutual parallel adjustment will be ultimately accepted:

(15) *Premise 1:* John's best friend threw him a basketball.
Premise 2 (implicated premise): A basketball can be used to stay afloat.
Conclusion (implicated conclusion): John's best friend was trying to save him from drowning.

In this derivation, the contextual assumption, which results from accessing the relevant emergent property of BASKETBALL (i.e. CAN BE USED TO STAY AFLOAT), in

[11] This approach seems to be inconsistently adopted in the following passage: "It therefore seems that the *global implication* that John's friend was trying to save him to some extent drives the search for relevant properties at the local conceptual level." (Rubio-Fernández, 2013, p. 740, *my emphasis* (DM)). It is worth noticing that the result of a local process of free enrichment cannot be a global implication.

conjunction with the explicature, inferentially grounds the expected contextual implication (i.e. implicature).

To sum up, considerations proposed so far support an account of the 'enrichments' at issue in terms of implicatures (at least with regard to why-cases). If my analysis is correct, then Rubio-Fernández's (2013) results call for a reinterpretation. As I mentioned in passing, Recanati (2004) explicitly endorses the relevance-theoretic notion of mutual parallel adjustment, maintaining that the associative and the inferential modes of interpretation do not have to operate sequentially. Their parallel activity, and their mutual shaping, enables inferential processes to contribute to utterance interpretation from an early stage in processing. However, while inferential processes operate at a global level, associative processes intervene at a local level. Associative and inferential processes are, thus, conceived of as parallel systems taking different 'scopes'. If we assume (15) to be the intended overall interpretation, accessing the emergent property CAN BE USED TO STAY AFLOAT corresponds to accessing the contextual assumption which warrants the implicature *JOHN'S BEST FRIEND WAS TRYING TO SAVE HIM FROM DROWNING*. A process of backward inference could thus facilitate the accessibility of the emergent property FLOATS of the word *basketball*. Expectations about the intended implicature (grounded on the more general expectation about helpful behaviour in the context at hand) could affect the recovery of premises so that they are constructed with the purpose of warranting the intended effects. The inferential mechanism described so far is entirely compatible with Recanati's (2004) account of implicature derivation and exhibits a global, rather than a local, character. When it comes to implicature derivation, both Recanati and Relevance Theory describe it as a genuinely inferential process, which may involve several instances of forwards and backwards inferences.

In light of the suggested reanalysis, Rubio-Fernández's (2013) empirical study does not seem to be able to choose between the competing accounts of primary pragmatic processes given by Recanati (2004) and Relevance Theory. While it reinforces the existing strong evidence in favour of a fully integrated view of inferential

mechanisms in pragmatic processing,[12] it falls short in its assessment of the contribution of (alleged) inferential processes at the level of what is explicitly communicated. However, it is worth emphasising that a more extensive analysis of the materials used in the experiment would be needed in order to assess the extent to which Rubio-Fernández's results are undermined by the proposed reanalysis. A clearer distinction between why-cases and how-cases, as noted above, would be crucial for this evaluation. It would potentially be interesting to isolate the data that resulted from the how-cases and assess their statistical significance. Results analogous to the one at issue, but confined to how-cases, would not be subject to some of the critical remarks developed above (i.e. the implicature treatment). However, they would still face the issue concerning the necessity for stronger evidence in favour of an inferential account of the accessibility of emergent properties. For instance, let us consider the rolling-pin example again:

> Lucy's grandmother was giving a big party tonight. She was cooking several dishes and baking bread. Now she was taking a break sitting at the kitchen counter. Lucy wanted to help, so she started working the bread dough **with a rolling pin**/**with an empty bottle**/**with an empty stool** and preheated the oven. She was full of ideas when it came to cooking.

In line with the discussion in 2.3.3, it may be argued that the facilitation of the unfamiliar but appropriate target (*empty bottle*) is due to the activation of a contextually relevant schema, which reinforces the activation of those emergent properties of the target word that make it a suitable candidate to fit its slot (e.g. CYLINDRICAL → ROLLS → CAN BE USED TO STRETCH SOMETHING OUT).

2.3.5 Accessibility and lexical adjustment

In what follows, I briefly discuss a study conducted by Rubio-Fernández (2007) on metaphor interpretation and its insights for assessing alternative approaches to lexical adjustment. Rubio-Fernández (2007) investigates the main empirical predictions of various accounts of metaphor interpretation, which all propose that metaphor

[12] Further evidence can be found, among others, in Breheny, Ferguson, and Katsos, (2012, 2013).

interpretation involves the construction of an ad hoc concept on the basis of the concept encoded by the metaphor vehicle. The predictions at issue (e.g. enhancement of relevant properties of the metaphor vehicle and suppression of irrelevant ones) are shared by these accounts and the differences between them are not directly tested. However, as Rubio-Fernández (2007, p. 348) underlines, "although these differences might not be open to direct empirical investigation, the result of the present study might shed some light on some of these issues". In particular, I would like to discuss the extent to which her results support Recanati's model of lexical adjustment in terms of accessibility.

Rubio-Fernández (2007) conducted a cross-modal lexical priming study in order to investigate the patterns of activation and suppression of relevant and irrelevant properties of the metaphor vehicle during metaphor interpretation. Participants were presented (in the auditory modality) with contexts biased in favour of metaphorical interpretation (e.g. *Nobody wanted to run against John at school. He was a cheetah.*). At the offset of the metaphoric prime (e.g. *cheetah*), participants had to make a lexical decision on a visual target. Critical targets were either metaphor-inconsistent superordinates of the metaphor vehicle (e.g. *cat*) or metaphor-relevant distinctive properties (e.g. *fast*). Targets were presented 0, 400 and 1000 ms from the offset of the metaphoric prime in order to investigate the time course of activations of the properties at issue. The results are shown in the following graph:

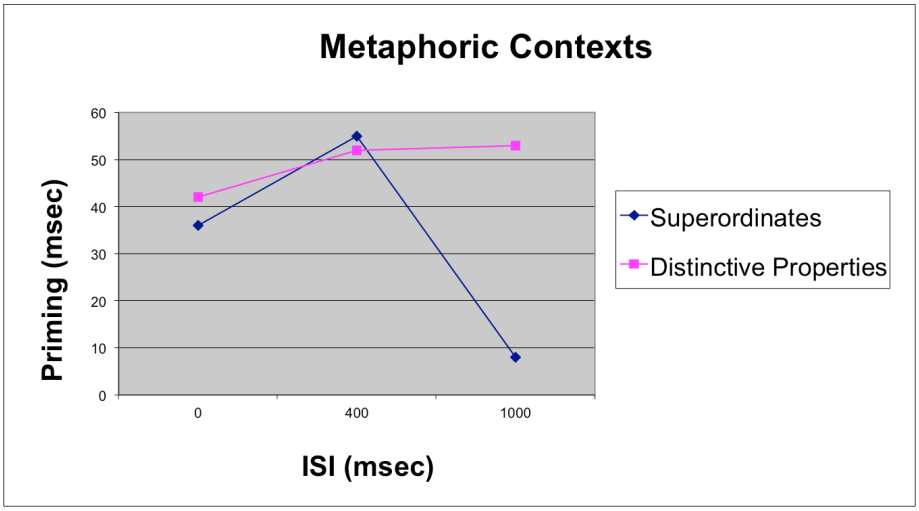

Figure 1 Activation curves of superordinates and distinctive properties in metaphoric contexts.

For the purpose of our discourse, it is interesting to notice that both metaphor-inconsistent superordinates and metaphor-relevant distinctive properties were active up to 400 ms from the offset of the metaphoric prime, with the former showing a slightly higher level of priming at that point in processing. Their patterns of activation differ significantly only at the 1000-ms delay.

According to Recanati (2004)'s accessibility-based framework of lexical adjustment, when processing a word, the literal meaning is accessed first and it triggers the activation of further meanings. These are associatively derived from the literal meaning through which they gain part of their activation and they are processed in parallel. The candidate which is retained and undergoes semantic composition is the *most accessible* one. It corresponds to the most active representation when the interpretation process stabilises. The question Rubio-Fernández addresses concerns the extent to which the activation patterns observed fit with Recanati's framework. According to Rubio-Fernández:

> For the present result to have fully supported Recanati's account of lexical interpretation [..], superordinates should have been active at the intermediate delay but at a lower level than distinctive properties, so that the latter would have been selected for interpretation while the former would have stood in conflict with interpretation. However, if the selection of meaning components was directed by considerations of

> relevance, and not by accessibility alone [..], it would still be possible to argue that superordinates had been discarded from the interpretation although still highly active at 400 ms. (Rubio-Fernández, 2007, p. 363)

In other words, the high activation of metaphor-inconsistent superordinates until late in processing seems to suggest that the selection of meaning components for deployment in the interpretation does not entirely depend on considerations of accessibility. This, in turn, would support the relevance-theoretic picture, which takes considerations of accessibility, on the one hand, and considerations of acceptability, on the other, to play joint but distinct roles in the interpretation process.

While this interpretation of the results could shed some light on the theoretical debate at issue, it is not to be taken as conclusive and must be endorsed cautiously. In particular, the vagueness of the empirical predictions derivable from Recanati's account does not allow us to fully embrace the suggested interpretation. For instance, Recanati's claim that the candidate which undergoes semantic composition is the most accessible one "when the interpretation process stabilizes" is still compatible with the result of Rubio-Fernández's study. The key question, which does not seem to have a definite answer, is the following: *when* does the interpretation process stabilise? Recanati could potentially argue that, since Rubio-Fernández's material comprises cases of *novel* metaphors, the interpretation process stabilises later than 400 ms from the offset of the metaphor vehicle. At that point, presumably falling between the 400-ms delay and the 1000-ms delay, the degree of activation of the figurative meaning of the vehicle (and its associated properties) would be higher than its lexically encoded meaning (and its superordinates), as predicted by his own framework. Thus, this pattern of results need not be interpreted as incompatible with an accessibility-based framework of lexical adjustment.

Recanati's account of primary pragmatic processes is based on the notion of accessibility. The dynamics of accessibility can be susceptible to highly context-dependent constraints. Far from reflecting mere associative priming (i.e. priming by associates which might or might not be semantically related), it is also affected by non-associative semantic priming. The latter is mediated by schemata, i.e. structured

patterns of information stored in our long-term memory. The notion of schemata has recently been extended beyond its original scope (Recanati, 2007b; Mazzone, 2011). As well as general world-knowledge, *particular* world-knowledge (e.g. knowledge about a particular individual, his abilities and preferences) can affect the dynamics of accessibility via the same blind mechanism of associations. The result is that "the smartness of an inferential system can be implemented in a dumb associative system" (Recanati, 2007b, p. 52). I suspect the distinction between such a 'smart' associative system, on the one hand, and an automatic, spontaneous inferential system, on the other, may be too fine-grained to be open to empirical investigation. Indeed, this decreases the chances of figuring out clear-cut empirical predictions, which could be tested experimentally. Thus, on the experimentalists' side, it requires extra careful attention in developing predictions which are genuinely warranted by the theories. However, this leaves wide space for further philosophical/theoretical analysis. A theoretical investigation of the conceptual tools employed by alternative frameworks, such as the notion of schemata or inferential heuristics, could provide a complementary means to evaluate these alternative approaches. As these considerations should suggest, the experimental turn, which has characterised the recent developments in the field of pragmatics, is to be integrated with, rather than to replace, its traditional methodologies. Arguably, the place for philosophical investigation has not only been preserved but is even more needed. Chapter 3 is devoted to such a philosophically-oriented approach to the debate at issue.

Chapter 3 Criticisms and defences

At the end of Chapter 1, I presented two of the foundational issues which are debated in the literature concerning the nature of pragmatic processes: (i) the question of which pieces of information affect the interpretative process, and, specifically whether information about the speaker's mental states needs to be actively taken into consideration in recovering the speaker's intended meaning; (ii) the question of whether pragmatic interpretation is *inferential* in the sense that it requires a stage/component of 'hypothesis confirmation', where interpretative hypotheses about the speaker's intended meaning are assessed against some criterion of pragmatic acceptability before being attributed to the speaker herself. In this chapter, I suggest a positive answer to both questions.

3.1 Carston's counterexample to Recanti's account

In order to demonstrate the inadequacy of the accessibility-based framework proposed by Recanati, Carston (2007, pp. 25-27) elaborates an example which could ultimately show that Recanati's account can even fail to explain such a simple pragmatic phenomenon as reference assignment. Furthermore, it is intended to show that, contrary to what Recanati (2004, p. 32) claims, the interpretative process cannot be blind to considerations about the speaker's beliefs and intentions. Consider the following scenario: Robyn is in one of her students' company. At some point in the conversation, the student, Sarah, utters

(16) Neil has broken his leg.

Suppose that Robyn knows two people called "Neil", her young son (NEIL$_1$) and a colleague in the linguistics department where she works (NEIL$_2$). Thus, there are two possible interpretations of (16) characterised by different reference assignments (leaving aside here any other primary pragmatic processes that might also be required):

(17) a. NEIL$_1$ HAS BROKEN HIS LEG.
b. NEIL$_2$ HAS BROKEN HIS LEG.

Assume that Robyn is aware that Sarah does not know anything about her private life, and that she does not even know that Robyn has a child. This being the case, our off-line intuitions suggest that (17b) is the selected interpretation of Sarah's utterance. If Robyn recognises that Sarah does not *know* (or *believe*) that she has a son called "Neil", she takes it that Sarah did not (and could not) *intend* to refer to him.

Carston (2007) claims that an accessibility-based account of primary processes is inadequate to deal even with such simple cases of reference assignment as the one in (16). Imagine that Robyn is constantly worried about her son, NEIL_1, who tends to get into a lot of trouble. For this reason, when processing the word "Neil", her NEIL_1 concept is more highly activated than her NEIL_2 concept. If accessibility is all that counts for a referent to be selected as the intended one, the interpretation (17a) should be selected as the correct one. There is nothing blatantly wrong with this output; the interpretation is consistent with Robyn's knowledge (she has not seen her son for several hours) and it is perfectly coherent. However, as off-line intuitions suggest, this is not the interpretation that gets selected.

Relevance theory, on the other hand, has the theoretical tools to explain how (16) is correctly interpreted in the context at hand. In fact, it predicts that the interpreter, Robyn, follows a comprehension process which stops when the expectation of optimal relevance raised by the utterance is satisfied. This means that competing interpretations are accessed in order of accessibility and one is selected if and only if it is optimally relevant (or, more precisely, it meets the hearer's occasion-specific expectation of relevance, which is underpinned by the presumption of optimal relevance carried by all utterances). As we have previously noticed, the concept of optimal relevance includes as a crucial component considerations of the speaker's abilities and preferences:

(2) **Presumption of optimal relevance**
   a. The ostensive stimulus is relevant enough to be worth the audience's processing effort.
   b. The ostensive stimulus is the most relevant one compatible with the communicator's abilities and preferences.

The ostensive stimulus is not expected to be the most relevant *tout court*, but the most relevant one compatible with the communicator's abilities and preferences. What do these notions amount to? On the one hand, 'abilities' refers to both linguistic competence and broader epistemic states (e.g. knowledge of or beliefs about the world). On the other hand, 'preferences' comprises linguistic preferences (e.g. preference for formal or indirect modes of expression), social behavioural preferences (e.g. compliance with social conventions) and other desires/goals (e.g. intention to withhold some information from the interlocutor).

The rationale behind clause (b) of the presumption of optimal relevance is that "communicators, of course, are not omniscient, and they cannot be expected to go against their own interests and preferences in producing an utterance" (Wilson & Sperber, 2004, p. 257). The constraints that considerations about the communicator's abilities and preferences impose on the addressee's expectations of relevance enable the addressee to cope with cases of *accidental relevance* like the one presented in (16) (Wilson, 2000, p. 421). The first interpretation that seems relevant enough to the addressee, (17a), is not the intended one. The interpretation (17a), which is accessed first due to its high accessibility, is thus dismissed since it is not optimally relevant: Robyn recognises that Sarah does not know (or believe) that she has a son called "Neil" and she takes it that, therefore, Sarah did not (and could not) intend to refer to him. That is, the interpretation is not compatible with the speaker's beliefs and knowledge about the world (her 'abilities'). As a consequence, further interpretative hypotheses, such as (17b), are accessed and assessed by the addressee and the one that meets the presumption of optimal relevance is accepted.

3.2 Recanati's reply

In his reply to Carston, Recanati (2007b) tries to resist the conclusion that accessibility is not ultimately an adequate notion to provide an account of hearers' grasp of 'what is said' (explicature) which is exclusively based on it. His strategy consists of providing an explanation of the example at issue in terms of an accessibility shift.

Assuming that Robyn's NEIL$_1$ concept is initially more highly activated than her

NEIL$_2$ concept, the only way in which Recanati can explain that the intended interpretation is (17b) NEIL$_2$ HAS BROKEN HIS LEG is to appeal to an accessibility shift from an initial stage *s* to a later stage of the interpretation process, *s′*, in which NEIL$_2$ has increased its degree of activation and, as a result, is the most accessible candidate referent. According to Recanati, an accessibility shift occurs when an initially highly activated conceptual representation fails to fit some activated schemata. Recanati (2007b, p. 50) suggests that it is the meta-representational schema "The speaker$_x$ says that ..." which is responsible for the required accessibility shift. Since the speaker (Sarah) is completely unconnected to Neil$_1$, (17a) fails to fit the schema. In other terms, "Sarah says that Neil$_1$ has broken his$_x$ leg" is not a coherent instance of the metarepresentational schema. Thus, (17a) loses its initial high accessibility while (17b) increases its degree of accessibility. In fact, not only is "Sarah says that Neil$_2$ has broken his$_x$ leg" a perfectly coherent instance of the meta-representational schema, but also the concept of NEIL$_2$ and the concept of SARAH mutually reinforce their activation owing to the connection between them in the knowledge base of the interpreter.

According to this explanation, the initial stage of the interpretation process is followed by a later stage, which Recanati calls 'the externalization of the explicature', when the primary meaning is embedded within the meta-representational schema and the accessibility shift is so triggered. The meta-represented explicature is seen as an essential input to secondary pragmatic processes. This view is substantially in line with the Gricean idea that the speaker's having said that P is the initial premise in the working-out schema of implicature derivation. In fact, it is wholly consistent with Recanati's adoption of the Gricean perspective on implicature derivation. As Recanati (2002a, p. 114) emphasises "conversational implicatures, in the strict sense, are inferentially derived from premises concerning the speaker's intentions in saying what he says." Hence, the externalization stage can be seen as providing an essential premise to the inferential process which delivers conversational implicatures.

3.3 A critical response to Recanati's account

In this section I argue that the notion of meta-representational schema advocated by

Recanati cannot trigger the predicted accessibility shift. Let us start by focusing on the following quotation:

> […] Sarah (the speaker) is *unconnected to* $Neil_1$, while she's got some connection to $Neil_2$. As a result of this, $Neil_2$ becomes the most accessible candidate at stage *s'*. […] This is *very similar* to the other sorts of accessibility-shift I have described in the works Carston refers to. (Recanati, 2007b, p. 50, *my emphasis* (DM))

The first point that needs to be emphasised is the assumed similarity between the example at issue and other cases of accessibility shift. It seems to follow that the meta-representational schema "The speaker has said that..." works according to the same dynamics of activations-and-associations that governs every world-knowledge schema. As I have already mentioned, schemata promote the search for coherent interpretations (where coherence is defined in terms of schema satisfaction): "One interpretation is the most accessible one, hence wins, at ***s***, but the interpretation fails to fit some schema, hence loses at some later stage ***s'***" (Recanati, 2004, p. 32). Applying this line of explanation to the Neil-example, we can say that the interpretation NEIL_1 HAS BROKEN HIS LEG (initially more accessible), fails to fit the meta-representational schema and it is then dismissed in favour of a more coherent interpretation (NEIL_2 HAS BROKEN HIS LEG). The key question that I would like to posit is the following: why does NEIL_1 HAS BROKEN HIS LEG fail to fit the meta-representational schema? Or, in other words, why should it count as an incoherent instance of the meta-representational schema itself?

According to the passage I quoted, it is the fact that the concept of Sarah and the concept of $Neil_1$ are unconnected to each other within the knowledge base of the interpreter (Robyn) which leads the interpretation NEIL_1 HAS BROKEN HIS LEG to fail. However, the lack of pre-existing associations of this kind can hardly be considered a sufficient reason to trigger a garden-path effect. In fact, it seems plausible to assume that most utterances express propositions containing concepts of individuals that lack any pre-existing association with the speaker (in the knowledge base of the interpreter). To clarify this point, let us consider the following example:

(18) a. *Peter:* What's your plan for tonight?

b. *Sarah:* Dom La Nena is giving a concert.

It is easy to imagine a context in which Sarah's utterance communicates that Madonna is giving a concert tonight and that Sarah is planning to go to it. Furthermore, it could easily be the case that Peter's knowledge base lacks any pre-existing link between the concept of Sarah and the concept of Dom La Nena (e.g. he has never heard of this singer and he does not know anything about Sarah's musical tastes). Nevertheless, this would not cause any difficulty in deriving the communicated content of her utterance. While the explicature of the utterance does not fit the metarepresentational schema, it seems unlikely that the lack of pre-existing associations in the knowledge base of the interpreter would disrupt the normal flow of interpretation.

These considerations seem to suggest that the recognition that the wrong referent (NEIL$_1$) has been assigned cannot be explained along the suggested line. We need a system that can detect an inconsistency between the initial reference assignment and the speaker's belief state; the notion of accessibility shift, together with the introduction of meta-representational schemata, does not seem to be adequate for this task. It seems that more is needed than the blind mechanism of associations-and-activations. What is required is an inference from the fact that the speaker has uttered "Neil has broken his leg", in addition to the assumption that she does not *know* (or *believe*) that I have a son called "Neil", that she did not (could not) *intend* to refer to Neil$_1$ and, therefore, that she meant that Neil$_2$ has broken his leg.

3.4 Mazzone's defence of an accessibility-based approach

Mazzone (2011) proposes a different solution to Carston's counterexample, one which is not committed to meta-representational schemata. He develops an important suggestion already embedded in Recanati's explanation though not yet playing a decisive role in it. As reported above, when Recanati describes the dynamics of accessibility, he fleetingly notices that the concept of NEIL$_2$ (Sarah's syntax lecturer) and the concept of SARAH mutually reinforce each other's level of activation. Their connection within the knowledge base of the interpreter is said to be responsible for this extra activation. Mazzone aims at demonstrating that the connection between the

two concepts is strong enough to drive the interpretation process in the expected direction. In other terms, the particular world-knowledge at issue (the fact that Sarah does not know $Neil_1$ while she is acquainted with $Neil_2$) could ground the expected interpretation of (16), independently from any embedding of the primary meaning within the meta-representational schema "The speaker says that ...".

Mazzone (2011, p. 2158) suggests that speaking with Sarah could automatically trigger Robyn's knowledge about her. This knowledge includes information about the environment where Robyn usually meets her, the people associated with that environment, the activities which generally take place there, and so on. According to this suggestion, Robyn's concept of Sarah should be deeply associated with a related environment (the university) including events and entities such as the other people who belong to that environment (lecturers, in the example). For this reason, Robyn's $NEIL_2$ concept receives extra activation which allows it to win the competition with $NEIL_1$.

This alternative account of Carston's counterexample seems to receive some support from Yeh and Barsalou's (2006) theory of 'situated concepts'. Their main thesis is that concepts are tightly associated with particular background situations.[13] In more specific terms, the representation of a concept in memory is associated with situational information about physical settings, events, and the subjective perspectives of agents. This means that when a concept is processed, associated situations become active (and *vice versa*).

We could think about background situations associated with a concept in terms of schemata which are activated by its occurrence. In line with this perspective, we should say that speaking with Sarah (the focal entity) activates a so-called "university schema" which corresponds to the background situation in which Robyn usually meets Sarah. It could be a 'situational' schema of this sort: $<$STUDENT (X_1, ...,

[13] A situation is defined as "a region of perceived space that surrounds a focal entity over some temporal duration, perceived from the subject perspective of an agent. The region of space surrounding the entity may include a variety of entities and events, and the agent's subjective perspective on the region may contain a variety of mental states." (Yeh & Barsalou, 2006, p. 353).

X_N) – LECTURER (Y_1,..., Y_M) – CLASS (Z_1, …, Z_L) – …>. The activation of such a schema raises the accessibility of the concept NEIL$_2$, a relevant semantic value for the slot LECTURER, and triggers the accessibility shift from NEIL$_1$ (initially more active) to NEIL$_2$. Furthermore, we may assume that the highly frequent co-occurrence of SARAH and this kind of background situation in Robyn's experience makes the "university schema" highly activated by Robyn's representation of Sarah.

To sum up, Mazzone (2011) provides an associative explanation of Carston's counterexample which seems to be less problematic than Recanati's one, and, in fact, much better fitted to Recanati's overall project. It does not require the intervention of some occasional supplements to primary pragmatic processes (e.g. via the meta-representational schema 'The speaker says that ...') but is completely based on the standard dynamics of association-and-activation. Despite this, I think there are some residual perplexities with Mazzone's account. Mazzone's line of argument seems to take for granted that 'situational schemata' have the force to trigger the required accessibility shift. Fitting the "university schema", Robyn's NEIL$_2$ concept is said to receive an extra activation which would outdo the level of activation of the individual concept of Robyn's son. However, the initial hypothesis that NEIL$_1$ is so much in her current thoughts casts some doubt about this alleged automatic shift. It could be objected that, although Robyn's NEIL$_2$ concept increases in accessibility thanks to the "university schema" evoked, its degree of activation is still not significantly higher than that of her NEIL$_1$ concept. In fact, Robyn's NEIL$_1$ concept could be even more active due to his mother's excessive apprehension. Therefore, once again, accessibility alone might not be sufficient to warrant the intended interpretation.

3.5 The speaker's preferences: A new counterexample

In the previous sections, we discussed Carston's counterexample and its possible solutions. Putting aside the final doubt raised above, Mazzone's solution seems to be able to contain the critical flow of the discussed counterexample against accessibility-based explanations in pragmatics. However, in this section I propose an adapted version of Carston's counterexample which should be resistant to any similar explanation in terms of associative processes.

Let us consider the following scenario. Pietro and Sara are two Italian MA students in a Linguistics department in London. They first met at university and they usually spend a lot of time together, attending the same classes and studying at the departmental library. Let us suppose that they are talking together about their university life, their impressions of their first term courses and their lecturers. Pietro then asks Sara if they could meet on Wednesday afternoon to prepare their oral presentation on Generative Grammar planned for the following week and she replies, uttering:

(19) No, sorry. I'm meeting Neil on Wednesday.

Let us further suppose that Pietro knows two people called "Neil", one of whom is their syntax lecturer, $Neil_2$, the other is a guy they met a few days before at a party, $Neil_1$. Thus, there are two possible interpretations of (19):

(20) a. I'M MEETING $NEIL_1$ ON WEDNESDAY.
b. I'M MEETING $NEIL_2$ ON WEDNESDAY.

Because of the topic of their conversation (the grammar presentation), Pietro's $NEIL_2$ concept is more highly activated than his $NEIL_1$ concept. For this reason, an accessibility-based framework predicts that (20b) is the selected interpretation, unless some accessibility shift occurs. Once again, there is nothing blatantly wrong within this interpretation: it is perfectly coherent with the speech situation and relevant. Nevertheless, this is not the interpretation Pietro assigns to the utterance. The reason is that he knows it is very unlikely that Sara would refer to their syntax lecturer using his first name rather than a more formal title like 'Dr. Simpson'. Recognising that social conventions do not allow Sara to do this, Pietro takes it that she did not intend to refer to their lecturer.

According to Relevance Theory, we can explain the selected interpretation, (20a), along the following line: (20a) is selected as the winning interpretation since it is the one which is optimally relevant. In fact, even if (20b) is accessed first because of its higher accessibility, it is dismissed by the interpreter, being incompatible with what he knows of the speaker's preferences. As mentioned above, the notion of optimal

relevance includes considerations of both the speaker's abilities and her preferences. In the example at issue, Sara's preferences include compliance with a (Italian) social convention which prescribes avoidance of first names when referring to older people known in a pretty formal context. For this reason, Pietro recognises that, given Sara's concern to be polite and respectful, she did not intend to refer to $Neil_2$ by uttering "Neil".

Before analysing the extent to which an accessibility-based account might try to explain this example, it is important to remove one potential objection to the explanation of the example given above. In fact, it could be replied that the explanation of the example does not work because it assumes that Pietro's $NEIL_2$ concept should be initially more activated. This objection could exploit the fact that, just as the existence of social conventions may prevent students from referring to their lecturers by their first names, so Pietro's concept of $NEIL_2$ would not be highly activated by an occurrence of "Neil". In order to resist the objection, we may refine the example at hand by providing some more details. Let us suppose that Pietro and Sara are very familiar with their syntax lecturer's first name. They have read a lot of papers published by $Neil_2$ which presented his full name, 'Dr Neil Simpson', among the headings, they have often visited his office which has a plaque bearing his full name hanging on the door, they have listened to other lecturers and British students speaking about $Neil_2$ using the name "Neil", they have often encountered his full name on their timetable, and so on. In these circumstances, we may assume that Pietro's (and Sara's) $NEIL_2$ concept is strongly associated with the word "Neil" even if neither of them actually use it to refer to $Neil_2$. These refinements should avoid objections of this sort and support the adequacy of the counterexample proposed.

In what follows, I try to sketch an accessibility-based explanation of the counterexample proposed following Mazzone's (2011) account based on situational schemata. My aim is to show how this kind of explanation fails with the example at hand. In line with the previous explanation, we may suppose that Pietro's representation of Sara is associated with several background situations (at least two): a "university situation" and a "party situation". For this reason, speaking with Sara should activate two different background situations or situational schemata:

the "university schema", <CLASSMATE (X_1, ..., X_N) – LECTURER (Y_1, ..., Y_M) – CLASS (Z_1, …, Z_L) – …>, and

the "party schema", <FRIEND (X_1, ..., X_N) – LOCATION (Y) – DRINK (Z) – …>.

Generally speaking, it is rather plausible to assume that a particular concept can be linked with more than one background situation. For this reason, we need an account of how different and competing situational schemata can drive the interpretation process. Let us consider an insightful passage from Recanati:

> […] a schema is activated by, or accessed through, an expression whose semantic value corresponds to an aspect of the schema. The schema thus activated in turn raises the accessibility of whatever possible semantic values for other constituents of the sentence happen to fit the schema. The schema itself gains extra activation from the fact that some other constituent of the sentence has a possible interpretation which fits the schema. [...] Coherent, schema-instantiating interpretations therefore tend to be selected and preferred over non-integrated or 'loose' interpretations. (Recanati, 2004, p. 37)

According to this remark, we could claim that, although all associated situational schemata are activated by a particular concept, it is only the one whose slots are filled by some other constituents of the uttered sentence that plays a role in the interpretation process.

However, it could be the case that two different associated schemata both gain extra activation from the fact that some other constituent of the uttered sentence has a possible interpretation which fits the two schemata. This seems to be the case of our example: both the "university schema" and the "party schema" can gain extra activation from the fact that the constituent "Neil" of the uttered sentence (19) has two possible interpretations (NEIL_1 and NEIL_2, respectively) each of which fits one of the schemata. In fact, NEIL_2 is a relevant semantic value for the slot LECTURER_I whereas NEIL_1 is a relevant semantic value for the slot FRIEND_J. Consequently, both (20a) and (20b) are interpretations which are coherent instances of an activated situational schema.

Let us further suppose that Pietro and Sara spend a lot of time at the university as

classmates, but in fact they seldom organise to go out together. In these circumstances, the "university schema" should be initially highly activated due to the fact that it corresponds to the background situation in which Pietro usually meets Sara. As long as nothing prevents it from obtaining extra activation during the interpretation process, the resulting interpretation should be (20b) I'M MEETING NEIL_2 ON WEDNESDAY. As noted above, this result strongly contrasts with our intuitions, given the nature of Sarah's social preferences, which are known to Peter. It seems, then, that we have a genuine counterexample to the wholly associative account of primary pragmatic processes.

3.6 Schemata and social conventions

In this section, I discuss a possible objection to my counterexample and propose a third counterexample to deal with it. The objection goes as follows: it is plausible to assume that the social convention at play (i.e. "Do not use a first name to refer to your lecturers") deactivates the concept of NEIL_2 as a candidate for "Neil". Social conventions represent frequently experienced patterns of behaviour (e.g. communicative behaviour). On this basis, we can assume the existence of a schema which encodes this regularity. The activation of such a schema could then play the expected role in the dynamics of activation-and-association.

This objection raises an interesting issue concerning the nature of schemata. In particular, it suggests that the notion of a schema is flexible enough to accommodate every regularity in our experience. However, in what follows, I show that the same argument against an accessibility-based account of reference assignment can be put forward without involving social conventions. In fact, a speaker's preferences which are not reducible to social conventions might play the same role.

Let us consider an adapted scenario from the previous example. Peter and Sarah are two English MA students in Linguistics. They have been asked to prepare an oral presentation on Generative Grammar for their Syntax course. Since it is their first time and they are not really experienced with Power Point, they ask for help from their English friend Mary who studies Statistics in another department and is a confident user of Power Point. The discussion takes place at the campus student bar:

(21) a. *Peter (to Sarah and Mary):* Shall we meet on Wednesday?
b. *Sarah:* No, sorry. I'm meeting Neil on Wednesday.

Let us suppose, once again, that Peter knows two people called "Neil", one of whom, $Neil_1$, is a friend of all three of them, and the other, $Neil_2$, is Peter and Sarah's syntax lecturer. Thus, there are two available interpretations of (21b):

(22) a. I'M MEETING $NEIL_1$ ON WEDNESDAY.
b. I'M MEETING $NEIL_2$ ON WEDNESDAY.

The argument is analogous to the previous case. Because of the topic of the conversation (i.e. generative grammar), Peter's $NEIL_2$ concept is more highly activated than his $NEIL_1$ concept. It follows that an accessibility-based framework would predict that (22b) is the selected interpretation, unless some accessibility shift occurs. Once again, there is nothing blatantly wrong with this interpretation, both in terms of coherence and relevance. Furthermore, it does not violate any social convention (they are English students who constantly use their lecturer's first name, "Neil", to refer to him). Nevertheless, this is not the interpretation that Peter gives to the utterance. The reason is that he knows it is unlikely that, in the context at hand, Sarah would refer to their syntax lecturer using his first name since she knows that Mary does not know him and would assume that Sarah is referring to their common friend $Neil_1$. Recognising that Sarah's interest in being understood by all the participants in the conversation (and in being transparent as much as possible to all of them) would prevent her from doing this, Peter takes it that she did not intend to refer to their lecturer.

Again Relevance Theory makes the right prediction: the most accessible interpretation (22b) is evaluated and dismissed since it is not compatible with Sarah's (the speaker) manifest preferences. It does not satisfy Peter's expectations of optimal relevance, which are met by the interpretation (22a) instead.

3.7 Implicature derivation and 'intentional schemata'

Mazzone (2009, 2011) maintains that Recanati's accessibility-based account of primary pragmatic processes can and should be extended to the derivation of

implicatures. In order to make this case, he introduces a significant extension of the notion of schema. The key assumption of his proposal is that cognitively coded regularities are not limited to word associations and world-knowledge structures. Rather, regularities concerning our intentional communicative behaviours and interpretative practises are coded within so-called 'intentional' schemata. Intentional schemata are thus defined as "specific patterns of intentional behaviour grounded in social regularities" (Mazzone, 2009, p. 321). The only example proposed by Mazzone is the pattern <YES/NO QUESTION (α) – YES/NO ANSWER (β)> (where α and β are variables for propositions). Given the pattern and a yes/no question, a yes-no answer is predicted. In order to clarify Mazzone's account, consider the following example:

(23) a. *Peter:* Would you like something to eat?
b. *Mary:* I've just had breakfast.

In a suitable context, Peter would interpret Mary's utterance as implicitly communicating that she would not like something to eat. Let us try to reconstruct the dynamics of activations-and-associations that could lead to the intended interpretation.[14] Suppose that Peter's question activates the intentional schema <YES/NO QUESTION (α) – YES/NO ANSWER (β)>, which in turn raises expectations about the intended answer (a yes/no answer). Suppose, then, that this intentional schema activates propositions that count as affirmative or negative answers fitting the schema at issue (e.g. MARY WOULD LIKE SOMETHING TO EAT and MARY WOULD NOT LIKE SOMETHING TO EAT). In addition, Mary's utterance, (23b), could activate the schema: <HAVING EATEN IN THE PRECEDING HOURS (*x*) – IS A REASON FOR – NOT WANTING SOMETHING TO EAT (*x*)> (where *x* is a variable for individuals). As a result, the concept NOT WANTING SOMETHING TO EAT would receive activation from both schemata, and the negative answer MARY WOULD NOT LIKE SOMETHING TO EAT would be selected as part of the intended interpretation in virtue of its high accessibility.

[14] Mazzone does not provide any comparable example in his papers. For this reason, this reconstruction, which is entirely mine, may be taken with a pinch of salt (even if Mazzone does consider it as plausible (personal communication)).

As the example suggests, the idea put forth by Mazzone is that different kinds of associations or schemata (e.g. world-knowledge structures, intentional schemata) constrain the dynamics of accessibility in such a way that it eventually leads to the selection of the appropriate speaker meaning. Through pattern recognition, anticipation and addition of activation, the hearer reaches an interpretation whose construction is based on considerations of accessibility alone. While an accessibility-based account of secondary pragmatic processes has not been developed further than this, the analysis should give a flavour of what it might look like. It is worth noticing that while an accessibility-based account of implicature derivation can be built upon an account of primary processes *à la* Recanati, it differs from it in some significant features. First, while primary pragmatic processes are local processes (i.e. they operate on subparts of the linguistic logical form), implicature derivation is a global process (i.e. it operates on fully propositional forms). Second, while some primary processes (e.g. disambiguation and reference assignment) are linguistically mandated or 'bottom up', hence obligatory (required in all contexts), implicature derivation is an optional process, in the sense that it is a 'top-down' process driven purely by context-specific pragmatic considerations. These two features, taken together, make the usual talk of competing candidates less appropriate for implicature derivation.[15] There is no pre-given set of candidate propositions that can be taken as competing for the status of the implicated content of the utterance. In this case, the dynamics of accessibility needs to construct, rather than select, an interpretation that is not given a priori (in contrast, for instance, with alternative lexically-encoded meanings of an ambiguous expression), and that is often ad hoc and one-off.

It is worth noticing that counterexamples, which mimic the ones presented in the previous sections, can be produced against an accessibility-based account of implicature derivation. That implicature derivation is constrained by considerations about the speaker's mental states has long been an uncontroversial assumption of every pragmatic theory.[16] What is at stake, then, is whether or not a 'blind' dynamics

[15] Thanks to Philippe De Brabanter for pointing this out to me. See Carston (2007, p. 27) for similar concerns with regard to "optional and constructive" primary processes such as those involved in metaphor interpretation and other kinds of free enrichment.
[16] But see Jary (2013) for a more nuanced view.

of accessibility can "mimic […] smart inferential processes" (Recanati, 2007b, p. 52) and lead to the derivation of the intended implicit content of the utterance (i.e. its implicatures). As discussed above, Mazzone's (2009) notion of 'intentional' schemata is supposed to capture regular patterns of intentional behaviour in communicative settings. Consider the following scenario. Peter, a young English man, goes to Greece for the wedding of his close Greek friend Nausicaa. The day after the party, Peter is chatting with Irini, Nausicaa's sister, who could not attend the party because she had a terrible flu.

(24) a. *Irini:* Was the party successful?
b. *Peter:* There were almost a hundred people.

Imagine that Irini, acquainted with the crowds that come to traditional Greek weddings (which can run to hundreds and hundreds of guests), does not consider a party with less than a hundred guests a successful party. Imagine, however, that Irini is aware that an English man like Peter, used to a different kind of wedding party, may not know that successful Greek weddings standardly involve hundreds of people (and that he may consider a party with almost a hundred guests a popular and successful one). It seems plausible to assume that while Irini ends up believing that the party was not a success, she does not attribute this implication to Peter. Rather, she attributes to him the opposite implication, that is, that the party was indeed a success.

An accessibility-based account of implicature derivation based on the notion of 'intentional' schemata seems to be inadequate to deal with examples like (24). Imagine that Irini's question, (24a), activates the intentional schema <YES/NO QUESTION (α) – YES/NO ANSWER (β)>, which in turn raises the accessibility of the following propositions: THE PARTY WAS SUCCESSFUL (i.e. yes answer) and THE PARTY WAS NOT SUCCESSFUL (i.e. no answer). Also, assume that Peter's answer, (24b) activates the culture-specific schema <LESS THAN 100 GUESTS AT WEDDING PARTY (*x*) – COUNTS AS – NOT A SUCCESSFUL PARTY (*x*)>, that is particularly salient for a Greek hearer like Irini. The concept NOT SUCCESSFUL PARTY would thus receive extra activation from both schemata, and the negative answer THE PARTY WAS NOT SUCCESSFUL would be the most accessible implication to Irini's mind. Once again, if

accessibility were all that counts in selecting an interpretation, the implication that the party was not successful would be attributed to Peter as an implicature of his utterance. This does not seem to be the case.

3.8 Accessibility *and* acceptability

The counterexamples presented in the previous sections are intended to raise the following issue. Not only are pragmatic processes (be they primary or secondary processes) constrained by considerations about the speaker's mental states (her beliefs and intentions), but also accessibility-based approaches to pragmatics lack the theoretical resources to account for this constraint. While schemata (e.g. world-knowledge structures, intentional schemata) modulate the dynamics of accessibility underpinning the interpretative process, their role does not suffice to explain the output of the interpretative process. It remains unclear how the hearer selects an appropriate interpretation when this is affected by consideration of the speaker's mental states. In what follows I argue that accessibility *and* acceptability are indispensable driving forces of the interpretative process.

Both Recanati (2007b) and Mazzone (2011) concede to Carston (2007) that considerations about the speaker's mental states do affect the interpretative process, but they argue that they do so by modulating the dynamics of accessibility in a 'dumb' and mechanical way:

> Schemata or scripts are an instance of *general* world-knowledge as opposed to *particular* world-knowledge […] I do not believe (and I never claimed) that *only* general world-knowledge can trigger the sort of accessibility shifts I talk about. Particular world-knowledge can play exactly the same role, and the crucial notion of associative 'links' between representations applies to representations at both levels. (Recanati, 2007b, p. 51)

This passage clearly suggests that the dynamics of accessibility can be modulated by particular, as well as general, world-knowledge. This is to say that information about the speaker's mental states, *qua* particular world-knowledge, can drive the dynamics

of accessibility towards the expected interpretation.[17] Information about the speaker's beliefs and intentions, for instance, can trigger an accessibility shift among competing representations.

Consider again example (16). The particular world-knowledge that SARAH DOES NOT KNOW NEIL_1 and that NEIL_2 IS SARAH'S SYNTAX LECTURER can affect the dynamics of accessibility by reversing the outcome of the competition between NEIL_1 and NEIL_2 (by adding extra activation to the latter and/or inhibiting the activation of the former). On this basis, Recanati (2007b) and Mazzone (2011) argue that examples like (16) represent a problem for accessibility-based accounts of pragmatics only under the (wrong) assumptions that information about the speaker's mental states does not alter the level of accessibility of competing representations and that it can only affect the interpretative process via mechanisms that are not based on a dynamics of activations-and-associations. Rather, since information about the speaker's mental states is represented as part of semantic/conceptual networks, it can affect the degree of accessibility of any representations that are part of the same network (François Recanati, personal communication). Thus, the information that SARAH DOES NOT KNOW NEIL_1 and that NEIL_2 IS SARAH'S SYNTAX LECTURER can affect the degree of accessibility of NEIL_1 and NEIL_2 by being part of the same semantic network.

Similar considerations may be adopted to explain the example in (24). In this case, the piece of information that PETER DOES NOT KNOW THAT IN GREECE A PARTY WITH LESS THAN A HUNDRED PEOPLE COUNTS AS AN UNSUCCESSFUL PARTY and/or the information that PETER BELIEVES THAT A PARTY WITH MORE THAN FIFTY GUESTS COUNTS AS A BIG AND SUCCESSFUL PARTY would be activated at some point during the interpretative process.[18] It would consequently trigger an accessibility shift from the interpretation THE PARTY WAS NOT SUCCESSFUL (initially highly activated) to the

[17] As recognised by Recanati himself, this contradicts his previous claim that primary pragmatic processes are blind to considerations about the speaker's mental states (Recanati, 2004, p. 32) and marks an interesting development of his theory.

[18] It is worth noting, though, that it is questionable to assume that absences of knowledge are registered in semantic networks. There are potentially indefinitely many of them for any individual and it seems unlikely that they are established as components of networks rather than being inferentially derived as and when they are relevant.

intended interpretation THE PARTY WAS SUCCESSFUL by inhibiting the culture-specific schema <LESS THAN 100 GUESTS AT WEDDING PARTY (x) – COUNTS AS – NOT A SUCCESSFUL PARTY (x)>.

In what follows, I discuss the viability of this line of explanation. The accessibility-based accounts proposed by Recanati (2004, 2007b) and Mazzone (2011) rely on spreading of activation, reinforcement and inhibition among conceptual representations (i.e. addition or subtraction of activation). The dynamics of different trains of activation interfere with each other by adding or subtracting activation to/from previously activated conceptual representations. In order for this to work, three different conditions need to hold (see Carston, 2007, p. 26): (i) there is a schema (or some particular world-knowledge) that can constrain the interpretative process in the intended way, (ii) such a schema (or particular world-knowledge) can be activated by the utterance situation at issue, (iii) its activation would outdo that of any other competing schemata (or particular world knowledge). My discussion focuses on (iii). This is not to suggest that the other two conditions are uncontroversial. As far as (i) is concerned, for instance, Mazzone himself claims that there is not at present sufficient evidence for 'intentional schemata' (Mazzone, 2009, p. 340). However, it is worth noting that a powerful objection to (iii) would undermine the very possibility of an accessibility-based account of pragmatics even if new evidence confirmed the existence of diverse kinds of schemata encoding various patterns of regularities (e.g. regular intentional behaviours). So, the focus on (iii) seems well motivated.

Accessibility-based accounts of pragmatics assume that, whenever it is supposed to play a role, information about the speaker's mental states has the force to trigger the required accessibility shift. That is, this information outdoes the activation of any other competing schema or piece of activated information. In (16), the activation of the piece of information that SARAH DOES NOT KNOW NEIL_1 triggers the accessibility shift from NEIL_1 to NEIL_2; in (24), the piece of information that PETER DOES NOT KNOW THAT IN GREECE A PARTY WITH LESS THAN A HUNDRED PEOPLE COUNTS AS AN UNSUCCESSFUL PARTY triggers the accessibility shift from THE PARTY WAS NOT SUCCESSFUL to THE PARTY WAS SUCCESSFUL. It may be questioned whether this is the

case: the initially activated representations may be so highly accessible that the activation of competing pieces of information may not be enough to prevent them from entering the selected interpretation (Carston, 2007; Mazzarella, 2011). More importantly, however, there is room to doubt that degree of accessibility of competing representations is all that is involved in reaching a correct interpretation.

Consider again example (24). As noted above, it seems plausible to assume that, in the context at hand, the hearer, Irini, would end up believing that the party was not successful. While she attributes to Peter the intention to communicate that the party was successful, she does not accept this piece of communicated information as true. This opens up interesting considerations about the relationship between comprehension and belief. Sperber et al. (2010) suggest that the believability of a piece of communicated information is epistemically assessed through cost-effective procedures that rely on the degree of activation of representations during the interpretative process:

> […] the comprehension process itself involves the automatic activation of background information in the context of which the utterance may be interpreted as relevant. […] We claim that this same background information which is used in the pursuit of relevance can also yield an imperfect but cost-effective epistemic assessment. (Sperber et al., 2010, p. 374)

They further suggest that when the assessment of a new piece of information against activated background information leads to a contradiction, some coherence checking may be needed. Leaving aside more theory-internal claims, the general idea that activated background information may play a role in the process of assessing (and potentially rejecting) communicated information is not particularly controversial. When considering example (24), it seems plausible to assume that Peter's communicated content is assessed against the highly activated background information that wedding parties in Greece do not count as successful if they have less than a hundred people (i.e. the culture-specific schema <LESS THAN 100 GUESTS AT WEDDING PARTY (x) – COUNTS AS – NOT A SUCCESSFUL PARTY (x)>). Since Irini holds this assumption with much conviction, the resulting contradiction leads to the rejection of the piece of (implicitly) communicated information. Irini does not

believe that the party was successful. Rather, from the explicit content of Peter's utterance (e.g. THERE WERE ALMOST A HUNDRED PEOPLE) and the contextual assumption/cultural specific schema, <LESS THAN 100 GUESTS AT WEDDING PARTY (x) – COUNTS AS – NOT A SUCCESSFUL PARTY (x)>, she infers that the party was not successful. On this basis, she may further infer many relevant conclusions (e.g. that the bride, Nausicaa, may be disappointed) and plan her future actions (e.g. calling Nausicaa to give her some support). Given this, it does not seem to be the case that the highly activated schema <LESS THAN 100 GUESTS AT WEDDING PARTY (x) – COUNTS AS – NOT A SUCCESSFUL PARTY (x)> and the implication THE PARTY WAS NOT SUCCESSFUL are somehow deactivated during the interpretative process. The inhibition mechanisms advocated by Recanati and Mazzone would prevent these representations from playing the role they seem to play in the assessment of the believability of the communicated information and in the process of inferentially deriving new conclusions and action plans.

To sum up the argument, a closer inspection of the role played by activated background assumptions in the processes of comprehending an utterance and of assessing the believability of its interpretation reveals that items of information that are inconsistent with the selected interpretation may be actively exploited in assessing its epistemic status. This casts doubt on the claim that their degree of accessibility is outdone by that of competing representations. But if this is not the case and they remain active during the interpretative process, accessibility cannot be the only constraint on the selection of the intended interpretation.

The above considerations suggest that information about the speaker's mental states may not affect the interpretative process by inhibiting representations that are incompatible with them (as suggested by Recanati and Mazzone). Rather, it may contribute to the interpretative process by triggering the discarding of interpretative hypotheses that are incompatible with them. In example (24), the information that Peter is not acquainted with traditional Greek wedding parties and/or that, according to his British standards, he considers a party with nearly a hundred people really successful may not inhibit the background schematic assumption that <LESS THAN 100 GUESTS AT WEDDING PARTY (x) – COUNTS AS – NOT A SUCCESSFUL PARTY (x)>.

However, this information about Peter's beliefs may contribute to the assessment of the interpretative hypothesis THE PARTY WAS NOT SUCCESSFUL. Given that this interpretation is incompatible with the speaker's beliefs, it would be dismissed in favour of an interpretation that is consistent with those beliefs. This amounts to saying that information about the speaker's mental states seems to contribute to the assessment of the acceptability of an interpretative hypothesis, rather than (or rather than only) to its accessibility. Thus, an analysis of the role of the speaker's mental states in pragmatic interpretation sheds some light on the limits of accessibility-based accounts of pragmatics and calls for an inferential explanation of utterance interpretation.

3.9 Acceptability through the backdoor?

In a recent paper, Mazzone (2013) proposes an amended version of his accessibility-based account of pragmatics. The aim is to cope with the objection, presented by Carston (2007) and Mazzarella (2011), that accessibility-based frameworks do not have the theoretical resources to account for the role of considerations about the speaker's mental states in pragmatic processing. As mentioned above, Carston (2007) and Mazzarella (2011) raise some doubts about the claim that the notion of 'accessibility shift' captures the interpretative contribution of information about the speaker's mental states. What guarantees that this information has the force to trigger the required accessibility shift?

Mazzone (2013) accepts this objection and explores further limitations of his account.[19] In particular, he suggests that, because of its nature, spreading of activation may not adequately capture the role played by information about the speaker's mental states. On the one hand, spreading of activation is a rapidly

[19] Note that a defendant of an accessibility-based approach to pragmatics may find a way out of this objection. In particular, he may argue that information about the speaker's mental states sometimes really does fail to trigger the required accessibility shift (i.e. it fails to affect pragmatic interpretation). This is in fact what we may expect in those circumstances in which the hearer reveals an 'egocentric' interpretative behaviour and fails to take the perspective of the speaker (see Brown-Schmidt and Hanna (2011) for an overview of the psycholinguistic literature on perspective-taking). This is the reason why in the previous section I presented a different objection to accessibility-based approaches to pragmatics that is not open to this line of reply. It aimed at showing that an accessibility shift might not even take place in reaching a correct interpretation.

decaying process and it does not guarantee that relevant information remains active until it can deliver the desired effect. On the other hand, information about the speaker's mental states needs to be monitored and actively maintained in order for it to play a systematic role in utterance interpretation.

To meet these concerns, Mazzone (2013) suggests that his account is to be implemented by integrating associative processes (e.g. spreading of activation, reinforcement and inhibition among activated representations) with attention and working memory:

> […] what I suggest is that attentional processes must be added to simple associative accounts in order to explain how speaker-related information comes to gain prominence in pragmatic processing and, especially, in the evaluation of consistency that either leads to mutual adjustment of premises and conclusions or causes the process to stop. (Mazzone, 2013, p. 113)

Let me briefly analyse this implementation before discussing its consequences for the debate between inferential and accessibility-based approaches to pragmatics. It is widely acknowledged in the pragmatics and social cognition literature that communication works by manipulating the hearer's attention (Sperber & Wilson, 1995; Tomasello, 1999, i.a.). However, Mazzone argues, little emphasis has been given to attentional processes in the explanation of pragmatic processing. In fact, attention to the speaker may account for a great deal of what is involved in utterance interpretation. Thus, the starting premise of Mazzone's proposal is that communication always pre-empts the hearer's attention to the speaker. Since attention involves the representation of inputs in conscious working memory, attention to the speaker will involve the representation of the speaker in conscious working memory. This representation, in turn, activates speaker-related information, which includes information about her mental states (e.g. her beliefs and intentions). It is worth noting that the account allows both bottom-up and top-down activations. While the object of attention is partly determined by the bottom-up activation and competition of conceptual representations (e.g. world-knowledge structures), the

representation of the speaker activates (in a top-down fashion) speaker-related information.[20] As a consequence, bottom-up and top-down activated items of information enter working memory where they are "attentively confronted with each other" (Mazzone, 2013, p. 112).

The central claim is that the attentional processes triggered by every ostensive stimulus (e.g. utterances) guarantee the active maintenance and conscious monitoring of information about the speaker's mental states. As a consequence, this information can reliably affect the recovery of the speaker's intended meaning. Such a recovery is carried out by the interaction between associative and attentional processes. Crucially, this account involves the addition of an extra component to the interpretative process, that is, conscious working memory and its "final evaluation of consistency":

> Processes of this sort could account for the intuition that pragmatic processing involves a *final evaluation of consistency of the attempted interpretation*, in which information about the speaker may play a crucial role *over and beyond simple associations* in the addressee's memory. (Mazzone, 2013, p. 112, *my emphasis* [DM])

In the remaining part of this section, the following question is addressed: where does this account fit in the debate between inferential and accessibility-based accounts of pragmatics? I suggest that, while it is proposed as a refinement of Mazzone's (2009, 2011) accessibility-based account of pragmatic processes, its implementation introduces a significant qualitative shift. In fact, it introduces (through the backdoor) the idea of a distinct stage/component in which the acceptability of interpretative hypotheses is tested.

Recall that the distinction between inferential and accessibility-based approaches to pragmatics is characterised by the kind(s) of stages involved in utterance interpretation (see Chapter 1). Specifically, inferential approaches to pragmatics involve two distinct stages, one of 'hypothesis formation' and one of 'hypothesis

[20] Note that this argument relies on an implicit assumption about the 'situated' nature of conceptual representation (Yeh & Barsalou, 2006). The discussion of this assumption goes beyond the purpose of the present thesis.

confirmation', the latter being governed by some sort of pragmatic criterion of acceptability. Considerations about the acceptability of an interpretative hypothesis determine the stopping point of the interpretative process (i.e. the interpretative process stops when an interpretative hypothesis is found acceptable). Accessibility-based accounts, on the other hand, do not include any confirmatory step in their description of how a certain interpretation is arrived at.

As noted above, the role attributed by Mazzone (2013) to attentional processes (i.e. conscious working memory) is to contribute to the interpretative process by providing a final "evaluation of consistency that either leads to mutual adjustment of premises and conclusions or causes the process to stop" (Mazzone, 2013, p. 113). It is clear that interpretative hypotheses need to be evaluated or tested before being attributed to the speaker and that such an evaluation, when positive, represents the stopping point of the comprehension procedure. The crucial components of any *inferential* model of pragmatic interpretation are thus present in Mazzone's (2013) framework.[21]

To sum up, I have presented some arguments against accessibility-based accounts of pragmatics and suggested the indispensability of a criterion of pragmatic acceptability based on considerations about the speaker's mental states (i.e. her beliefs and intentions). In support of this claim, I have shown that a closer analysis of Mazzone's (2013) account of pragmatic processing reveals that a 'rational justification' of the interpretation derived requires that the interpretative hypothesis is evaluated (and confirmed) in light of the speaker's mental states before it can be attributed to the speaker.

This conclusion opens up interesting questions, which will be addressed in the second part of this thesis. Once it is established that pragmatic interpretation requires interpretative hypotheses to be checked against (what the hearer takes to be) the speaker's mental states, the issue of the cognitive implementation of this checking

[21] Whether or not attentional processes are associative in nature is an issue that I do not discuss here. Such a discussion would pertain to the third level of analysis described in Chapter 1, that is, the level of analysis of the *mechanisms* involved in pragmatic processing.

process arises. Specifically, it is an open question whether the distinct stages of construction and confirmation of interpretative hypothesis are underpinned by a single pragmatic system or by distinct, albeit interacting, cognitive mechanisms.

Part II

Chapter 4 Epistemic vigilance

4.1 Communication and speaker's intentions

"Human communication is characterised, among other things, by the fact that communicators have two distinct goals: to be understood, and to make their audience think or act according to what is to be understood" (Sperber et al., 2010, p. 364). The distinction between understanding and the further cognitive effects that communicators intend to produce in their audience is captured by Grice (1989) with his definition of speaker meaning. As discussed in Chapter 1, according to Grice, to mean something by uttering X, the speaker must display the following set of nested intentions. S must intend

(a) S's utterance of X to produce a certain response *r* in a certain audience A;

(b) A to recognise S's intention (a);

(c) A's recognition of S's intention (a) to function as at least part of A's reason for A's response *r*.

Intention (a) is the basic-level or embedded intention to produce a certain response in the audience. The nature of this response varies according to the force of the utterance. For instance, imperative-type utterances are intended to produce actions, whereas assertions are intended to produce beliefs. Crucially, the recognition of intention (a) is enough for understanding to occur.

Building on Grice's definition of speaker meaning, Sperber and Wilson (1986/1995, p. 29) suggested that ostensive-inferential communication is characterised by the fact that communicators produce an utterance (or another ostensive stimulus) with the following two intentions: (i) the *informative intention* to inform the audience of something, and (ii) the *communicative intention* to inform the audience of one's

informative intention.[22] More technically, the *informative intention* is defined as the intention to make manifest or more manifest to the audience an array of propositions I, whereas the *communicative intention* is defined as the intention to make it mutually manifest to audience and communicator that the communicator has this informative intention. In order to clarify the import of these definitions, we shall focus on the notion of 'manifestness'. A proposition is manifest to an individual at a given time to the extent that he is likely to some positive degree to entertain it and accept it as true. The set of propositions that are manifest to an individual at a given time is called by Sperber and Wilson (1986/1995) the individual's *cognitive environment*. A cognitive environment is shared between two or more individuals when it is a cognitive environment of each of them. A shared cognitive environment might include propositions which identify the people who share that environment. When this is the case, the shared cognitive environment is also a mutual cognitive environment, and it comprises propositions that are *mutually manifest* (Sperber & Wilson, 1986/1995, pp. 41-42).

Communicators aim at changing the manifestness of each other's assumptions. When a speaker says "It is raining", she aims at making the assumption (proposition) that it is raining (typically, here and now) manifest or more manifest to her audience (as well as making manifest further assumption concerning, for instance, how they should spend a rainy afternoon, etc.). In other terms, communicators aim at changing the audience's cognitive environment. Crucially, communicators do this in an ostensive, i.e. overt, way. That is, they make it evident that they want the audience to recognise that they are expressing this intention. This raises an interesting question, which is addressed by Sperber and Wilson (1986/1995, pp. 61-62): why should someone who has an informative intention want to engage in *ostensive* communication? Two main reasons seem to underpin ostensive communication: first, ostension directly affects the audience's attention, which is directed to that stimulus because of the presumption of optimal relevance that it carries; second,

[22] Sperber and Wilson reject the idea that the communicator must have a third-level intention that the addressee's recognition of her informative intention should be at least part of his reason for fulfilling this. This allows Sperber and Wilson's account of ostensive-inferential communication to cover the whole continuum from pure cases of 'showing' to pure cases of 'meaning'. See also Sperber and Wilson (forthcoming/2015).

ostension has a crucial social importance. As Sperber and Wilson (1986/1995, p. 62) suggest, "[a] change in the mutual cognitive environment of two people is a change in their possibilities of interaction (and, in particular, in their possibilities of future communication)".

The social importance of ostensive communication seems to be confirmed by recent research in developmental psychology. Grosse, Behne, Carpenter, and Tomasello (2010) investigated whether 18-, 24- and 30-month-old children reacted differently in a request interaction depending on whether they had been understood correctly, independently of whether they obtained the requested object from the experimenter. In the critical condition (see below), the adult misunderstood the request but handed over the requested object. For instance, when asked for a ball, the adult turned to the distractor object saying "Oh, you want the [paper]" while distractedly placing the requested object in the target position and subsequently uttering "Yes, I'll give it to you in a moment".

Happy Accident Condition

Figure 2 Happy accident condition

Interestingly, children across all groups displayed a significant tendency in repairing the misunderstanding even if they had achieved their material goal of obtaining the requested object. This result was taken to show that children do not regard a request as successful solely on the basis of its material outcome. Rather, understanding seems to be an essential component of children's communicative aims. With regard

to the distinction between communicative and informative intentions discussed above, we could say that successful communication relies not only on the fulfillment of the speaker's informative intention (the elicitation of the appropriate audience's response, e.g. handing over the requested object), but also on the fulfillment of the speaker's communicative intention (which requires the audience to recognise, for instance, that the speaker is requesting a [ball] rather than a [paper]). As emphasised by the authors, "the fact that children do repair in cases in which they did obtain the requested object but have been misunderstood clearly indicates that they already have a separate intention to be understood" (Grosse et al., 2010, p. 1721).

4.2 Comprehension, acceptance and manifestness

Understanding requires the fulfilment of the communicative intention, that is, the recognition of the speaker's basic level/informative intention, but it might not require the fulfilment of the latter. This is the case with assertions. Understanding is achieved when it becomes mutually manifest to the communicator and the audience that the speaker has asserted that P. However, this does not require the assumption P to increase its manifestness, that is, to become more likely to be entertained and *accepted as true*. If we limit ourselves to assertions, we could say that the informative intention is about getting the audience to believe, and the communicative intention is about getting them to understand. Understanding is a matter of recognising what the speaker intends you to believe.

Critically, an audience can understand an utterance without believing what they have understood. In this case, the communicative intention is fulfilled without the corresponding informative intention being fulfilled. As Wilson and Sperber (2004, p. 611) suggest, typically "[w]hether the informative intention itself is fulfilled depends on how much the audience trusts the communicator". In section 4.4, I will focus on the mechanisms which bridge the gap between comprehension and acceptance (believing the proposition(s) communicated). For the time being, I intend to show that the possibility of comprehension without acceptance is warranted by the definition of communicative and informative intentions provided by Sperber and Wilson.

To begin with, it is worth noting that the definition of manifestness provided by Sperber and Wilson is epistemic in nature. As previously discussed, a proposition is manifest to an individual at a given time to the extent that he is likely to some positive degree to entertain it and *accept it as true*. Given that communicative and informative intentions are conceived of as intentions to change the degree of manifestness of certain assumptions, the following question seems to arise. Is the rejection of the communicated content compatible with the idea that manifestness involves acceptance on the interpreter's part?[23] In order to answer this question, it is worth stressing that in the circumstances in which the addressee understands a piece of communicated information without ending up believing it, the communicative intention is fulfilled without the corresponding informative intention being fulfilled. The informative intention is *recognised* (rather than fulfilled). The fulfilment of the communicative intention entails that the fact that the communicator has a certain informative intention, let's call it "i_i", is made mutually manifest. That is, the fact that the communicator has the intention i_i is likely to be accepted as true by the audience (and the communicator) and this is itself manifest. This does not seem to be problematic: while the audience is likely to accept as true the fact that the speaker has the intention i_i (i.e. the intention to make manifest or more manifest to the audience an array of propositions I_i), the audience is not bound to accept as true any of the propositions which are included in I_i. The audience accepting as true the array I_i would correspond to the fulfilment of the informative intention (but it is not a condition for its recognition).

Importantly, the output of the pragmatic system is metarepresentational in nature. According to Relevance Theory, utterance interpretation is a process that starts with a metarepresentation of an attributed utterance ('The speaker uttered *u*') and ends with a metarepresentation of an attributed thought or set of thoughts ('The speaker communicated I'). Understanding requires entertaining and accepting as true this metarepresentational output.

---

[23] Thanks to Steve Oswald for raising this question at the *PragLab Research Colloquium in Linguistics* in Fribourg (May 2015).

4.3 Comprehension and acceptance: two distinct processes?

The relationship between 'understanding' (comprehension) and 'believing' (doxastic acceptance) has been vigorously discussed for decades in philosophy, and more recently in psychology. This debate has focused on the following two competing models. On the one hand, the 'Cartesian model' suggests that understanding and believing are two distinct and sequential psychological operations: that is, belief acquisition is the result of an epistemic assessment that follows comprehension of the asserted content. On the other hand, the 'Spinozan model' claims that belief is acquired automatically: that is, people initially accept every assertion that they understand. Epistemic assessment is an optional process which occurs, if at all, only after information has been added to the 'belief box' and it can result in 'unbelieving' those asserted contents which are found to be not epistemically warranted. In what follows, I discuss some reasons to question the adequacy of the two most prominent models of testimony-based belief acquisition in philosophy and psychology and propose, in line with Sperber et al. (2010), that comprehension and acceptance may be *distinct* albeit *parallel* processes.

4.3.1 The 'Spinozan' model

The philosopher Ruth Millikan has provided one of the most prominent defences of the 'Spinozan' view (see Millikan, 2004, 2005). According to Millikan, comprehension typically involves directly believing what has been asserted or directly wanting to comply with what has been requested: "Speech is a form of direct perception of whatever speech is *about*. Interpreting a speech does not require making any inference or having any beliefs […] about the speaker's intentions" (Millikan, 1984, p. 62).

This view has received some empirical support from Gilbert and colleagues' studies on belief formation via testimony. As summarised in Gilbert (1993), Gilbert and colleagues conducted a series of studies to investigate the hypothesis that comprehension includes initial acceptance/belief in the information comprehended. In Gilbert's experimental paradigm, participants are presented with a series of sentences during an initial learning phase. Sentences are accompanied by an

indication of their truth-value (e.g. by explicitly mentioning the truth-value before/after the presentation of each sentence, or by coding their truth value via sentence colour). In a second phase, participants are exposed to the sentences previously learnt and asked to judge their truth-value. Crucially, during the learning phase, a group of participants is exposed to the same stimuli while being involved in a distracting task (e.g. discriminating between different tones). That is, information is acquired under cognitive load. The results showed that participants of this second group systematically reported false information as true, but not vice versa (Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993). This was taken to show that rejection, but not acceptance, of communicated information mobilises additional cognitive resources, which go beyond the ones routinely exploited in the comprehension process. The idea is that participants would automatically accept any incoming information and subsequently assess and 'unbelieve' false information. When they are prevented from investing extra cognitive resources in this subsequent assessment process, default acceptance is not questioned and information is stored in memory as true information.

As suggested by Sperber et al. (2010), Gilbert's studies exhibit some important limitations. From the point of view of the participants, the information presented in the learning phase is mainly irrelevant. For instance, in a representative experiment, participants were told that they had to learn some Hopi words and presented with sentences like "A Monishna is a star" (which were tagged as 'true' or 'false'). Even if participants could muster some interest towards Hopi words (and there is nothing in either their background knowledge nor in the experimental situation that makes this likely to be the case), knowing that "A Monisha is a star" is 'false' would still be blatantly irrelevant to them. Manipulating the relevance of the stimuli, Hasson, Simmons and Todorov (2005) showed that when people read relevantly informative statements which are presented as false, the acceptance rate for such statements is not affected by cognitive load or time pressure. For instance, participants presented with statements like *John owns a television* tagged as 'false' tended to judge as true a statement about John that could be stereotypically inferred from *John does not own a television.* That is, statements that are informative when false are not represented as

if they were true. This means that the relevance of the incoming information prompts its epistemic assessment.

It is debatable whether this represents conclusive evidence against the 'Spinozan' model. For instance, Kissine (2009, 2013) has argued that Gilbert's results still demonstrate that in certain circumstances, incoming information is automatically believed: "That this happens when the communicated message is below a certain threshold of relevance makes no difference to the fact that epistemic vigilance is optional in acquiring hearsay beliefs" (Kissine, 2013, p. 87). According to Kissine, epistemic assessment is an optional process which may be disrupted under cognitive load. When this is the case, automatic acceptance cannot be corrected and incoming information is automatically stored as true.

The question which arises though, and that 'Spinozan' models do not seem to provide an adequate answer to, is what motivates belief revision after initial default acceptance? If people are trustful by 'default' and end up rejecting upcoming information only when circumstances call for it, it is an open question how they recognise such circumstances to begin with. Blind trust, which results in default acceptance, would prevent them from exercising that critical stance towards communicated information which allows them to filter out false information (Sperber et al., 2010, p. 363).

4.3.2 The 'Cartesian' model

An alternative model of belief formation via testimony is what has been referred to within the literature as the 'Cartesian' model (see, e.g., Gilbert (1991)). According to this model, comprehension does not involve automatic acceptance. Rather, epistemic assessment is needed in order to establish whether any piece of communicated information should eventually be accepted or rejected. That is, acceptance is not granted any priority over rejection: both propositional attitudes are the result of a stage of epistemic assessment, which evaluates the plausibility and the reliability of incoming information. This is in line with Descartes's (1664/1983) suggestion that individuals initially withhold their judgement concerning the validity of a proposition during comprehension and that only through subsequent thoughtful analysis they do,

or do not, grant its validity. Importantly, epistemic assessment is seen as an optional process, that is, there is not inevitable assessment: representations can be stored without being tagged as true or false. Gilbert (1991) develops the following metaphor to clarify the relationship between comprehension and epistemic assessment suggested by the Cartesian model:

> A new book (new information) appeared in the library (was represented in the mind), its contents were read (assessed), and the book was then tagged (recorded or represented) as either fiction (false) or nonfiction (true). New books (unassessed ideas) lacked a tag, of course, and thus were not identifiable as either fiction or nonfiction until they had been read. Such new and unread books were "merely" represented in the library. (Gilbert, 1991, pp. 108-109)

It follows that if the process of belief formation is delayed or stalled, perhaps due to lack of sufficient cognitive resources to carry out the assessment process, the Cartesian model predicts that the individual will be left in a state of *non*belief (as opposed to belief or *dis*belief).

4.3.3 Assessment of the Cartesian and Spinozan models

The Cartesian and Spinozan models are both dual-stage models: they assume that people *first* comprehend a piece of communicated information (whether or not understanding involves default belief), and *then* assess its believability (whether or not information is routinely evaluated). Epistemic assessment is thus seen as a subsequent and resource-demanding stage that follows effortless comprehension.

	Spinozan model	**Cartesian model**
REPRESENTATION STAGE	Comprehension and acceptance	Comprehension
ASSESSMENT STAGE	Certification or unacceptance	Acceptance or rejection

Recent research clearly suggests that some form of spontaneous epistemic assessment occurs in parallel with comprehension. For instance, Richter, Schroeder, and Wöhrmann (2009) conducted a series of studies to test the hypothesis that relevant background knowledge is automatically and routinely used in order to assess

incoming information. In one study, they adopted Gilbert's paradigm but manipulated the nature of the stimuli: they used realistic factual statements that were either *strongly* related to participants' knowledge (e.g. *Soft soap is edible*) or *weakly* related to it (e.g. *Toothpaste contains sulfur*). Interestingly, results showed that while the verification of statements that were only weakly related to participants' knowledge followed the pattern identified by Gilbert (with false statements being systematically reported as true under cognitive load, but not vice versa), the additional cognitive load did not have any impairing effects on the verification of both true and false statements when these were strongly related to the participant's background knowledge. Richter et al. (2009) further investigated the role of automatic epistemic assessment by adopting a Strooplike paradigm. Participants were asked to perform orthographical judgements on individual words which appeared one by one on a screen, that is, they were asked to judge whether the word was spelled correctly. Sequences of words formed simple assertions that were either true or false. The critical trials were those in which the target word was placed at the end of a sequence that expressed an assertion strongly related to the participants' background beliefs. Crucially, the results showed a significant interference effect: that is, participants took longer to verify the orthography of the target word when this was part of a false assertion that strongly contradicted the participants' background knowledge (and the other way around: negative answers were slower when the word was embedded in a true assertion). The automatic tendency to epistemically assess incoming information slowed down responses in the unrelated orthographical task. According to Richter et al. (2009), this provides direct evidence that epistemic assessment is an automatic process which is carried out non-strategically (that is, it does not require a specific processing goal, i.e. validation, to be triggered).

Further evidence of spontaneous epistemic assessment can be found in the ERP studies conducted by Hagoort, Hald, Bastiaansen and Petersson (2004). ERPs indicate that violations of world knowledge are detected as quickly as semantic anomalies. Participants' brain activity was recorded while reading three different versions of a sentence like *The Dutch trains are yellow/white/sour*. On the assumption that Dutch participants have strong background knowledge concerning the actual colour of Dutch trains, the sentence *The Dutch trains are yellow* was

semantically coherent and consistent with the participants' world knowledge, whereas the other two versions of the sentence contradicted either the participants' world knowledge/background beliefs (*The Dutch trains are white*) or their semantic knowledge (*The Dutch trains are sour*). Crucially, results showed an N400 effect for both world knowledge and semantic violations, which was identical in onset and peak latency and very similar with regard to amplitude and topographic distribution. As emphasised by Hagoort et al. (2004, p. 439), "[t]his finding is strong empirical evidence that lexical semantic knowledge and general world knowledge are both integrated in the same time frame during sentence interpretation, starting at 300 ms after word onset.".

The empirical evidence discussed in this section contradicts the assumption inherent in the 'Cartesian' model of belief-formation that epistemic assessment is a resource-demanding process which optionally follows comprehension. More generally, it casts some doubt on the idea, shared by both 'Cartesian' and 'Spinozan' models, that comprehension and epistemic assessment represent two distinct and sequential stages in the overall process of acquiring belief via testimony.

4.3.4 The 'pragmatic' model

Sperber et al. (2010) have recently proposed a model of the relation between comprehension and epistemic assessment that, for reasons which will soon become apparent, we shall call the 'pragmatic' model. According to Sperber and colleagues, comprehension and epistemic assessment are *parallel* processes which are triggered by the very same act of ostensive communication. While comprehension is underpinned by a relevance-guided comprehension procedure (see section 1.2), epistemic assessment is carried out by a dedicated mechanism which contributes to the capacity for 'epistemic vigilance'. Before discussing the nature of epistemic vigilance mechanisms (see 4.4 below), it is worth focusing on Sperber et al.'s (2010) rejection of a two-stage view of the relationship between comprehension and epistemic assessment.

First of all, Sperber et al. (2010) suggest that considerations of believability play a crucial role in the comprehension process itself:

> "We claim that, whether he ends up accepting it or not, the hearer interprets the speakers as asserting a proposition that would be relevant enough to him provided that he accepted it" (Sperber et al., 2010, p. 386)

This is a 'hypothetical' role: comprehension initially proceeds *as if* the interpretative hypotheses under construction were to be accepted as true. Given such an assumption, the first hypothesis that satisfies the addressee's expectations of relevance is attributed to the speaker as her intended meaning. No *actual* assessment of the believability of that hypothesis needs to take place at this stage. As discuss in Chapter 1, Relevance Theory claims that comprehension is driven by the expectations of relevance which are raised by every ostensive stimulus. Specifically, every ostensive stimulus conveys a presumption of its own optimal relevance, that is, the expectation that the stimulus will be relevant enough to the addressee (to be worth processing) and that it is the most relevant one compatible with the speaker's abilities and preferences. This presumption need not to be true or accepted as true: the speaker might fail to achieve relevance either because of incompetence or malevolence (Sperber & Wilson, 1986/1995, pp. 158-159). However, the very fact that the presumption is communicated is enough to guide the interpretative process. To illustrate this point, Sperber and colleagues discuss the following example. Andy and Barbara have decided to throw a party and Barbara has asked Joan to bring a bottle of champagne.

(25) a. *Andy (to Barbara):* A bottle of champagne? But champagne is expensive!
b. *Barbara:* Joan has money.

Imagine that Andy had previously assumed that Joan was a junior underpaid academic. In the context at issue, Barbara's utterance would make a relevant contribution to the discussion if Andy interpreted it as communicating that *Joan has enough money to be easily able to afford champagne*. The interpretation that *Joan has some money* (which is not only true but also compatible with Andy's background belief) would make little sense as a conversational move at this point of the conversation. Considerations of relevance lead Andy to interpret Barbara's utterance in the expected way: the interpretation that *Joan has enough money to be easily able to afford champagne* is relevant to Andy provided that he accepts it as true. Andy may decide to reject the communicated information (for instance, because he thinks

that Barbara does not know that Joan is only a teaching assistant who is paid on an hourly basis) but, whether or not he ends up believing it, he will interpret it in order to optimise its (intended) relevance.

In line with this, Origgi (2008) suggests that interpretation involves a 'stance of trust' that our interlocutors will provide relevant information for us. Any departure from the satisfaction of our expectations of relevance may result in a revision or a withdrawal of the initial trust with which we approach the interpretative process: this is why the stance of trust is 'dynamic' – it is only tentative and labile, but it plays a crucial role in determining the output of the comprehension system.

Interestingly, it is by appealing to this stance of trust that Sperber and colleagues have attempted to explain the results of Gilbert's psychological experiments. The stance of trust involved in comprehension may be responsible for a bias in favour of actual acceptance of communicated information. That is, when there are no high stakes and the cognitive resources devoted to epistemic assessment are kept to a bare minimum, acceptance of the communicated information may be favoured. As Sperber et al. (2010) recognise, a bias towards acceptance would be entirely rational since most communication is honest (presumably because of the very existence of epistemic vigilance mechanisms).

The 'pragmatic' model is thus grounded on the claim that the search for relevance, which drives comprehension, requires considerations about the believability of the interpretation under construction to enter the picture. Furthermore, as Sperber et al. (2010) propose, it is this very search for relevance which triggers spontaneous forms of epistemic assessment. As discussed in section 1.2, the search for relevance corresponds to the search for an optimal balance between cognitive effects and processing effort. Importantly, there are three types of cognitive effects: contextual implications (i.e. implications that can be derived from the input and the context, but from neither input nor context alone), strengthening of available assumptions and contradiction and elimination of available assumptions. It follows that the search for relevance involves the automatic activation of background information (as well as the on-line construction of contextual assumptions) with regard to which the utterance may be interpreted as relevant. This limited subset of activated background

information represents the background against which incoming information might be assessed via "an imperfect but cost-effective epistemic assessment" (Sperber et al., 2010, p. 374).

This proposal is in line with Richter et al.'s (2009) and Hagoort et al.'s (2004) studies demonstrating the central role of background beliefs in automatic epistemic assessment. Furthermore, it offers a convincing answer to the question of which background beliefs are involved in this assessment. Obviously, it would not be psychologically plausible to assume that it is the whole mental encyclopaedia of the individual that constitutes the background against which the plausibility of incoming information is checked. The current suggestion not only has the advantage of identifying how the relevant subset of background beliefs involved in epistemic assessment is selected, but it also avoids postulating any ad hoc activation. Epistemic assessment relies on those background beliefs which are activated in the pursuit of relevance and does not require any extra cost (which would be incurred in mobilising additional information that play no role in the comprehension process itself).

4.4 Epistemic vigilance mechanisms

Epistemic vigilance is an ability underpinned by "a suite of cognitive mechanisms", which is targeted at the risk of misinformation in communication. Each of the mechanisms is likely to be specialised in one of the many kinds of considerations relevant to warranting (or undermining) epistemic trust.

But what exactly is 'epistemic trust'? It can be defined as the willingness to believe the communicator and accept her claims as true. Communicators are not always competent or benevolent and communication is thus open to the risk of misinformation. A competent communicator possesses genuine information (rather than misinformation or no information), whereas a benevolent communicator is willing to share the information he has (as opposed to asserting false information because of indifference or malevolence). If communication has to remain advantageous on average (as its pervasiveness in our social interaction suggests it is), humans have to deploy an ability to calibrate their epistemic trust. This ability is 'epistemic vigilance'.

Sperber et al. (2010) conceive of epistemic vigilance as a cognitive adaptation for social exchange. As Cosmides and Tooby (1992, p. 166) suggest, "each cognitive specialisation is expected to contain design features targeted to mesh with the recurrent structure of its characteristic problem type". Thus, a closer investigation of its 'problem type' will shed some light on the nature and function of the cognitive mechanisms underpinning epistemic vigilance as a whole.[24]

The 'problem type' that represents the target of epistemic vigilance is the risk of misinformation in communication. Misinformation can be either accidental or intentional. The former is often the result of speaker's incompetence, the latter of speaker's malevolence. An incompetent speaker may communicate information that is false because she takes it to be true; a malevolent speaker may communicate false information with the intention of deceiving her interlocutor.

These alternative and recurrent features of misinformation suggest that some of the epistemic vigilance mechanisms should check for the reliability of the source of information, where reliability is a function of both speaker's competence and speaker's benevolence. In other terms, epistemic vigilance should help us with monitoring *who* to believe (i.e. individuals who are both competent and trustworthy).

The reliability of the source of information, however, is not the only factor affecting the believability of a piece of communicated information. The content of information may itself be more or less believable, independently of its source (with tautologies and logical contradictions lying at the two extremes of a continuum of believability). Thus, Sperber et al. (2010) argue for the existence of a second cluster of epistemic vigilance mechanisms, that is, mechanisms which assess the quality of the incoming information (i.e., *what* to believe). They check its factual plausibility by assessing its consistency with existing knowledge and its degree of evidence.

---

[24] Both Sperber and Cosmides and Tooby advocate the massive modularity view of the mind, that is, the view that the mind is a system of evolved cognitive mechanisms that are dedicated to a particular task (hence domain-specific) and interact with each other in constrained ways. I will discuss this view in Chapter 5, section 5.6.

Epistemic vigilance mechanisms towards the source of information can deliver either general impressions of trustworthiness (e.g. on the basis of facial clues, see Willis and Todorov (2006)) or more costly assessments that result from context-sensitive evaluations of the reliability of the speaker. For instance, they may assess whether the speaker may want the audience to believe the set of communicated propositions P_1, .., P_n for reasons that do not concern their (alleged) truth (e.g. because of some deceptive intention). Or they may detect that this set of propositions P_1, .., P_n is warranted by some beliefs of the speaker that are, in fact, false. In both cases, they would prevent P_1, .., P_n from entering the belief box of the interpreter.

A growing body of research on the development of the epistemic vigilance capacity towards the source indicates that this ability emerges very early in development (for a review, see e.g. Harris (2012), Robinson and Einav (2014)). Some form of epistemic vigilance may be present from the very age infants have actually been tested. For instance, as reported by Koenig and Harris (2007), when 16-month-olds saw pictures of familiar objects and heard accurate/inaccurate labels from (a) a human looking at the picture, (b) a human with her back to the picture, (c) an audio speaker, they tended to be surprised when label (a) was false, when label (b) was true, and not surprised either way by (c).

By 2 to 4 years of age, children employ a number of criteria for evaluating the reliability of the speaker. They show selective trust based on past accuracy (2-year-olds, see Koenig and Harris (2007)**),** speaker's attitude (indications of certainty/hesitation) (3 year-olds, see Matsui, Rakoczy, Miura and Tomasello, (2009)), true knowledge vs. past accuracy (4-year-olds, see Einav and Robinson (2011)).

As far as epistemic vigilance towards deception is concerned, children become able to cope with intentional deception from 4 to 6 years of age (Mascaro & Sperber, 2009). This capacity requires sophisticated mindreading abilities, as the interpreter needs to combine a first-order attribution of belief ('The speaker believes that not-P') with a second-order attribution of intention ('The speaker wants me to believe that P').

Epistemic vigilance mechanisms towards the content assess the believability of the incoming information relative to the context of the addressee's existing beliefs (which are themselves, of course, open to revision). As discussed in section 4.3.4, according to Sperber et al. (2010), the beliefs against which the communicated information is tested are those that are automatically activated by the comprehension process and used in the pursuit of relevance. These are a subset of the mental encyclopaedia of the addressee, and provide the ground for an "imperfect but cost-effective epistemic assessment (Sperber et al. 2010, p. 374). When the result of this assessment is a contradiction, there are three possible outcomes: (i) if the source is taken as trustworthy and the background beliefs of the interpreter that conflict with the incoming information are not held with much conviction, these beliefs are corrected; (ii) if the source is not regarded as trustworthy, the new information is rejected; (iii) if the source is regarded as authoritative and the conflicting background beliefs are held confidently, some process of (typically conscious) coherence checking is triggered. The choice among (i), (ii), and (iii) partly depends upon the output of epistemic vigilance mechanisms focused on the source (the speaker).

In conclusion, according to Sperber et al. (2010), the gap between comprehension and acceptance is bridged by epistemic vigilance mechanisms, which play a significant role in filtering incoming information with the aim of minimising the risk of misinformation. In the next chapter, I will address the question of what role (if any) epistemic vigilance mechanisms may play in the comprehension process itself.

Chapter 5 The role of epistemic vigilance in comprehension

In Chapter 3, I argued that pragmatic interpretation is 'inferential' in the sense that it relies on two distinct stages of 'hypothesis formation' and 'hypothesis confirmation'. In the second stage, interpretative hypotheses are assessed against a criterion of pragmatic acceptability based on consideration of the speaker's mental states (i.e. her beliefs and intentions). In this chapter, I propose that this two-stage process is underpinned by the interaction of two distinct systems: a relevance-guided comprehension procedure (Wilson & Sperber, 2004) and epistemic vigilance mechanisms, which assess the quality of incoming information and the reliability of the individual who dispenses it (Sperber et al., 2010).

While the scope of this interaction has not been much explored, its centrality has already been recognised:

> […] the abilities for overt intentional communication and epistemic vigilance must have evolved together, and must also develop together and *be put to use together*. (Sperber et al., 2010, p. 360, *my emphasis* (DM))

This passage suggests three different perspectives that are relevant to the investigation of epistemic vigilance in communication: an evolutionary perspective, a developmental perspective, and a 'pragmatic' perspective. The main focus of my investigation will be the pragmatic perspective, but I will also discuss its implications for the development of the pragmatic abilities.

As discussed in section 4.3.4, Sperber and colleagues suggest that comprehension and epistemic assessment are two distinct processes which work in parallel with each other:

> […] several mechanisms may work in parallel or in competition. For instance, it could be that any piece of communicative behaviour activates two distinct processes in the addressee: one geared to identifying the relevance of what is communicated on the assumption that it is trustworthy, and the other geared to assessing its trustworthiness. (Sperber et al., 2010, p. 364)

The question that arises is whether the capacity for epistemic vigilance plays a role in comprehension (i.e. the process of identifying the speaker's meaning). Sperber et al. (2010) seem to suggest a negative answer. While comprehension and epistemic assessment are seen as parallel processes triggered by the same piece of communicative behaviour ('The speaker has uttered *u*'), their interaction has a limited scope. The only role of the epistemic vigilance system is to assess the believability of the interpretation *resulting from* the comprehension process (in light of considerations about both communicator's reliability and content's believability). In what follows I suggest that the interaction between comprehension and epistemic assessment has a wider scope than previously assumed and that, as a result, it may be more finely articulated.

5.1 Interpretative strategies

It is worth beginning our investigation of the relationship between comprehension and epistemic assessment by considering examples that give rise to clear off-line intuitions about how our interpretative practice might be affected by considerations about the moral and epistemic trustworthiness of our interlocutors.

Sperber (1994b) invites us to consider the following scenario. Imagine that Carol and John are going to a party and they have planned to leave their child at home with the baby sitter. The baby sitter usually leaves at midnight. That day, however, thinking that the party would be great fun, Carol has made a special arrangement with the babysitter and she will stay until one. Crucially, Carol does not know that John is aware of this. Later that night, Carol is not enjoying the party and, at around 11.30pm, she says to John "It's late" expecting him to think that it is time to go home because of the baby sitter. The interpretation of Carol's utterance "It's late" depends on whether the addressee (John) *trusts* the speaker (specifically, on whether he assumes that Carol is behaving benevolently). If John assumes Carol's benevolence, he will be bound to misunderstand her. She could not intend to communicate something that she knows to be false, so he would take it that she intended to communicate that it is late with respect to some other schedule or expectation (for

instance, that it is late if they want to catch the last train to get home). On the other hand, if he recognises that Carol is trying to deceive him, he will correctly attribute to her the intention to communicate that it is time to go home because of the baby sitter.

Let us consider a modified version of this scenario. Imagine that Carol and John are at the party and no special arrangement has been made with the baby sitter, who will leave, as usual, at midnight. Unbeknownst to Carol, John is very worried about a delivery that should have been made that very day. At 11.30pm, Carol says to John "It's late" expecting him to think that it is late because of the baby sitter. Because he is caught up in his thoughts, the first interpretation to come to John's mind is that the delivery is late. Once again, the interpretation of the utterance "It's late", which is eventually attributed to Carol, depends on whether the addressee trusts the speaker (this time, on whether John assumes that Carol is competent, that is, she possesses genuine information as opposed to misinformation or no information). If he does, he will be bound to misunderstand her by mistakenly attributing to her the intention to communicate that the delivery is late. If he realises that she could not intend to communicate something that she does not know, he would take it that she intended to communicate that it is time to go home because of the baby sitter.

These examples, taken together, suggest that considerations of the speaker's benevolence and competence may affect the way in which we interpret what she says. An adequate account of pragmatic processing should shed some light on how this happens to be the case.

With regard to this, Sperber (1994b) suggested that competent interpreters have sophisticated interpretative strategies at their disposal, which allow them to cope with deliberate deception or to avoid misunderstandings due to speaker's incompetence. This proposal has been given only a relatively marginal role within the development of Relevance Theory in the following years. In what follows, I present the interpretative strategies proposed by Sperber (1994b) and suggest the existence of an interesting link between these strategies and the operations of epistemic vigilance mechanisms (section 5.2). This suggestion is thus grounded on

the well-established relationship between interpretation and trust, but it offers a new cognitively oriented perspective in which to frame such a relationship.

Sperber (1994b) suggests the existence of three interpretative strategies, which he labels 'naïve optimism', 'cautious optimism' and 'sophisticated understanding', which can be seen as different versions of the relevance-guided comprehension procedure:

(3) **Relevance-guided comprehension procedure**
   a. Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.
   b. Stop when your expectations of relevance are satisfied.

As emphasised by Wilson and Sperber, clause (b) of the relevance-guided comprehension procedure "[..] allows for varying degrees of sophistication in the expectations of relevance with which an utterance is approached." (Wilson & Sperber, 2004, p. 625). Importantly, the difference between the three strategies relies on different assumptions about the communicator's competence and benevolence, which in turn raise different expectations of relevance (hence determine different stopping points in interpretation). A naïvely optimistic hearer takes for granted that the communicator is behaving both benevolently and competently: he takes the communicator to be competent enough to avoid misunderstanding, and benevolent enough not to lead him astray. Thus he expects 'actual optimal relevance'. In contrast, a cautiously optimistic interpreter assumes the communicator to be benevolent, but not necessarily competent. As a consequence, he looks for 'attempted optimal relevance'. Finally, a sophisticated interpreter drops not only the assumption that the communicator is behaving competently, but also that she is behaving benevolently. Then the expectations of relevance that guide the comprehension procedure and determine its stopping point are expectations of 'purported optimal relevance'. The following table illustrates the three different versions of the relevance-guided comprehension procedure (which differ with regard to clause (b)):

Three versions of the relevance-guided comprehension procedure:

	(a)	Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses in order of accessibility.
Naïve optimism	(b_1)	Stop when your expectations of *actual optimal relevance* are satisfied (i.e. stop at the first relevant enough interpretation)
Cautious optimism	(b_2)	Stop when your expectations of *attempted optimal relevance* are satisfied (i.e. stop at the first interpretation that the communicator *might have thought* would be relevant enough to you)
Sophisticated understanding	(b_3)	Stop when your expectations of *purported optimal relevance* are satisfied (i.e. stop at the first interpretation that the communicator *might have thought would seem* relevant enough to you)

To appreciate the difference between these interpretative strategies, consider again the examples discussed above. With regard to the delivery-example, if John were a naively optimistic interpreter, he would attribute to Carol the first interpretative hypothesis that is relevant enough to him. The first interpretation that comes to John's mind is that the delivery is late. Given its relevance to John, a naïve interpreter would retain it and mistakenly attribute it to the speaker. But what if John adopted the cautiously optimistic version of the relevance-guided comprehension procedure? John would not take for granted Carol's competence and he would be vigilant to the possibility that Carol may not know what he knows (and may consequently fail in her attempt to make the relevant information that she intends to convey more accessible than any other possible interpretation). John would realise that Carol could not have intended the interpretative hypothesis *The delivery is late* to occur to him (precisely because she does not know that he is waiting for a delivery). Carol could not have thought that this interpretation would be relevant enough to him as she, Carol, has no thoughts of any sort involving this delivery. Thus, the comprehension procedure would go further and test the next most accessible interpretative hypothesis. For instance, it would access and assess the interpretation that it is time to go home because of the baby sitter. Since John takes it that Carol might have thought this interpretation to be relevant enough to him (as in fact it is), the interpretation is selected and attributed to Carol. With regard to the

deceptive version of this example, it is possible to show how both a naïve and a cautiously optimistic interpreter would fail in attributing to the speaker the intended interpretation. Let us assume that the first interpretation to come to John's mind is the (intended) interpretation that it is time to go home because of the baby sitter. However, John immediately realises that this is not the case, as he knows that Carol and the baby sitter have made a special arrangement for that night and that the baby sitter will leave later than usual. If John were a naively optimistic interpreter, he would discard that interpretation, as it is not relevant to him (he knows it to be false). If he were a cautiously optimistic interpreter, he would also discard it given that it is not an interpretation that Carol might have thought would be relevant to him (having made the arrangement herself, Carol knows that it is not the case that it is time to go home because of the baby sitter). Only the adoption of a more sophisticated interpretative strategy would allow John to correctly attribute this interpretation to Carol. John would realise that Carol might have thought that this interpretation would *seem* relevant to him (as she does not know that John is aware of this special agreement) and, if he had reasons to think that she might want to deceive him, he would end up attributing that interpretation to her.

5.2 Sophisticated strategies and epistemic vigilance

Utterance interpretation may depend on considerations about the speaker's competence and/or benevolence. The issue of what brings such considerations to bear on the interpretative process, however, has not been addressed within the literature. My proposal is that the expectations of relevance which guide the comprehension procedure and determine its stopping point are directly modulated by the operations of epistemic vigilance mechanisms. That is, epistemic vigilance mechanisms can modulate the hearer's expectations of relevance (i.e. from 'actual' to 'attempted' or 'purported' optimal relevance) and assess whether the interpretative hypothesis under construction satisfies these expectations.

If the interpreter is vigilant towards the speaker's competence, for instance, he will expect 'attempted' optimal relevance. As a consequence, he will stop at the first relevant interpretation that the speaker might have thought would be relevant to him (as described in the cautiously optimistic version of the relevance-guided

comprehension procedure). If the interpreter is vigilant towards the speaker's competence as well as her benevolence, he will expect 'purported' optimal relevance. In this case he would stop at the first interpretation that the speaker might have thought would seem relevant enough to him.

In the same vein as my proposal, Padilla Cruz (2012) has suggested that epistemic vigilance should be considered as the trigger for a shift in interpretative strategies. For instance, if epistemic vigilance detects that the interlocutor is not a very competent communicator, it may trigger a shift from naïve optimism to cautious optimism. I believe, though, that the recent work on epistemic vigilance should be seen as encompassing Sperber's (1994b) original proposal. Once epistemic vigilance is brought into the picture, the three interpretative strategies are found to be redundant. For instance, a cautiously optimistic interpreter may be seen not as an interpreter who is prompted to adopt a particular strategy by his epistemic vigilance mechanisms (as Padilla Cruz suggests), but rather as an interpreter who is actively monitoring the speaker's competence through his epistemic vigilance mechanisms. A very interesting and plausible picture emerges: the three interpretative strategies described above may simply be an *epiphenomenon* of the interaction between a single comprehension procedure and epistemic vigilance mechanisms.

Before developing this proposal with regard to the examples under discussion, it is worth noting that Sperber and colleagues suggest that epistemic vigilance mechanisms towards the source of information can deliver either general impressions of trustworthiness or more costly assessments that result from context-sensitive evaluations of the reliability of the speaker. With regard to the latter, they note the following:

> Clearly, the same informant may be competent on one topic but not on others, and benevolent towards one audience in certain circumstances, but not to another audience or in other circumstances. This suggests that trust should be allocated to informants depending on the topic, the audience, and the circumstances. (Sperber et al., 2010, p. 369)

The definition of trustworthiness provided by Sperber et al. (2010) is thus intrinsically context-dependent; and it could not be otherwise. To elaborate on this

point, let us focus on speaker's competence. For every speaker, there is always some information that she does not possess and some false assumptions that she takes to be true. However, this is not what 'competence' is about. If this were the case, every speaker would have to be classified as incompetent and would not be entitled to receive our epistemic trust. Competence has a narrower and context-sensitive scope: the same communicator may be competent on one topic but not on others. The investigation of epistemic vigilance mechanisms that can assess competence (as well as benevolence) in a context-sensitive way will prove to be crucial for a general understanding of epistemic vigilance, and my proposed interaction with the comprehension system.

To clarify the dynamics of the hypothesised interaction between the relevance-guided comprehension procedure and epistemic vigilance mechanisms, let us consider again the example "It's late", where the first interpretative hypothesis to come to the hearer's mind is that the delivery is late but this does not correspond to the intended interpretation.

My suggestion is that the construction of an interpretative hypothesis provides a hypothesised topic of conversation. This, in turn, serves as input to epistemic vigilance mechanisms which assess the competence of the speaker on a particular topic. In this case, the interpretative hypothesis that the delivery is late provides a hypothesised topic of conversation (i.e. the delivery) with regard to which epistemic vigilance mechanisms assess Carol's competence. These mechanisms access the piece of information that Carol does not know that John is waiting for a delivery. As a consequence, an incompatibility between the speaker's system of beliefs and the interpretative hypothesis under construction is detected. This inhibits the comprehension procedure and prompts it to access (and assess) the next most accessible interpretative hypothesis.

Let us now consider the deceptive version of this example, where Carol's utterance "It's late" is intended to be interpreted as implicitly communicating that it's time to go home because of the baby sitter. Suppose that the first interpretation to come to John's mind in this context is the interpretation that it is time to go home because of the baby sitter (where this conclusion is warranted by the explicature of "It is late"

and the implicated premise *The baby sitter typically leaves at midnight*). Once again, this provides a hypothesised topic of conversation, that is, the baby-sitter arrangement, which, as in the previous example, triggers epistemic vigilance mechanisms which assess the speaker's competence on that topic. These mechanisms access the piece of information that Carol believes that it is not the case that the baby sitter will leave at midnight. Given that the interpretative hypothesis is not compatible with the speaker's epistemic state, a cautiously optimistic hearer would discard this hypothesis and look for a different interpretation. However, if the hearer's epistemic vigilance mechanisms targeted at assessing the speaker's benevolence detect that the speaker has the deceptive intention of making the hearer (falsely) believe that it is time to go home because of the baby sitter (and if this intention is also compatible with what the hearer believes the speaker takes to be the hearer's epistemic state, i.e. Carol does not know that John knows that the baby sitter will stay until one), the interpretation would be retained and attributed to the speaker. As Sperber and colleagues suggest, "[w]hen epistemic vigilance is targeted at the risk of deception, it requires an understanding not only of the communicator's epistemic states but also of her intentions, including intentions to induce false beliefs in her audience" (Sperber et al, 2010, p. 372). In our example, the hearer understands that Carol believes that not-P but she wants him to believe that P (P = it is time to go home because of the baby sitter).

As the discussion of these examples illustrates, not only do epistemic vigilance mechanisms affect the *believability* of a piece of communicated information (as proposed by Sperber and colleagues), but they also contribute to the assessment of the *acceptability* of interpretative hypotheses. It is crucial to appreciate here the conceptual distinction between 'believability' and 'acceptability'. The first notion concerns the extent to which an interpretation attributed to the speaker (i.e. the output of the comprehension procedure) is allowed to enter the 'belief box' of the interpreter (i.e. whether the interpreter ends up believing it or not). The second notion, 'acceptability', concerns whether an interpretative hypothesis about the speaker's meaning is retained and attributed to the speaker as the intended

interpretation (i.e. whether it ends up being the output of the comprehension module or not).[25] As far as the latter is concerned, epistemic vigilance mechanisms may filter out interpretative hypotheses that, although relevant, are incompatible with the speaker's mental states (i.e. her beliefs and desires). In this case, they may prompt the comprehension process to continue and assess further interpretative hypotheses. In other circumstances, they may prevent the comprehension procedure from abandoning an interpretative hypothesis that is irrelevant (to the hearer, e.g. he knows that it is false) but compatible with the speaker's mental states (e.g. her intention to induce false belief in the hearer).

It is worth investigating the implications of this proposal for the relationship between comprehension and epistemic assessment. Specifically, the question of whether the epistemic assessment that is involved in comprehension exhausts the validation-process for communicated information shall be addressed. According to my proposal, the effects of epistemic vigilance on comprehension and acceptance can be either *simultaneous* (when no further epistemic assessment beyond that involved in comprehension is needed) or *serial*. When the addressee arrives at the intended interpretation via the recognition that the communicator is trying to deceive him, it is plausible to assume that no further epistemic assessment will be undergone and the output of the comprehension process will be automatically prevented from entering the addressee's belief box. This seems to be the case in the deceptive version of the example discussed above. If John correctly reaches the intended interpretation that it is time to go home because of the baby sitter via the recognition of Carol's deceptive intention, he won't further assess the believability of this interpretation after attributing it to Carol. However, when the role played by epistemic vigilance is that of warranting an interpretation that is compatible with the speaker's system of beliefs, it is still an open issue whether or not the interpretation selected is worth being accepted as true (i.e. believed). For instance, in the delivery-example discussed above, John recognises that Carol could not intend to communicate that the delivery

[25] To avoid a further terminological confusion, it is worth noting that the term 'acceptance' is typically used within the literature (and in this thesis) in relation to the output of the pragmatics system (which can be more or less believable, and consequently accepted as true (or not)). This is not to be confused with the term 'acceptability' which refers to interpretative hypotheses (rather than interpretative outputs).

is late, as she is not aware that he is waiting for a delivery. As a consequence, he looks for an interpretation that is compatible with Carol's system of beliefs, e.g. that it is late if they want to catch the last train. Once this interpretation has been attributed to Carol as the intended interpretation, it is still an open question whether John should accept it as true or not. For instance, it might contradict some of his strongly held beliefs concerning the recent introduction of a 24-hour train service and could thus be rejected as false. In this case, rejection (or acceptance/belief) will be the result of a process of epistemic assessment that takes place on the *output* of the comprehension system.

To conclude this section, let us explore the implications of this hypothesised interaction between the comprehension system and epistemic vigilance mechanisms with regard to the relationships among Sperber's (1994b) three interpretative strategies:

> Much of everyday communication takes place between people who are benevolent to one another and who know one another well enough. In such circumstances, cautious, and even naïve optimism can serve as default interpretation strategies […] Still, *when the optimistic strategies fail*, a competent hearer resorts to the sophisticated strategy. (Sperber, 1994b, p. 198, *my emphasis* (DM))

In line with this, if we assume that more sophisticated interpretative strategies are implemented by the operations of epistemic vigilance mechanisms, we can endorse the idea that optimistic interpretative strategies might represent a 'preferred' option, but reject the claim that optimistic strategies must fail in order for the interpreter to resort to 'sophisticated understanding'. Rather, addressees may start by being vigilant and only grant trust if they have no reasons to doubt the communicator's competence and benevolence. If they do have reasons to doubt her trustworthiness, they will downgrade their initial expectations of 'actual' optimal relevance to expectations of 'attempted' or 'purported' optimal relevance. As emphasised by Sperber et al. (2010), epistemic vigilance is not the opposite of trust, but it is the opposite of *blind* trust. To clarify this point, Sperber et al. (2010, p. 346) develop the following enlightening analogy: when we walk down the street through a crowd of people, we typically do not have any hesitation about walking among them, despite

the risk of being accidentally or intentionally hit by them. However, we do monitor the trajectory of others and if we detect the presence of a careless or aggressive individual, we raise our level of vigilance. This low-level, unconscious, vigilance allows us to enjoy our stroll while preventing us from any risk of collusion. In other words, our mutual trust is buttressed by our mutual vigilance.

5.3 Implications for developmental pragmatics

Sperber (1994b) suggests that the three versions of the relevance guided comprehension procedure ('naïve optimism', 'cautious optimism' and 'sophisticated understanding') may correspond to different stages in pragmatic development. That is, children may start out as naïvely optimistic interpreters and progressively acquire the ability to monitor the speaker's competence and benevolence and to adapt their interpretative behaviours to these.

This gives rise to the interesting question of what allows the progression from naïve optimism to the further developmental stages. Sperber (1994b) claims that the three versions of the relevance-guided comprehension procedure require the interpreter to manipulate increasingly higher order representations of mental states. For this reason, Carston (1997) and Wilson (2005) have suggested that the development of more sophisticated interpretative strategies may correlate with the emergence of more complex mind-reading abilities: the move from naïve optimism to cautious optimism may correlate with the emergence of first-order mind-reading ability, the one from cautious optimism to sophisticated understanding with the emergence of second-order mind-reading abilities.

While this suggestion is certainly worth exploring further, the recent work on epistemic vigilance by Sperber et al. (2010) seems to open further interesting scenarios for developmental pragmatics. In light of the hypothesised interaction between the comprehension system and epistemic vigilance mechanisms, it seems plausible to assume that these three stages in the development of pragmatic abilities may follow a similar developmental trajectory to that of epistemic vigilance capacities. Naïve optimism, cautious optimism and sophisticated understanding involve different assumptions about the communicator's competence and

benevolence. As noted above, epistemic vigilance mechanisms focused on the source of information (*who* to believe) monitor the speaker's epistemic and moral reliability, that is, her competence and benevolence. Thus, it seems plausible to hypothesise that the emergence of epistemic vigilance mechanisms targeted at the assessment of the speaker's competence may correlate with (and prompt) the development of a cautiously optimistic interpretative strategy. Similarly, the move to sophisticated understanding may be triggered by the emergence of epistemic vigilance mechanisms monitoring the speaker's benevolence.

While the development of epistemic vigilance has been the subject of extensive experimental investigation (see section 4.4), an explicit comparison between the development of epistemic vigilance and of pragmatic competence remains to be carried out. Furthermore, there is a growing body of research on children's ability to track the communicator**'**s epistemic state and use this to infer what she intends to refer to (Carpenter, Call, & Tomasello, 2002; Southgate, Chevallier, & Csibra, 2010; i. a.). The implications of this literature for Sperber's (1994b) theoretical distinction between 'naïve optimism', 'cautious optimism' and 'sophisticated understanding' have not been assessed yet. Crucially, Southgate et al. (2010) show that 17-month-old infants can take account of the speaker's epistemic state (i.e. her false-belief) in reference resolution. In Southgate et al.'s study, the infants see the experimenter place two novel objects in different boxes and leave the room. An accomplice changes the position of the objects in her absence. When the experimenter returns, she points towards one of the boxes and says to the infant: "Do you know what's in here? Shall we play with it? Shall we play with it? Let**'**s play with it!". Finally she says, "Can you get it for me?". The issue here is which of the two objects the infant would take the word 'it' to refer to. The results showed that infants as young as 17-month-old were significantly more likely to choose the object in the box that the experimenter had not pointed to. What I would like to point out here is that the infant assigning the appropriate referent to the pronoun 'it' seems to require a cautiously optimistic interpretative strategy. That is, the infant should not stop at the first relevant referential interpretation (which corresponds to the one where the referent is taken to be inside the pointed-to box), but rather at the first relevant referential interpretation that the experimenter could have thought would be relevant to him.

This requires the infant to take account of the experimenter's epistemic state (i.e. her false belief that the intended object is in the pointed-to box) and reason that she could have not intended to refer to the object in the pointed-to box because she does not know that it has been swapped with the object in the non-pointed-to box. The results raise the question of whether there is in fact any developmental stage corresponding to naïve optimism or whether naïve optimism is a theoretical construct without any empirical counterpart.

If my proposal is on the right track, children's ability to pass the false belief task with referring expression will be built on the progressively improving capacity for epistemic vigilance which manifest itself from 16 months after birth (see, e.g. Koenig and Harris (2007)). It is also worth noting that further studies have shown that children become able to cope with intentional deception between 4 and 6 years of age and that this ability strongly correlates with success in irony comprehension (see, e.g., Winner & Leekam, 1991; Sullivan, Winner & Hopfield, 1995; Winner et al., 1998, Mascaro & Sperber, 2009). Given this, it would be interesting to test whether this period corresponds to one in which the passage from cautious optimism to sophisticated understanding occurs.

5.4 Relevance and epistemic vigilance

The discussion so far has focused on how the interpretative process is constrained by considerations about the speaker's mental states. Addressees look for interpretations that are compatible with the speaker's system of beliefs and intentions. The rationale behind this is that speakers cannot be expected to go beyond their competence or against their interests. The core idea is that an interpretative hypothesis is tested against the information retrieved by epistemic vigilance mechanism towards the source and, if no incompatibility is detected, the interpretative hypothesis is attributed to the speaker as the intended interpretation. The information against which the interpretative hypothesis is tested concerns the speaker's epistemic state (e.g. her beliefs) and other mental states (e.g. her desires and intentions), monitored by epistemic vigilance.

At this point a subtle but crucial remark is needed: the set of information against which an interpretative hypothesis is tested is not the speaker's system of beliefs but *what the interpreter takes to be* the speaker's system of beliefs. The latter does not generally coincide with the former, at least not entirely: this is not only because the set of beliefs on a particular topic that the interpreter attributes to the speaker is usually smaller than her actual set of beliefs on that topic, but also because the interpreter may be mistaken. He may assume that the speaker believes that P, while she may have no beliefs about P (or even believe that *not*-P).

Importantly, the set of beliefs that the interpreter attributes to the speaker is constantly updated and revised in light of new evidence. This revision can occur through communication in two different ways. On the one hand, the speaker may explicitly state that she does not believe that P (in a context in which the interpreter assumed that she believed that P). On the other hand, the interpreter may be forced to revise his assumption that the speaker believes that P in order to make sense of something that the speaker has said (whose only sensible interpretation is incompatible with that assumption).

In what follows, I investigate this second scenario. The aim is to shed some light on the intricate interaction between the interpretative process and epistemic vigilance mechanisms. Such interaction is to be explored in a 'bi-directional' way. Not only what the interpreter takes to be the speaker's system of beliefs can affect the interpretative process, but the interpretative process can modify what the interpreter takes to be the speaker's system of beliefs.

To illustrate this point, let us consider again example (25) from Sperber et al. (2010, p. 368). Imagine that Barbara has asked Joan to buy a bottle of champagne to the party that Barbara and Andy are organising. After reporting this to Andy, the following exchange takes place:

(25) a. *Andy (to Barbara):* A bottle of champagne? But champagne is expensive!
b. *Barbara:* Joan has money.

As discussed in section 4.3.4, if we imagine that Andy had previously assumed that Joan was an underpaid junior academic, the only interpretation that is consistent with such an assumption is that Joan has *some money* (as opposed to no money). Despite this, he would interpreter Barbara's utterance as communicating that Joan has *enough money to be easily able to afford champagne* since this is the only interpretation that can satisfy Andy's expectations of relevance in the context at hand. He can then decide whether or not to believe it (and whether or not to abandon his own assumption about Joan's financial situation), but he interprets Barbara's utterance as asserting a proposition that would be relevant enough to him provided that he believes it.

Let us imagine now that the same conversation occurs in a modified scenario. In this scenario, Andy had not only assumed that Joan was an underpaid junior academic, but also that Barbara believed the same. If this were the case, the interpretative hypothesis that Joan has *enough money to be easily able to afford champagne* would conflict with what the interpreter, Andy, takes to be Barbara's system of beliefs. This means that the interpretative hypothesis would conflict with the information about the speaker's epistemic state retrieved and deployed by epistemic vigilance.

I previously argued that the incompatibility between an interpretative hypothesis and the information retrieved by epistemic vigilance should result in the abandonment of the interpretative hypothesis at issue and in the assessment of further interpretative hypotheses. But what if no interpretative hypothesis derivable from the relevance-theoretic comprehension procedure is consistent with what the interpreter takes to be the speaker's system of beliefs? If interpreting Barbara's utterance as conveying that Joan has *enough money to be easily able to afford champagne* is the only way to reach a relevant enough interpretation (provided he believes it), Andy will consequently adjust his previous assumptions about Barbara's beliefs.

Barbara could not have tried to be optimally relevant if she had thought that Joan was a junior underpaid academic and, despite this, communicated that Joan had *some* money (as opposed to no money) in reply to Andy's remark (i.e. "But champagne is

expensive!"). If Andy has no reason to think that Barbara is being dishonest, he will then adjust his own beliefs about Barbara's beliefs.[26]

The general conclusion to be drawn is that epistemic vigilance mechanisms assessing the competence of the communicator are fallible; they may retrieve information about the speaker's beliefs that is not correct (e.g. it is not true *that Barbara believes that Joan is an underpaid junior academic*). This is the reason why, in some circumstances (e.g. when the benevolence of our interlocutor is not in question), it is rational to give it up. For instance, when *the only* interpretation which satisfies the hearer's expectation of relevance is incompatible with what the hearer takes to be the speaker's system of beliefs, the interpreter will revise his assumption in order to reach a (sufficiently relevant) interpretation of the speaker's utterance. This is why the interpretative process "involves a readiness to adjust one's own beliefs to a relevance-guided interpretation of the speaker's meaning, as opposed to adjusting one's interpretation of the speaker's meaning to one's own beliefs" (Sperber et al. (2010, p. 368). This also includes a readiness to adjust one's own *beliefs about the speaker's beliefs* to a relevance-guided interpretation of the speaker's meaning.

5.5 Inferential approaches to pragmatics and epistemic vigilance

Mazzone (2009, 2013) has presented two sorts of objection against the role attributed by Relevance Theory to expectations about 'optimal relevance' as a criterion of pragmatic acceptability. In particular, he has questioned whether Relevance Theory provides an adequate account of how considerations about the communicator's mental states affect pragmatic interpretation:

> Intention-reading is not thought to drive the search for intended interpretations from the beginning; rather, it is described as a *filter* on interpretations which are detected by the comprehension procedure. This is another way to state that intention-reading is distinct from the assessment of relevance, and probably *subsequent* to it. However, we are

[26] Andy may come up with reasons to think that Barbara is trying to deceive him. For instance, he may suspect that Barbara is trying to make him think not only that Joan has a lot of money but also that *she thinks* that Joan has a lot of money (while, in fact, she knows that Joan is an underpaid junior academic). If this is the case, Andy will not eventually end up revising his system of beliefs about Barbara's beliefs.

never told *how that filter could work*. (Mazzone 2009, p. 325, *my emphasis* (DM))

The first objection concerns the stage at which considerations about the communicator's mental states are supposed to affect the interpretative process. The second concerns the account of the cognitive mechanisms that are involved in recognising the communicator's mental states and putting them to use. I address these objections in turn and provide a reply to both of them in light of my proposed interaction between the comprehension system and epistemic vigilance mechanisms.

5.5.1 *When* do considerations about the communicator's mental states enter the picture?

According to Mazzone (2009), Relevance Theory conceives of the process of utterance comprehension as involving two different components: one responsible for forming interpretative hypotheses (based, in part, on considerations of accessibility), the other for consideration of the communicator's mental states that bear on the interpretation. The latter would thus discard interpretative hypotheses that are found to be incompatible with the communicator's beliefs and other mental states, by acting as a filter on unwanted interpretations. Indeed, this seems to be the path followed by addressees in interpreting examples like the ones discussed above. Let us consider again example (16) (see Chapter 3, section 3.1). The context hypothesised was one in which, at some point during the conversation between Robyn and her student Sarah, Sarah utters:

(16) Neil has broken his leg.

Crucially, Robyn knows two people called "Neil", her young son (NEIL_1) and a colleague in the linguistics department where she works (NEIL_2). While Sarah is acquainted with NEIL_2, she does not know that Robyn has a son called "Neil". Assuming that the first interpretation that comes to Robyn's mind is that NEIL_1 HAS BROKEN HIS LEG, it would be discarded as incompatible with the speaker's mental states (i.e. her epistemic state).

Mazzone (2009) claims that this analysis restricts the stage at which considerations about the communicator's mental states can affect pragmatic interpretation: an interpretative hypothesis is constructed, independently of any consideration of the communicator's mental states, and it is *subsequently* tested against them. For this reason, it may be seen as making specific empirical predictions about the time-course of the integration of information about the speaker's knowledge and beliefs in utterance comprehension. Specifically, it may generate the prediction that information about the communicator's mental states cannot be immediately integrated in on-line language processing. I do not explore this implication here but the recent debate on perspective-taking in the psycholinguistics literature sheds some doubt on the adequacy of such a prediction.[27] So this could be a problem for Relevancy Theory, if this prediction indeed followed from it. I suggest, though, that Relevance Theory is not committed to this, and that the picture sketched by Mazzone (2009) does not exhaust the ways in which considerations about the communicator's mental states are allowed to affect utterance interpretation within the relevance-theoretic framework. Let us focus on the relevance-guided comprehension procedure:

(3) **Relevance-guided comprehension procedure**

a. Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.
b. Stop when your expectations of relevance are satisfied.

Mazzone's (2009) suggestion that "if something metapsychological has to happen, it must take place outside the procedure" seems to imply that the path of least effort through which interpretative hypotheses are constructed and tested could never be a path that involves consideration of the communicator's mental states. However, this is simply not the case. Relevance Theory is perfectly compatible with the idea that, in many circumstances, such information may be so highly activated in the interpreter's mind that the 'least effort' interpretative hypothesis is exactly the interpretative hypothesis that is built on the basis of it. This hypothesis would thus be

[27] See Brown-Schmidt and Hanna (2011) for an overview.

accessed on the first processing pass through the communicator's utterance and no subsequent adjustment on the basis of considerations about the communicator's mental states would be required.

What bearing might this have on the operation of epistemic vigilance mechanisms? Epistemic vigilance mechanisms may play a role not only in the assessment but also in the *construction* of interpretative hypotheses. Epistemic vigilance mechanisms work in parallel and interact with the relevance-guided comprehension procedure. An interpretative hypothesis, or some parts of it, can be fed up to the epistemic vigilance mechanisms for assessment while comprehension is still in process. If this interaction is plausible, then there seems to be no principled reason why it should take place only at a particular stage of the interpretative process (e.g. when an interpretative hypothesis is assessed in order to decide whether or not it can be attributed to the speaker). Rather, it is plausible that in those circumstances in which our epistemic vigilance is particularly alerted or the speaker's epistemic state (on a certain topic) particularly salient, epistemic vigilance can constrain the construction of interpretative hypotheses from the very beginning.

Examples of early effects of epistemic alertness may be found in different communicative settings, but there are certain settings that may be more likely to manifest this feature. For instance, conversational settings that display an asymmetry between the interlocutors, such as pedagogical settings or, more generally, communicative interactions between adults and children, may be characterised by a higher awareness on the teacher/adult's part of the risk of *accidental* irrelevance accompanied by a higher activation of epistemic vigilance mechanisms.

I believe that a closer investigation of the effect of epistemic vigilance mechanisms on the construction of interpretative hypotheses may shed new light on different pragmatic phenomena. For instance, the phenomenon of 'scalar implicatures' might constitute a fertile ground for the application of this idea. Breheny, Ferguson and Katsos (2013) have recently shown that the on-line incremental derivation of quantity implicatures is constrained by information about the speaker's knowledge state – the so-called 'epistemic step' in the derivation of quantity implicatures (e.g. the derivation of *Not all of the X* from an utterance of 'Some of the X'). Implicature

derivation is reduced when the speaker is assumed to lack knowledge concerning the stronger alternative (that is, whether or not all of the X is the case). This suggests that information about the speaker's competence may constrain the interpretative processes from the very beginning (i.e. incrementally).

To conclude, I suggest that the interaction between epistemic vigilance mechanisms and processes of utterance interpretation displays a dynamic and complex range of effects at different stages of interpretation: the construction of interpretative hypotheses, the assessment of their pragmatic acceptability and the assessment of the believability of interpretations attributed to the speaker.

5.5.2 Which cognitive mechanisms are involved?

Mazzone's second objection is that Relevance Theory does not provide an adequate account of the cognitive mechanisms by which information about the speaker's mental states is recognised and put to use in the interpretative process. He claims that the appeal to the notion of expectations about optimal relevance as a pragmatic criterion of acceptability is not enough to show *how* this information enters the picture.

Mazzone (2013) distinguishes between two different levels at which an explanation about a pragmatic phenomenon can be given: what he calls the "functional level", on the one hand, and a lower "cognitive level", on the other. The former provides a conceptual analysis of the phenomenon, whereas the latter describes the actual cognitive mechanisms underpinning it. With this distinction in mind, Mazzone's objection can be rephrased along the following line: Relevance Theory offers a functional description of how expectations about 'optimal relevance' work as a pragmatic criterion of acceptability, but it does not offer a cognitively-specified description.

Before addressing this objection, it is worth discussing Mazzone's (2013) own proposal about how Relevance Theory might attempt to implement its framework in order to cope with this explanatory gap. As Mazzone himself shows, his proposal is not effective for the purpose at issue and he concludes, on that basis, that Relevance

Theory cannot offer an adequate cognitive explanation of the notion of optimal relevance. While I agree that Mazzone's proposal is inadequate, I show that the conclusion he draws does not necessarily follow. Let us start by introducing the proposal and illustrating the notions it involves:

> One should rather show that MAIS *[Mutual Adjustment between Inferential Steps]* hypothesis has the resources to account for the role that, at a functional level, the notion of optimal relevance assigns to speaker-related information. (Mazzone, 2013, p. 110)

What Mazzone refers to as "mutual adjustment between inferential steps" is what Relevance Theory usually calls "mutual parallel adjustment" (see section 1.2). According to Wilson and Sperber (2004), the relevance-guided comprehension procedure, (3), subsumes three different sub-tasks concerning the construction, respectively, of appropriate hypotheses about explicit content, intended contextual assumptions (in relevance-theoretic terms, implicated premises) and of intended contextual implications (or implicated conclusions). These sub-tasks are not sequentially ordered. Thus, the interpreter is not required to *first* recover the explicit content of the utterance, *then* select a useful range of contextual assumptions, and *finally* derive the intended contextual implications. When the hearer has fairly precise expectations about the intended implications, the comprehension procedure can be effect-driven: explicatures are constructed with the purpose of warranting the intended effects. Let us consider again the following example from Wilson and Sperber (2004):

(4) a. *Peter:* Did John pay back the money he owed you?
b. *Mary:* No. He forgot to go to the bank.

The interpretation of Mary's utterance is driven by the expectation that it will achieve relevance by answering Peter's question. One of the logical forms (not yet disambiguated) of the utterance provides access to the contextual assumption that forgetting to go to the bank may prevent someone from repaying his debt. This can be used in order to derive the relevant contextual implication that John did not pay back the money, provided that Mary's utterance is interpreted as explicitly communicating that John forgot to go to the BANK_1 (where BANK_1 refers to the

financial institution and BANK_2 to the sloping side of a river). The explicit content of Mary's utterance is thus constructed with the purpose of warranting the expected conclusion concerning whether John did or did not pay the money back to Mary.

Now, let us consider the details of Mazzone's line of argument. Pragmatic inferences are constrained by considerations about the communicator's mental states. According to him, these considerations play the role of contextual assumptions in the rational reconstruction of the inferences at issue. From the point of view of the theorist, then, they can modulate the construction of interpretative hypotheses about the explicit content of the utterance and its implicated conclusions through a process of mutual parallel adjustment. However, Mazzone argues, this leaves open the following question: how are premises about the speaker's mental states injected into the derivation during actual pragmatic processing? According to Mazzone, Relevance Theory can only appeal to considerations about the accessibility of such information in order to answer this question. Contextual assumptions concerning the speaker's mental states may be injected into the derivation by "following a path of least effort". However, as discussed in Chapter 3, relevance-theorists (Carston, 2007; Mazzarella, 2011; Wilson & Carston, 2007) have shown that accessibility-based approaches to pragmatics fall short of explaining how information about the speaker's mental states gets prominence during pragmatic interpretation. It follows that Relevance Theory cannot provide an adequate answer to the question of how considerations about the communicator's mental states affect pragmatic interpretation.

In what follows, I support Mazzone's claim that 'mutual parallel adjustment' does not offer an adequate explanation of the cognitive underpinnings of the relevance-theoretic criterion of pragmatic acceptability (based on expectations about 'optimal relevance'). However, I suggest that 'mutual parallel adjustment' is not treated as, nor was it ever intended to be, an acceptability criterion within the relevance-theoretic framework.

Interpretative hypotheses about the explicit and the implicit content are 'mutually adjusted' so that the derivation of the latter from the former is sound.[28] Soundness requires that the conclusions of the derivation follow from (are warranted by) its premises. This is a necessary but not sufficient condition for an overall interpretative hypothesis about the communicator's meaning to be retained and attributed to her. As Sperber and Wilson (2002, p. 609) suggest, the interpretative process stabilises when hypotheses about explicit content and implicatures are "mutually adjusted, and jointly adjusted with the hearer's expectations of relevance". I illustrate this with an example. Let us consider again example (16), but with a slight modification of the scenario described. Imagine that the student Sarah runs into Robyn's office, while she is having a meeting with a colleague, and exclaims:

(16) Neil has broken his leg.

For the reason previously discussed, we may assume that the first interpretation to come to Robyn's mind is that $Neil_1$, her son, has broken his leg. This interpretation is not the intended one and may not be eventually attributed by Robyn to Sarah (because she realises that Sarah does not know $Neil_1$). However, the first interpretative hypothesis about the explicitly communicated content of Sarah's utterance (i.e. *that $Neil_1$ has broken his leg*) may warrant the intended implicated conclusion that Robyn needs to interrupt her meeting:

(26) Explicature: *$Neil_1$ has broken his leg.*

Implicated premises: *You should interrupt a meeting in case there is an emergency.*

Implicated conclusion: *The fact that Neil1 has broken his leg is an emergency. Robyn should interrupt her meeting.*

This example shows that inferential soundness is not enough for an overall interpretative hypothesis to satisfy the hearer's expectations of relevance, which

[28] Soundness is to be interpreted "in a sense that applies to non-demonstrative inferences" (Sperber & Wilson, 1998, p. 194).

drive the comprehension procedure (and determine its stopping point). Thus, inferential soundness guaranteed by mutual parallel adjustment does not coincide with the acceptability criterion proposed by Relevance Theory. Nicholas Allott (personal communication) expresses this point very clearly:

> In Sperber and Wilson's account, the implicature is warranted by being part of the best explanation for the behaviour, not (in general) by being seen to be supported by the proposition expressed. Rather, that the explicature is supportive of the implicature is a *constraint* on the hypothetical explanations generated. (*my emphasis* (DM))[29]

The difference between the roles played by 'constraints' on hypothesis formation, on the one hand, and by 'criteria' for hypothesis confirmation in pragmatic interpretation, on the other hand, suggests that 'mutual parallel adjustment' cannot (and should not) be seen as the cognitive mechanism underpinning the notion of 'optimal relevance'.

The model proposed in section 5.2 sheds some light on the question of how information about the speaker's mental states are put in use in pragmatic interpretation. Epistemic vigilance mechanisms towards the source would access these pieces of information and, as a result, modulate the expectations of relevance the drive the comprehension process. This illuminates the cognitive underpinnings of the process of 'hypothesis confirmation', which is a crucial and indispensable component of pragmatic interpretation.

5.6 Massive modularity, the pragmatics module and epistemic vigilance

In what follows I explore the implications of my proposal for the cognitive architecture of the mind. Specifically, I address the issue of whether the proposed extensive interaction between the comprehension system and epistemic vigilance mechanisms is compatible with a modular view of pragmatics. I aim to show that it is, but to do so will require careful analysis of how a system can be both appropriately encapsulated and sensitive to relevant system-external information.

---

[29] This passage comes from Allott's notes for the PhD course "Communication and Inference" (CSMN, UiO, 2013) which he has kindly sent to me.

5.6.1 The pragmatics 'module'

Sperber and Wilson (1986/1995) brought to the attention of the pragmatics community the question of the place of pragmatic abilities in the overall architecture of the mind. At that time, Fodor had already suggested that human cognitive architecture is partly modular by introducing the functional and architectural distinction between modular perceptual and linguistic processors, on the one hand, and non-modular higher-level central processes (of belief fixation), on the other (Fodor, 1983). This gave rise to the interesting question of whether pragmatics is to be thought of as a domain-specific modular system or as part of a domain-general central system.

Fodor's (1983) distinction between modular input systems (perception and language) and non-modular central thought processes is based on a precise characterisation of the nature of 'mental modules'. Modular mental systems are task-specific and relatively autonomous systems which contingently share the following properties: they operate in a mandatory and fast way, they are domain-specific, informationally encapsulated and generally associated with a fixed neural architecture, and they exhibit specific breakdown and developmental patterns. Fodor places particular emphasis on informational encapsulation: "The informational encapsulation of the input systems is […] the essence of their modularity" (Fodor, 1983, p. 70). A mental system is informationally encapsulated (or cognitively impenetrable) if it is rigidly restricted in its access to the full range of the organism's knowledge, beliefs and desires. Such a system cannot take account of (potentially relevant) information that does not belong to its proprietary database, that is, which lies outside its own task specific body of information for processing its particular domain of stimuli.

The Fodorian notion of 'module' seems hardly applicable to pragmatic processing given that the central undisputed property of utterance interpretation is sensitivity to a wide range of contextual or background information. A constant tenet of post-Gricean approaches to pragmatics (e.g. Relevance Theory and other contextualist accounts) is the assumption that linguistic meaning underdetermines not only what is meant, but also what is said or explicitly communicated, that is, the truth-conditional

content of the utterance (Carston, 2002). To illustrate this, let us consider again example (16).

(16) Neil has broken his leg.

The hearer of (16) must decide who the referential expressions 'Neil' and 'his' refer to. Does 'Neil' refer to NEIL_1 (the hearer's son) or NEIL_2 (her colleague in the linguistics department)? Does 'his' refer to the referent of 'Neil' or to a different male individual who got a broken leg from a scuffle with Neil_1/Neil_2? Furthermore, the hearer must decide when the event took place (at some generic time in the past? a few days ago? this morning?). An utterance of (16) may also convey some implicatures (intended implications): for instance, that Neil_2 cannot participate in the staff meeting because he is still at the hospital, or that Neil_1 cannot run in the school marathon because he has not entirely recovered from the bad accident in which he broke a leg. The linguistically encoded meaning of the utterance, thus, plays a minor, albeit crucial, role in the recovery of the communicated content of the utterance, both at the explicit and at the implicit level. Crucially, there is no principled restriction on the kinds of information that pragmatic processing can call on: perceptual information, background information stored in long-term memory, information that is part of the linguistic context of the utterance. Pragmatic processes appear to be 'informationally UNencapsulated' and an account of the structure and function of pragmatics in the mind needs to respect this feature of utterance interpretation.

In line with this, pragmatics has traditionally been conceived of as a non-modular central inferential process, i.e. a non-deterministic process of rational belief-fixing. Specifically, it has been described as the process of arriving at an interpretation of the utterance, that is, the process of fixing a belief about the speaker's communicative intention. This was indeed the standard position of Relevance Theory up to the late '90s (Sperber & Wilson, 1986/1995; Wilson & Sperber, 1986).[30]

[30] But see Kasher (1991) for a modular view of certain components of pragmatic knowledge (i.e. basic speech acts and talk-in-interaction). For a critical discussion, see Carston (1997).

However, Sperber and Wilson (2002) suggest that the interpretative process is carried out by a dedicated pragmatics or comprehension 'module', with its own principles and mechanisms. How does this fit with the traditional view of pragmatics as an informationally unencapsulated system? In order to answer this question, we first need to consider the conceptual transformation that has characterised the notion of 'mental module' itself. Although pragmatics is now conceived of as a module, it is not a Fodorian module. For this reason, it is worth taking a step back to look at the wider picture of the mind within which Sperber and Wilson's (2002) proposal is located. This is the view of the mind as 'massively modular', a position pioneered by the evolutionary psychologists Cosmides and Tooby (1992, 1994) and advocated by Sperber himself (e.g., Sperber, 1994a, 2001, 2005).

Starting from evolutionary considerations, Sperber (1994a) suggests that the mind is modular through and through, that is, that cognition is based on dedicated domain-specific mechanisms, as opposed to domain-general central processes. This view, which takes the name of 'massive modularity', subverts Fodor's architectural taxonomy of psychological processes, and introduces 'conceptual modules' in addition to perceptual ones. While a discussion of the massive modularity thesis goes beyond the purpose of the present thesis, it is worth focusing on some of its implications: first, the introduction of a revised and looser notion of mental module, and, second, the hypothesis of a complex network of perceptual and conceptual modules "interconnected in ways that would make an engineer cringe" (Sperber, 1994a, p. 46).

The assumption that cognitive mechanisms, like biological mechanisms, are adaptively specialised for the solution of particular kinds of task is what grounds the conception of 'mental modules' as domain-specific and autonomous computational mechanisms. They are attuned to the regularities of their specific domain and employ dedicated procedures which are justified by those regularities. Significantly, they do not (necessarily) manifest all the properties that Fodor attributes to mental modules: their operations may not be mandatory (in the sense that an appropriate input may not be sufficient to trigger its own processing (Sperber, 2005)) and their informational encapsulation may be conceived of as a matter of degree:

> it may be that we have to rethink the concept of module and allow for a kind of continuum, from peripheral perceptual systems, which are rigidly encapsulated (not diverted from registering what is out there), through a hierarchy of conceptual modules, with the property of encapsulation diminishing progressively at each level as the interconnections among domain-specific processors increase. (Carston, 1997, p. 20)

This passage interestingly highlights the connection between this new notion of mental module and the hypothesis that modules are highly interconnected. The output of a perceptual or conceptual module can be fed to other conceptual modules, whose outputs can in turn be fed to further conceptual modules, and so on and so forth. The result consists of a chain of inferences that integrates the contribution of each individual module.

With this picture in mind, let us turn to Sperber and Wilson's (2002) proposal that utterance interpretation is carried out by a dedicated inferential mechanism or 'comprehension module'. Sperber and Wilson suggest that pragmatics is a sub-module of the general mind-reading module, which is responsible for providing explanations of individuals' behaviours in terms of attributed mental states (e.g. beliefs, intentions, etc.) [31]. Recognising the intention behind the speaker's communicative behaviour is a particular case of mind-reading (as Grice pointed out long ago). But, while utterances are a type of action and a speaker's meaning is a type of intention (i.e. a communicative intention, which is a second-order informative intention), according to Relevance Theory, the domain of overt communication exhibits such specific regularities and is so important in human life that, instead of employing general mind-reading procedures, it deploys its own dedicated comprehension procedure.

The Communicative Principle of Relevance expresses the kind of regularity that characterises the domain of overt communication and it is this which, according to Sperber and Wilson (2002), motivates the adoption of the relevance-guided

[31] For a detailed defence of this claim see Sperber and Wilson (2002) and Wilson (2005). According to this view, mind-reading is not a single, homogeneous system but a collection of autonomous dedicated mechanisms, or sub-modules (e.g. the Eye Direction Detector module).

comprehension procedure, (3), which is a dedicated inferential procedure (that works according to an in-built presumption of optimal relevance).

(3) **Relevance-guided comprehension procedure**
   a. Follow a path of least effort in computing cognitive effects: Test interpretative hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.
   b. Stop when your expectations of relevance are satisfied.

This procedure guides the construction of appropriate hypotheses about explicit content, intended contextual assumptions and intended contextual implications (implicatures) within the overall comprehension process. Note that the regularity expressed by the Communicative Principle of Relevance is specific to the domain of overt communication: in general, an observer is not entitled to expect that the intentional behaviours of others would have any particular level of relevance to him. But since this expectation is warranted in the case of overt communicative behaviours, the relevance-guided comprehension procedure tends to yield reliable conclusions.

5.6.2 The pragmatics systems and epistemic vigilance mechanisms: modules?

The massive modularity framework of the mind recognises that modules are highly interconnected with each other, that is, they form a network of systems and subsystems connected in such a way that they may take as input the outputs of several other modules. In what follows, I focus on the pragmatics module and its connection with epistemic vigilance module(s).

The pragmatics module takes as input an ostensive stimulus and delivers as output an interpretative hypothesis about the communicator's meaning. The metarepresentational output of the pragmatics module, i.e. 'The speaker$_x$ communicated that I' (where I is the set of propositions communicated), provides the input to two different kinds of epistemic vigilance mechanisms: mechanisms that focus on the source of information (*who* is to be believed) and mechanisms that focus on the informational content itself (*what* is to be believed).

From the perspective of the information flow through the architecture of the cognitive systems, the role of epistemic vigilance in the comprehension processes (see section 5.2) suggests that the epistemic vigilance module(s) does not receive its input only when the comprehension process is over. Rather, during the comprehension process, subparts of the interpretation are fed to the epistemic vigilance module(s) for its assessment. As a consequence, it may filter out interpretative hypotheses that are incompatible with the speaker's mental states (i.e. her beliefs and desires).

A similar account is proposed by Sperber and Wilson (1986/1995, pp. 186-187) with regard to the relationship between a language decoding module and a central inferential system. As suggested by Deirdre Wilson (personal communication), the relation between linguistic decoding and inferential comprehension and between comprehension and epistemic vigilance mechanisms could be framed in the same way. In general, epistemic vigilance mechanisms that monitor the speaker's competence and benevolence may restrict and direct the operations of the comprehension module.

As suggested in section 5.5.1, epistemic vigilance mechanisms may modulate the accessibility of the interpretative hypotheses under construction, and so affect 'hypothesis formation' as well as 'hypothesis confirmation'. This suggests, at least *prima facie*, an even more penetrative and intrusive role for epistemic vigilance in the processes of the pragmatics module than the one proposed by Sperber and Wilson (1986/1995) for pragmatics into the linguistic processing system. I believe, though, that this further role is easily accommodated within the model sketched above. It is worth noting that, as Wilson and Sperber (2012) write, pragmatic interpretation is "an inferential process which takes as input the production of an utterance by a speaker, *together with contextual information*" (Wilson & Sperber, 2012, p. x, *my emphasis* (DM)). Among these pieces of contextual information, there may well be information which has been accessed by epistemic vigilance mechanisms at an earlier stage. This information, which is already available at the beginning of the interpretative process, may be fed to the pragmatics module together with other relevant contextual information. When this is the case, it affects

the ‘path of least effort’ that the addressee follows in constructing interpretive hypotheses about the speaker’s meaning.

5.7 The cost of being vigilant

As for every cognitive mechanism, deployment of epistemic vigilance comes at a cost. In particular, it seems reasonable to assume that the kind of context-sensitive monitoring of the speaker’s competence and benevolence that plays a role in both comprehension and acceptance may require a great deal of processing effort. This opens up to the following question: are interpreters always willing and able to pay the price?

Sperber et al. (2010) briefly describe some of the factors that may modulate the activation of epistemic vigilance mechanisms. They confine their analysis to those mechanisms involved in assessing the believability of a piece of communicated information. Nevertheless, I try to apply such insights to the extended domain of epistemic vigilance proposed in this chapter.

As discussed in section 4.3.1, the first factor that is likely to affect the investment of energy required by epistemic vigilance is the *potential relevance* of a piece of communicated information. This hypothesis has received some support from Hasson, et al. (2005), who experimentally showed that increasing the relevance of a piece of communicated information modulates the importance of believability for the cognitive system. That is, epistemic vigilance mechanisms are likely to be less activated when the incoming information is not relevant to the hearer: the hearer would not invest extra energy in deciding whether or not to believe a piece of irrelevant information.

Sperber et al. (2010) mention a few other factors that may affect the activation of epistemic vigilance mechanisms. In discussing the parallel activation of the interpretative process, on the one hand, and epistemic vigilance, on the other, they suggest that “either process might abort for lack of adequate input, or because one process inhibits the other, or as result of distraction.” (Sperber et al., 2010, p. 364). This passage interestingly relates to the idea of there being competition between

cognitive mechanisms for the allocation of cognitive resources. This competition involves the comprehension module and epistemic vigilance mechanisms, but it is not limited to them:

> From a modularist point of view, attentional selection might be best seen, not as the output of a distinct attention mechanism allocating resources to specific modules, but as the result of a process of *competition for such resources among modules*. Some modules, for instance danger detectors, may be permanently advantaged in this competition because their inputs have a high expected relevance. Other modules may be advantaged at a given time because of a decision to attend to their potential inputs. For instance, face recognition is on the alert when waiting for a friend at the train station. Leaving aside these permanent bottom-up biases and temporary top-down biases, modules with the highest level of immediate activation both from upstream and downstream modules should be winners in the competition (with ongoing changes in these levels of activation resulting in shifts of attention). (Mercier & Sperber, 2009, pp. 151-152, *my emphasis* (DM))

With this picture in mind, we may speculate about the kind of circumstances in which epistemic vigilance mechanisms geared to assessing the speaker's reliability would not be favoured in the inter-modular competition. In such circumstances, they would thus fail to interact with the comprehension module and the interpreter would attribute an intended interpretation without taking into consideration information about the speaker's epistemic state.

Let us considered example (16), with the following important modification: Sarah runs into Robyn's office while she is having a meeting with a colleague and, without knocking at the door, she enters the room and excitedly utters:

(16) Neil has broken his leg.

As already discussed at length, epistemic vigilance mechanisms should prevent Robyn attributing to Sarah the first interpretation that comes to her mind, that is, that $Neil_1$ (i.e. her son) has broken his leg, and should allow her to recover the intended interpretation that $Neil_2$ (i.e. her colleague) has broken his leg. However, it is not implausible to imagine that, in such a circumstance, Robyn could be so alarmed as not to realise that Sarah could not have intended to refer to her son. She would take

Sarah to communicate that $Neil_1$ has broken his leg. Robyn might realise that that is the wrong interpretation afterwards (e.g. after running towards the corridor and finding $Neil_2$ lying on the floor with an injured leg). In this case, however, epistemic vigilance would not be responsible for triggering such recognition; rather, this would be due to the processing of some other (perceptual) information.[32]

How could Robyn's interpretative behaviour be explained? One possible explanation for this breakdown in communication is that the activation of 'danger detectors', and of those cognitive mechanisms that take as input the output of danger detectors, can overwhelm epistemic vigilance in virtue of their permanent advantage in the inter-modular competition. If this is the case, epistemic vigilance will fail to affect pragmatic interpretation because of the lack of cognitive resources available to complete its job.

This line of explanation is easily generalisable and provides an interesting working hypothesis to explain why interpreters can be blind to the speaker's mental states in some communicative settings: if epistemic vigilance cannot recruit enough cognitive resources to monitor the speaker's reliability, the interpreter will manifest an egocentric bias (to borrow the terminology of Keysar, Lin, and Barr (2003)).

[32] Considerations about the speaker's mental states (e.g. her beliefs) could play a subsequent role in reassessing the previously attributed interpretation. This role is different from the one that this chapter focuses on, that is, the role that they play through epistemic vigilance mechanisms in *on-line* pragmatic interpretation.

Chapter 6 Comprehension and acceptance in experimental pragmatics

The distinction between comprehension and acceptance (or belief) has been vigorously discussed by philosophers in the ordinary language tradition of Austin (1962), Grice (1957) and Strawson (1964b) and, more recently, by psychologists such as Daniel Gilbert. As discussed in Chapter 4, while comprehension involves the pragmatic ability to infer the speaker's meaning from linguistic and contextual cues (Sperber & Wilson, 1986/1995), acceptance involves what Sperber et al. (2010) call a capacity for 'epistemic vigilance', which enables hearers to avoid being accidentally or intentionally misinformed. While epistemic vigilance has become part of several research investigations, the comprehension/acceptance distinction has not been the focus of much attention with regard to pragmatic processing, with important methodological consequences, as I explain in this chapter.

6.1 Scalar inferences

The phenomenon of 'scalar inference' has been the subject of considerable debate within pragmatic theory and, more recently, has led to extensive empirical work in experimental pragmatics. Scalar inferences are pragmatic inferences like (27b), which seem to be drawn on a regular basis across different contexts.

(27) a. Some of the guests have arrived.
b. Not all of the guests have arrived.

While the encoded (semantic) meaning of the scalar expression 'some' is compatible with 'all', (27a) is often taken to suggest that *not all* the guests have arrived. In general, scalar inferences seem to arise when the speaker uses an expression which is not the strongest one on an informativeness scale, like the following: < some, many, most, all > (Horn, 1972, 1984).[33] The explanation goes as follows: while the encoded (semantic) meaning of 'some' is *at least one and possibly all*, the pragmatic

[33] In what follows I use the expression 'scalar inference' as opposed to 'scalar implicature'. This terminological choice is neutral as to whether the scalar inference is to be thought of as arising at the level of the explicit content of the utterance or as an implicature.

inference occurs because the speaker's use of a relatively weak expression is taken to indicate that the speaker believes the stronger option to be false. In other words, the addressee assumes that the speaker would have said 'all' if she thought the statement with 'all' was true; the choice of 'some' thus implies the negation of the stronger alternative.

Much of the debate has focused on the role of context in the derivation of scalar inferences. The question is whether they arise independently of context or as a function of contextual features which happen to be shared by many contexts and thus give rise to the impression that the scalar inference is regular. On the one hand, default accounts (e.g., Chierchia, 2004; Levinson, 2000) conceive of scalar inferences as default inferences, which are lexically associated but defeasible if contextually inappropriate. On the other hand, 'context-sensitive' accounts (e.g., Carston, 1990, 1998; Grice, 1975/1989; Sperber & Wilson, 1986/1995) suggest that scalar inferences are derived only when contextually appropriate. These competing accounts assign different roles to context: according to the former, context cancels inappropriate scalar inferences which arise independently of contextual support, whereas, according to the latter, it triggers appropriate scalar inferences which otherwise don't arise.

Default and context-sensitive accounts give rise to different psycholinguistic processing predictions with regard to 'lower bound' contexts. Lower bound contexts are those in which "all that is relevant, or can be known, is the lower bound" of scalar expressions (Horn, 1984, p. 13). For instance, consider the following example:

(28) A: Is there any evidence against them?
B: Some of their documents are forgeries.

(Levinson, 2000, p. 51)

In the context set up by A's question, all that it is relevant to know for the purpose of the discourse (i.e. answering to A's question) is that *at least some* of their documents are forgeries (that constitutes sufficient evidence against them). When contextual manipulations make the scalar inference irrelevant, the expected speed of interpretation of scalar terms (e.g. 'some') varies according to the two models. Default accounts predict that the time taken to derive the interpretation would

increase in virtue of the extra cognitive effort involved in cancelling the inappropriate scalar inference (which has arisen automatically). Context-sensitive accounts, on the other hand, predict the opposite: since the scalar inference (e.g. *Not all of their documents are forgeries*) is not relevant, it won't be derived; no additional processing cost is thus involved.

Much evidence supports the context-sensitive account. As has been reported and summarised elsewhere (e.g., Noveck & Reboul, 2008; Noveck & Sperber, 2007; Noveck & Spotorno, 2014), linguistically encoded readings are often sufficient for making on-line interpretations with utterances containing scalars and it is the more refined readings that come with extra processing costs (e.g., Bott, Bailey, & Grodner, 2012; Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; De Neys & Schaeken, 2007; Huang & Snedeker, 2009). Evidence comes from developmental, sentence processing, and neurological literature, all of which show how the linguistically encoded reading very often suffices for out-of-the-blue sentences and how extra processing is associated with the derivation of the scalar inference, in cases with supporting context. For example, using a self-paced reading task, which is exemplified in (29) and (30), the slashes indicating where participants manually advanced the text by pressing a key, Breheny et al. (2006) showed that when a scalar expression ("some of his relatives") occurs in an Upper-bound context (as in (29)), it takes longer to read because the context encourages a narrowed reading that excludes the more informative alternative ("some but not all of his relatives"), than in a Lower-bound context (as in (30), which encourages the lower-bound interpretation (*at least some*):

(29) ***Upper-bound context:*** Mary/asked John/whether he intended to host/all his relatives/in his tiny apartment./John replied/that he intended to host/**some of his relatives**./**The rest**/would stay/in a nearby hotel.
(30) ***Lower-bound context:*** Mary was surprised/to see John/cleaning his apartment/and she asked/the reason why./John told her/that he intended to host/**some of his relatives**./**The rest**/would stay/in a nearby hotel.

Conversely, reading times for a subsequent anaphoric expression ("The rest") were shorter in the Upper-bound context than in the Lower-bound context. That is, the interpretation of "The rest" was facilitated when the scalar inference had already

been made before encountering the target phrase (see also Katsos, Breheny and Williams, (2005); Katsos (2008)).

6.2 Scalar inferences and face-threatening contexts

In a very interesting further development, Bonnefon, Feeney and Villejoubert (2009) and Feeney and Bonnefon (2012) have suggested that other contextual manipulations, beyond those explored by Breheny et al. (2006)), may prove to be equally fruitful for testing default and context-sensitive accounts of scalar inference. In particular, they propose that, along with lower-bound contexts, *face-threatening* contexts may make the scalar inference contextually inappropriate. Face-threatening contexts are contexts in which the face of the addressee may be damaged, where an individual's 'face' corresponds to his/her public self-esteem, a property/quality which is cultivated (and defended) in social interactions (Brown & Levinson, 1987). For instance, in a face-threatening context like (31), the addressee might take the use of 'some' to be a polite device adopted by the speaker in order not to hurt his feelings. If this is the case, Bonnefon et al. (2009) suggest, A won't interpret B's utterance as conveying that some *but not all* the guests at the dinner thought that he drank too much.

(31) A: What impression did I make during dinner?
B: Some thought that you drank too much.

(Bonnefon et al., 2009, p. 250)

Indeed, they even suggest that A might interpret B's utterance as communicating that *all* the guests thought that he drank too much.

The hypothesis put forth by Bonnefon et al. (2009) is that an utterance of the form 'Some *X*-ed' (e.g. "Some thought that you drank too much") is less likely to be interpreted as communicating the scalar inference *Not all X-ed* when *X*-ing is something that threatens the face of the listener (e.g. thinking that the listener drank too much). In these circumstances, the speaker's face-saving concerns might motivate her to use the term 'some' even if she is in an epistemic position to use the term 'all'. As a consequence, the hearer might refrain from concluding that the

speaker does not believe the stronger (and more face-threatening) statement to be true.

In order to test this hypothesis, Bonnefon et al. (2009) ran a series of three off-line studies, in which behavioural responses were compared across face-threatening and face-boosting contexts. In the first experiment, participants were presented with two stories, one in a face-boost condition, and the other in a face-threat condition. The following represents a sample story:

(32) Imagine that you have joined a poetry club, which consists of five members in addition to you. Each week one member writes a poem, and the five other members discuss the poem in the absence of its author. This week, it is your turn to write a poem and to let the others discuss it. After the discussion, one fellow member confides to you that "Some people **hated/loved** your poem".
(Bonnefon et al., 2009, p. 251)

After reading each story, participants were asked to answer the following Yes/No question: 'From what this fellow member told you, do you think it is possible that everyone hated/loved your poem?' The percentage of negative answers was compared across the two conditions.

The results show that 42% of participants answered 'Yes' when asked whether it was possible that everyone hated their poem when told that 'some' did (as opposed to 17% in the face-boost condition). In other words, participants were more likely to think that everyone hated their poem when told that some did, than to think that everyone loved their poem when told that some did.

In the second experiment, a variant of the first one, participants were required to rate, on a 10-point scale anchored at *totally unlikely* and *totally likely*, how likely it was that the speaker would have used the word 'some' if she knew that the number of people who hated *x*/loved *x* was in fact 1. The question was repeated for all numbers up to 6 (where 6 corresponds to *all*). The membership function of the number 6 (i.e. to what extent does *all* enter in the concept of 'some'?) was significantly higher in the face-threat condition (5.6/10) than in the face-boost condition (4.1/10). This suggested that the quantity 6 (i.e. all) in the face-threat condition was considered as

more representative of the concept denoted by 'some' than the quantity 6 (i.e. all) in the face-boost condition.[34]

Overall, the results of these two experiments were taken to show that when *X* threatens the face of the listener, "then 'some *X*-ed' is less likely to be interpreted as implying 'not all *X*-ed'" (Bonnefon et al., 2009, p. 254). Bonnefon et al. (2009) suggest that in face-threatening contexts the scalar inference from 'some…' to 'not all…' is less likely to be derived.

Finally, in the third experiment, participants were presented with four stories (two in the face-boost condition, two in the face-threat condition), where the epistemic state of the speaker was explicitly manipulated. In one version, the speaker was said to know that the proportion of people who *X*-ed was 3/6, in the second, 6/6. The following is a sample story for the 6/6 version:

(33) Imagine you gave a speech at a small political rally. You are discussing your speech with Denise, who was in the audience. There were six other people in the audience. You are considering whether to give the same speech to another audience. Hearing this, Denise tells you "Some people hated/loved your speech". Denise knows that 6 people out of 6 loved/hated your speech.
(Bonnefon et al., 2009, p. 254)

After reading each story, participants were asked to rate, on four different 10-point scales, how *accurate*, *considerate*, *honest* and *nice* it was of the speaker to use the word *some* in that context. The ratings revealed that the use of *some* (e.g. "Some people hated your speech") by a speaker who knows that everyone *X*-ed is considered inaccurate and dishonest in both face-threatening and face-boosting contexts (although the effect was slightly stronger in the former). However, in these circumstances, the use of 'some' was considered nice and considerate only if the context was face-threatening.

---

[34] In virtue of the addition of a Target Group condition, Experiment 2 shows that the effect of a face-threatening context is detectable only when the target of the face-threat is the listener (as opposed to a third party, one that neither the speaker nor the listener care about.) The two groups were inserted into the design of the experiment to control for the possibility that face-threatening contexts might simply correspond to lower-bound contexts (favouring 'at least some' interpretations), and face-boosting contexts to upper-bound ones (favouring 'some but not all' interpretations). The interaction effect ruled this possibility out.

Overall, the results of the three experiments were interpreted as follows:

> When *X* in 'some *X*-ed' threatens the face of the listener, individuals are less likely to infer that the speaker *meant* or knew that not all *X*-ed – and this is because they consider the possibility that the speaker might want to be nice more than to be precise. (Bonnefon et al., 2009, p. 255, *my emphasis* (DM))

Interestingly, a similar pattern of results was found by Feeney and Bonnefon (2012) with regard to the scalar connective 'or'. Adopting an experimental design similar to the one in Bonnefon et al. (2009), Feeney and Bonnefon (2012) concluded that the scalar connective 'or' is less likely to give rise to an exclusive interpretation (i.e. 'A or B but not both') in a face-threatening context ("You'll take a pay cut or your vacation period will be reduced") than in a face-boosting context ("You'll get a pay rise or your vacation period will be increased"). They interpret these results as strengthening the conclusion of Bonnefon et al. (2009). Given their similar experimental design, the methodological considerations presented in the next section are intended to apply to both studies.

6.3 Comprehension and acceptance: methodological implications

The primary aim of this section is to show the methodological importance of the theoretical distinction between comprehension and acceptance, which has been arguably neglected in the experimental literature on scalar inferences. Because comprehension does not require acceptance of the communicated content, it is important to distinguish between, on the one hand, what the addressee takes to be the communicated content of the utterance (that is, what the speaker meant) and, on the other hand, what the addressee accepts as true. These may differ from (and even be incompatible with) each other in significant ways. In what follows, I discuss the implications of taking the distinction between comprehension and acceptance into consideration for the current literature on politeness and scalar inferences. I suggest that there are two alternative lines of explanation that can account for the available empirical evidence: (i) face-threatening contexts block the derivation of scalar inferences (as suggested by Bonnefon and colleagues), or (ii) in face-threatening

contexts the scalar inference is in fact derived but is less likely to be accepted (as true).

6.3.1 Assessing the believability of scalar inferences

In this section, with the help of some examples, I illustrate the relationship between comprehension and acceptance, as well as the distinction between inferences that are part of the communicated content of the utterance and inferences that are not.

To begin with, let us consider the following example from Noveck and Sperber (2007). Jane and Henry are hosting a dinner party. The doorbell rings and Jane shouts at him from the living room:

(27) a. Some of the guests have arrived.

In what follows I elaborate on three possible scenarios. In the first scenario the communicated content of the utterance is accepted as true, while in the second and third ones it is rejected (on the basis of consideration of, respectively, the speaker's abilities and the speaker's preferences). I show that in each case what the addressee infers about the actual state of the world does not coincide (at least not entirely) with the communicated content of the utterance.

While Henry is cooking in the kitchen, Jane utters (27a). Let us assume that what makes Jane's utterance relevant is that it implies that Henry should go out and get the dessert. This is something they previously agreed on, so Henry is likely to interpret (27a) as communicating that some of the guests have arrived (where the meaning of 'some' is constructed as having a vague cardinality compatible with any number of guests) and (weakly) implicating that Henry should go out and get the dessert.[35] Henry has no reason to doubt Jane's trustworthiness and he believes what she communicates to him. Assume, though, that Henry has also planned to buy a bottle

[35] In fact, Noveck and Sperber (2007) assume that this scalar inference arises at the level of the explicit content of the utterance, as modulation of the meaning of 'some' (rather than as an implicature). However, this does not affect the point I am making here. I will discuss this point and other details of the relevance-theoretic account of scalar inferences in section 6.4.

of champagne at the local store on the way to the bakery as a surprise for Jane and all the other guests. Jane's utterance, (27a), will thus license the further inference that he should go and buy the champagne before driving to the bakery. Of course, Henry would not take this inference to be intended (Jane is not even aware of the surprise he is organising). It should be clear, then, that the inferences licensed by Jane's utterance go beyond those that Jane may intend to communicate. More generally, the addressee may form a wide range of beliefs about the world through an act of communication, only some of which the speaker may have intended him to form.

Now, consider a different scenario. While Henry is in the kitchen, the doorbell rings. Jane opens the door and utters (27a) to him. Since they agreed that he should throw the pasta in the boiling water only when all the guests have arrived, he would take Jane's utterance to implicate that not all the guests have arrived (as well as a range of weakly communicated implications about what he should do to greet the guests, e.g. bringing the finger food and/or some drinks). Imagine, though, that Henry can see from the kitchen's windows some cars parking outside the house. He knows that Jane could not see them from the front door. So, while he understands what she communicates, he has reason to doubt her epistemic trustworthiness. As a consequence, he does not accept that not all the guests have arrived. On the contrary, he has evidence to think that it is likely to be the case that all the guests have arrived.

Finally, consider a different scenario. Henry and Jane have had a fight earlier that day and are not on good terms. After welcoming all the guests, Jane utters (27a) to him. She wants to trick Henry into thinking that not all the guests have arrived, so he will feel under pressure in the kitchen when he later realises that everyone is waiting for the main course which is not ready yet. Suppose that Henry suspects that Jane is being mean and uncooperative towards him. As a consequence, he does not accept her communicated implication *Not all the guest have arrived*. On the contrary, because of his suspicion, he may end up thinking that all the guests have actually arrived.

The last two scenarios suggests that not only may the addressee have reasons to doubt the truth of the communicated content (and, as a consequence, fail to accept it as true), but that he may also infer that a state of affairs, which is incompatible with

the one described by the communicated content, is likely to hold, on the basis of those very same reasons.

6.3.2 A methodological note

Let us look now at the implications that this distinction between what is communicated by the speaker and what is accepted/believed by the hearer has for the experimental studies carried out by Bonnefon et al. (2009). I suggest that Experiment 1 and Experiment 2 provide evidence only in support of the claim that, in face-threatening contexts, the addressee is more likely to believe that the possibility that all *X*-ed holds when told that 'Some *X*-ed'. In Experiment 1, participants were more likely to answer "Yes" when asked if it was possible that everyone *X*-ed in a face-threatening context (e.g. "Some people hated your poem") than in a face-boosting context (e.g. "Some people loved your poem"). Experiment 2 provided the same kind of evidence, suggesting that participants judged it as more likely that the speaker could have used the word 'some' while knowing that *all* in a face-threatening context than in a face boosting-context. The results of these experiments clearly indicate that addressees are more reluctant to rule out the possibility that *all X-ed*, when told that 'Some *X*-ed' in a face threatening context.

This, however, is not enough to support the stronger claim that they do so because, in face-threatening contexts, speakers do not communicatively intend and addressees do not derive the scalar inference *Not all X-ed*. The same pattern of results could be explained by assuming that while addressees do, in fact, derive the scalar inference *Not all X-ed* as part of the (face-saving) communicated content, they do not accept it as true. The addressee may have reasons to doubt the truth of what the speaker communicates, for example because he thinks that she is trying to be nice and polite rather than strictly honest (as Experiment 3 suggests). If this is the case, he may fail to accept what she communicates (e.g. *Not all X-ed*), and even infer that an alternative state of affairs (e.g. *All X-ed*) is likely to hold on the basis of the same considerations. Interestingly, this explanation receives some support from the results of Experiment 3. These results reveal that, in the face-threatening contexts at issue, the addressee thinks that the speaker is trying to be nice and polite rather than strictly honest. The addressee may reason that the communicator does not want to hurt his

feelings by telling him the whole truth and consequently fail to accept what she communicates as true.

The same point applies to the results of Feeney and Bonnefon (2012). Their participants were presented with a scenario containing an utterance of the form 'A or B'. They were subsequently asked whether, in their opinion, the scenario ruled out the possibility that A *and* B. The following is a sample scenario for the face-threatening condition (in which A and B are both undesirable events):

(34) Imagine that Clare is a children's author. Over lunch her publisher tells her that the sales of her last book have been so poor that she will receive decreased royalties or she will be denied an upfront payment for her next book. In your opinion, does this rule out the possibility that Clare will receive decreased royalties and be denied an upfront payment for her next book?

(Feeney & Bonnefon, 2012, p. 185)

The 10-point scale they were asked to use for their ratings was anchored at *Does not rule out the possibility at all* (lower end) and *Completely rules out the possibility* (higher end). Low ratings can be explained either by assuming that participants adopt an inclusive interpretation of the connective 'or' (i.e. they do not derive the scalar inference *A or B but not both*) or by assuming that they derive the scalar inference but reject it as unlikely to be true. In other words, lower ratings may correspond to a lower acceptance rate of the derived scalar inference rather than to the absence of the inference from what is taken to have been communicated by the kind/polite speaker.

It may be asked why a speaker who is concerned to avoid complete factual accuracy in order to save the hearer's face, e.g. by implicitly communicating that *Not all X-ed* when in fact all *X*-ed should not straightforwardly utter 'Not all *X*-ed'. After all, if the speaker were willing to sacrifice honesty for kindness, why would she avoid saying explicitly to the addressee "Not all the people hated your poem"? I believe that the choice of uttering 'Some X-ed' in these circumstances can be explained by appealing to the notion of 'deniability' (Lee & Pinker, 2010). In a context in which the speaker knows that all of the people hated the addressee's poem, she could implicitly communicate that not all of them did by uttering "Some people hated your poem", while at the same time keeping open the possibility of denying that she intended to communicate such an implicature. If openly challenged, the speaker

could deny that she meant that content, e.g. by suggesting that the semantic (encoded) meaning of 'some' (which is compatible with 'all') is all that she wanted to communicate. Crucially, speaker denial would require the speaker to pretend that a different conversational context is in place, namely, one in which the assumption that the stronger alternative 'All *X*-ed' would be more relevant to the addressee is not mutually manifest.

More generally, the search for deniability is often motivated by a conflict of interests/goals on the speaker's side. For instance, in the contexts at issue, while the speaker may want to spare the addressee's feelings and be polite, she may not want to run the risk of being accused of lying. Implicating is indeed a very important way of merely misleading rather than lying (e.g., Saul, 2012). If openly challenged, the speaker can object that what she said was true.[36]

In conclusion, I suggest that the design of Bonnefon et al. (2009) and Feeney and Bonnefon's (2012) studies misses the important theoretical distinction between what the addressee takes to be the communicated content of the utterance (i.e. the speaker's intended meaning), on the one hand, and what he infers about the reality (the state of affairs in the world), on the other hand. As a consequence, their results are compatible with two alternative lines of explanation. According to the first, put forth by the authors themselves, they are explained by suggesting that addressees are less likely to derive scalar inferences in face-threatening contexts. According to the second, advanced in this chapter, they can be accounted for by assuming that addressees are less likely to accept certain (face-saving) pieces of communicated information as true in a face-threatening context. The crucial distinction between the two is that, while the former assumes that the speaker of a face-threatening utterance of the form 'Some X-ed' is less likely to communicatively intend the scalar inference *Not all X-ed*, the latter maintains that the speaker does intend to communicate this inference but that the addressee may not believe it. It should be clear that these

[36] For more on the notion of 'saying' which is relevant for the lying-misleading distinction, see Saul (2012).

alternative explanations make incompatible claims with regard to the *pragmatic interpretation* of scalar utterances in face-threatening contexts.

Interestingly, the line of explanation explored here opens up a new way of thinking about certain cases of understatement.[37] Understatement (or meiosis) is a rhetorical trope which is typically analysed as the opposite of hyperbole. We think of understatement as a way to communicate something stronger in an indirect way (e.g., of a man known to have broken up all the furniture, one says: "He was a little intoxicated", Grice, 1975/1989, p. 34). However, let us focus on the subset of 'polite understatements', that is, understatements which might serve a face-saving function. Consider the following example.

(35) Not everybody thought that you were sober.

The usual view would be that, by uttering (35), the speaker is implicating the stronger proposition that *Nobody thought you were sober*. In line with the analysis of face-threatening scalar utterances proposed above, there is an alternative account of examples like this. While the semantic (encoded) meaning of 'not everybody' is compatible with 'nobody', it is possible to argue that (35) has a (non-prototypical) scalar implicature *It is not the case that nobody thought that you were sober*, which is truth-conditionally equivalent to *Some thought that you were sober*. That is, polite understatements do have a 'polite' implicature, but this may not be accepted as true by the addressee. Specifically, the 'polite' implicature would not be taken to be sincere and, as a consequence, the addressee would not accept its content as true and would further infer that it is likely to be the case that *nobody* thought that he was sober. Interestingly, this line of reasoning might be anticipated, even intended, by the speaker, without the inference to *Nobody thought that you were sober* falling under her communicative intention (in the sense of 'ostensive-inferential' communication proposed by Sperber and Wilson 1986/1995).

[37] Thanks to Dan Sperber for pointing this out to me.

6.3.3 "When some is actually all"

Bonnefon et al.'s (2009) face-management analysis of scalar utterances goes beyond the claim that face-threatening contexts may block the derivation of the scalar inference. They suggest that addressees may interpret a face-threatening utterance like "Some people hated your poem" as *communicating* that *all* people hated his poem. According to this suggestion, when the speaker utters "Some people hated your poem", the addressee may take the proposition *All the people hated your poem* to be communicatively intended. Bonnefon and colleagues take it as supporting evidence for this analysis that participants (in experiments 1 and 2) are more reluctant to rule out the possibility that *All X-ed* when told that 'Some *X*-ed' in face-threatening contexts than in face-boosting contexts. In the previous section, I advocated a more cautious analysis of Bonnefon et al.'s results. I questioned whether they provide any conclusive evidence as far as the *interpretation* of the speaker's utterance is concerned. In line with this, I believe that they do not provide any evidence that the addressee takes the proposition *All X-ed* to be *communicatively intended* (i.e. meant) by the speaker. In what follows, I present some arguments that support my rejection of this analysis.

The scalar inference from 'Some *X*-ed' to *Not all X-ed* can be seen as a special case of a more general phenomenon, that of 'quantity inferences' or 'quantity implicatures' (e.g., Breheny et al., 2013; Geurts, 2011). For example, in a context in which you know that a woman has two sets of objects, forks and spoons, and you know that she is placing them in either one of two boxes, A and B, if a third party tells you "The woman put a spoon in box A and a spoon and a fork in box B", you are likely to infer that the speaker wanted to communicate that the woman put a spoon *and nothing else* in box A (Breheny et al., 2013, p. 425). The inference 'and nothing else' closely resembles the lexically triggered inference *Not all X-ed* from 'Some *X*-ed' and seems to be derived via the same pattern of reasoning. Now, imagine the following face-threatening context. A and B are working in the same company. During the last year A's sales volume has considerably decreased. The director has informed all the employees that two kinds of cuts will be considered in cases of poor sales performance: retirement benefit and holiday allowance. They are all worried about this. A is speaking with B, the director's secretary.

(36) A: What did the director decide to do?
B: She will cut your holiday allowance.

Assuming that scalar inferences are a particular case of quantity inferences, Bonnefon et al.'s (2009) analysis of scalar inferences in face-threatening contexts should generalise to (36). The quantity inference *The director will cut your holiday allowance and nothing else* should be less likely to be derived because of the face-threat towards the addressee. Furthermore, B's utterance should be possibly interpreted as communicating the more informative proposition that *The director will cut your holiday allowance and your retirement benefit*. This seems to be an implausible interpretation of B's utterance, but it follows from the face-management analysis provided by Bonnefon and colleagues with regard to scalar inferences. So, it is either that such an analysis generates some counterintuitive predictions (when extended to closely related quantity implicature cases) or that it misses an important generalisation (since the inferential pattern across the cases seems to be the same).

Furthermore, the claim that in face-threatening contexts 'Some X-ed' is interpreted as communicating *All X-ed* is at odds with the empirical findings of Experiment 3. The results of the study show that participants judged the use of 'some', when the addressee knows that *all*, as dishonest and inaccurate both in the face-threatening and in the face-boosting conditions (with the effect being slightly stronger in the face-threatening condition). However, the use of 'some' in the all-condition was considered nice and considerate only in the face-threatening condition. The point worth noting here is that low honesty ratings in the face-threatening condition are arguably incompatible with the claim under discussion. The reason is that if the addressee attributes to the speaker the intention to communicate *All X-ed* in a situation in which it is indeed the case that all X-ed, we wouldn't expect him to judge the communicative act as dishonest. We would expect the addressee to judge the scalar utterance as honest in the face-threatening context but not in the face-boosting one.

To sum up, the claim that 'Some X-ed' is interpreted as *All X-ed* when the context is face-threatening faces two independent challenges: first, it fails to fit within a unified account of quantity implicatures; second, it is not supported by the empirical

evidence just surveyed.

6.3.4 Scalar inferences and honesty

As well as exploring the effect of politeness on the derivation of scalar inferences, Feeney and Bonnefon (2012) investigate the role of (self-rated) honesty. Their hypothesis is that more honest people may be more likely to derive scalar inferences (e.g. to interpret the connective *or* in 'A or B' as *A or B but not both*). As previously discussed, there are reasons to doubt that the experiments at issue address the *interpretation* of scalar utterances (rather than the believability of the inference they may warrant). I present Feeney and Bonnefon's analysis of the results and suggest an alternative explanation of the data.

Let us first investigate the theoretical motivations behind the hypothesis that honesty and scalar inferences positively correlate. They revolve on two claims: (i) because of 'social projection' (Robbins & Krueger, 2005), people who consider themselves honest are more likely to expect other speakers to manifest the same honesty,[38] (ii) scalar inferences are facilitated by an assumption of honesty. Feeney and Bonnefon suggest that honest speakers are more cooperative speakers, that is, they try to be as informative as possible given their epistemic state. When using an expression that is relatively low on a salient scale, such as 'or', they can be expected not to be in the epistemic position to make a stronger claim (e.g. 'A and B'). Thus, hearers can confidently derive the scalar inference *A or B but not both*. It follows that since more honest people may tend to attribute a higher degree of honesty to their interlocutors (see (i)), they may be more likely to derive scalar inferences when their interlocutors use a scalar expression.

The experiments seem to support this hypothesis. Regardless of context (be it face-threatening or face-boosting), more honest people are more likely to rule out the possibility that A and B, given an utterance of 'A or B'. In other terms, while both Face and Honesty had a main effect, no interaction was detected. In what follows, I put forth an alternative explanation of the main effect of Honesty. Once again, the

[38] In what follows, I assume the validity of (i). For more details, see Robbins and Krueger (2005)).

principal aim of this suggestion is to show that the experimental evidence provided does not warrant Feeney and Bonnefon's explanation over this alternative.

My argument focuses on the relationship between honesty and trust (Sperber et al., 2010). Honest people are usually trustworthy: they are willing to share information that they possess and regard as true. This is the reason why they are more likely to be granted epistemic trust. Epistemic trust can be defined as the willingness to believe the communicator and accept her claims as true. It follows that hearers are more likely to accept a piece of communicated information when this comes from an interlocutor they judge to be honest.

As suggested in section 6.3.2, the experimental paradigm adopted by Bonnefon et al. (2009) and Feeney and Bonnefon (2012) does not allow us to draw firm conclusions about the selected *interpretation* of scalar expressions (in face-threatening and face-boosting utterances). Given a sentence of the form 'A or B', participants are asked whether the possibility that A *and* B is ruled out. While the authors suggest that lower ratings on the scale '*Does not rule out the possibility at all* (lower end) - *Completely rules out the possibility* (higher end)' correspond to an inclusive interpretation (i.e. to the absence of the scalar inference), I suggested that they may depend on a lower acceptance rate of the scalar inference (i.e. the exclusive interpretation). The scalar inference would be derived and attributed to the speaker (as part of her intended meaning) but not accepted as true.

If this is the case, however, the correlation between honesty and higher ratings comes as no surprise. If honest people are more likely to accept a piece of communicated information because they are more likely to grant epistemic trust to their interlocutor (because of social projection), they will tend to rule out the possibility that A and B when the speaker has uttered 'A or B'. This would not depend on the fact that they are more likely to derive the scalar inference than dishonest people, but rather that they are more likely to accept it as true when it is communicated.

This explanation is compatible with the pattern of results and with the absence of interaction between Face and Honesty. Honest people are generally more likely to accept a piece of communicated information (regardless of context) but they are not

"any less sensitive to the demands of politeness" (Feeney & Bonnefon, 2012, p. 187) than less honest individuals.

6.4 Relevance Theory and scalar inferences

In their general discussion, Bonnefon et al. (2009) examine whether Relevance Theory is compatible with their face-management analysis of the empirical data. They claim that "the only way […] for Relevance Theory to account for our results is to make an ad hoc assumption about processing effort" (Bonnefon et al., 2009, p. 257). In order to explain their results, Bonnefon and colleagues suggest, Relevance Theory would need to assume that interpreting an utterance of 'Some *X*-ed' as communicating that *All X-ed* requires less processing effort in a face-threatening context than in other contexts. Specifically, in a face-threatening context, the interpretation *All X-ed* would be less costly than the interpretation *Some (but not all) X-ed*, and, because of this, it would be the first interpretation to satisfy the hearer's expectations of optimal relevance and so be attributed to the speaker.

I have already suggested reasons to doubt that hearers interpret scalar utterances in face-threatening contexts as communicating *All X-ed*. In what follows, I argue that Relevance Theory would not predict such an interpretation but rather the alternative analysis of Bonnefon and colleagues' data which I am advocating in this chapter. Before doing this, however, I present the relevance-theoretic account of scalar inferences and discuss the contextual premises that are taken to licence the derivation of scalar implicatures.

6.4.1 Lexical modulation and scalar expressions

Relevance Theory has long offered a 'deflationary' account of scalar inferences (Carston, 1998; Noveck & Sperber, 2007; Sperber & Wilson, 1986/1995). That is, scalar inferences do not comprise a 'natural kind': not only are they not confined to utterances containing expressions on fixed entailment scales like 'some'/'all' (see Hirschberg (1985) for an account of ad hoc scales), but they are not even inferences of the same type. In particular, they might or might not be part of the communicated content of the utterance, and, when they are, they might arise either at the level of

what is explicitly communicated, or as an implicature (i.e. part of the implicit content of the utterance).

According to Relevance Theory, most of the examples of 'scalar inferences' discussed in the literature arise at the level of the explicit content of the utterance (i.e. its explicatures, in relevance-theoretic terms). Specifically, they are conceived of as the result of 'lexical modulation', a pragmatic process "which applies spontaneously, automatically and unconsciously to fine-tune the interpretation of virtually every word" (Wilson, 2004, p. 343). As a result, the concept communicated by the use of a certain word, which emerges as a by-product of the search for relevance, can be significantly different from its lexically encoded meaning. For instance, consider the following example from Wilson and Carston (2007, p. 232):

(37) I'm not *drinking* tonight.

Depending on the circumstances, a speaker could utter (37) to communicate that she is not going to drink (a) any liquid, (b) any alcohol, or (c) any significant amount of alcohol, etc. It follows that the interpretation of (37) might be more or less narrowed, depending on the contextually manipulated denotation of the lexically encoded concept DRINK.

Noveck and Sperber (2007) suggest that the same kind of lexical modulation affects many instances of so-called 'scalar expressions', such as 'some'. The lexically encoded concept SOME does not include an upper bound (i.e. it denotes any subset containing *n* elements, where *n* is at least 2 and at most the total number of elements).[39] However, as for DRINK in (37), it can be used to communicate a contextually modulated, more specific concept SOME*, whose denotation differs from that of SOME. For instance, it can be used to convey an upper-bounded interpretation of 'some' that is, AT LEAST TWO AND FEWER THAN ALL. While this is certainly a common narrowing down of the encoded meaning of 'some', it is not the only lexical modulation available and it does not carry any special status. Following Noveck and

[39] Other authors, including Feeney and Bonnefon (2012), suggested that the semantic meaning of 'some' corresponds to AT LEAST 1 AND POSSIBLY ALL.

Sperber (2007, p. 192), consider again the following scenario. Jane and Henry are hosting a dinner party. They have agreed that Henry will go out and get the dessert from the pastry shop as soon as the guests start arriving. The doorbell rings and Jane shouts at him from the living room:

(27) a. Some of the guests have arrived.

Henry does not know how many guests have arrived and whether Jane has already opened the door. Significantly, what makes Jane's utterance relevant is that it implies that Henry should go and get the dessert. This implication is warranted by any construal of 'some'. Thus, Henry's interpretation of (27a) is likely to be compatible with any number of guests, even just one. In this context, the lexically encoded concept SOME works as a clue to derive an ad hoc concept SOME**, whose denotation is arguably broader than that of the encoded SOME.

In general, Relevance Theory suggests that the encoded meaning serves as a clue to indicate the speaker's meaning, which is inferred on the basis of the occasion-specific expectations of relevance that guide the comprehension process (e.g. the expectation that Jane's utterance will inform Henry about whether he needs to go and get the pastry), underpinned by a general presumption of optimal relevance:

(2) **Presumption of optimal relevance**
   a. The ostensive stimulus is relevant enough to be worth the audience's processing effort.
   b. The ostensive stimulus is the most relevant one compatible with the communicator's abilities and preferences.
   (Sperber & Wilson, 1995, p. 270; Wilson & Sperber, 2004, p. 612)

Word meanings are modulated in context in order to yield implications which satisfy the addressee's expectations of relevance. The contextual modulation of the meaning of 'some', whose denotation can be either broadened or narrowed down (or both), affects the explicit content of the utterance (i.e. its explicatures). In all these cases, an ad hoc concept SOME* is inferred and attributed to the speaker as part of what she intends to communicate explicitly.

It is worth noticing, though, that lexical modulation does not exhaust the phenomenon of 'scalar inferences'. In some, limited, cases, the 'scalar inference' occurs both at the explicit and at the implicit level of communication:

> [..] the class of cases described in the literature as scalar inferences is characterized by an enrichment at the level of the explicature (where, for instance, *some* is interpreted in a way that excludes *all*) and only in a small class of these is the exclusion of the more informative concept not just entailed but also implicated. (Noveck & Sperber, 2007, p. 193)

This means that while the narrowing down of the meaning of 'some' to SOME* is pretty common, the implication that *[For all the speaker knows] not all X-ed* is not always communicated by the speaker (even if it can be derived by the hearer). The crucial distinction to draw here is the one between implication and implicature, which I illustrate with the help of an example ((Noveck & Sperber, 2007, p. 192). Consider again (27a), uttered in the following context. While Henry and Jane are waiting for their guests to arrive, Henry is cooking alone. Jane comes in and utters (27a).

(27) a. Some of the guests have arrived.
b. Not all of the guests have arrived.

Henry may consider that he should go and greet the guests and bring them some drinks. The fact that Jane's utterance satisfies the hearer's expectations of relevance without bringing to mind consequences more typical of the arrival of *all* the guests (e.g. that he should add the pasta to the boiling water), allows Henry to construct the meaning of 'some' with a vague cardinality above one and below all. In this context, the explicit content of (27a) implies (in fact entails) (27b) but this may not be actively entertained by Henry (let alone taken to be part of the communicated content of the utterance). Suppose, now, that Henry is wondering whether all the guests have arrived. Given Jane's utterance, (27a), he may well infer that not all the guests have arrived, i.e. (27b). However, it is only in a context in which there is a mutually manifest reason for the interlocutors to wonder about whether all the guests have arrived, that the denial of this stronger alternative will be implicated. For instance, if Jane has uttered (27a) in response to the question "Have all the guests arrived?",

Henry will take (27a) to implicate that *[For all that Jane knows] Not all of the guests have arrived*.

As the examples above suggest, there is a distinction to be drawn among 'entailment', 'inferred entailment', and 'communicated entailment' (or 'implicated entailment'). While the first notion pertains to a semantic theory, the others arise within a cognitive/pragmatic theory (Carston, 2002, pp. 112-113). Crucially, an entailment whose propositional content is actually entertained (i.e. inferred) by the addressee is not always communicated as an implicature. This occurs only when the utterance achieves relevance by providing an answer to an explicit or tacit question as to whether the stronger alternative (e.g. all the guests have arrived) is true. According to Relevance Theory, scalar inferences are "not scalar, [and] not necessarily implicatures" (Noveck & Sperber, 2007, p. 193).

6.4.2 'don't know' vs. 'don't want to say' implicatures

The derivation of genuine scalar implicatures occurs only if the context makes the stronger alternative a relevant possibility. This is in line with the claim that particularised approaches conceive of scalar inferences as context-sensitive. A scalar implicature is derived only if it is contextually relevant to the addressee, that is, only if the context is an upper-bound context. However, according to Relevance Theory, the irrelevance of the stronger alternative is not the only contextual factor that may prevent the derivation of scalar implicatures. Considerations about the speaker's 'abilities' (e.g. what she does or does not know) and 'preferences' (e.g. what information she is willing to communicate) may equally block the derivation of scalar implicatures.

The role played by considerations of the speaker's epistemic state (i.e. her 'abilities') has been widely recognised beyond Relevance Theory, being referred to in the literature as the 'epistemic-step' in the derivation of scalar implicatures (e.g., Chierchia, Fox, & Spector, 2012; Sauerland, 2004). Consider the following example:

(38) a. *Teacher*: Some of the students did well on the test.
b. [The speaker believes that] Not all of the students did well on the test.

In order to derive (38b) from (38a), the addressee has to assume that the teacher has an opinion about whether or not all of the students did well on the exam. In other terms, he has to assume that the teacher believes either that all the students did well on the exam or that not all of them did. This should be clear from the rational reconstruction of the derivation of (38b) offered by Breheny et al. (2013, p. 424):

I. The teacher has said that some of the students did well on the test. For all that is said, it could be true that all of the students did well.

II. However, given that (i) the utterance is telling us about how the students did on the test and (ii) the mutually assumed expectation is that the teacher will give as much information as is relevant modulo her own knowledge and preferences,

III. it would clearly be deficient of the speaker to have said what she did if she had known that all of the students did well.

IV. So we can conclude that the teacher does not know that all of the students did well.

V. Given that the speaker knows all about how the students did, we can conclude that not all of the students did well.

VI. The speaker intends me to reason as above.

As indicated by V, the addressee has to assume that the speaker is opinionated as to whether the stronger alternative is true or false in order to derive the scalar implicature. Interestingly, when the speaker's epistemic state does not warrant the 'epistemic step' (i.e. passing from IV to V), the addressee derives the 'weaker' implicature that the speaker does not know whether all of the students did well (Geurts, 2011).[40] If it is manifest enough that the speaker (e.g. the teacher) is

[40] The weak implicature is derived from IV, plus the assumption that the speaker is not competent. While IV corresponds to the proposition $\neg \text{BEL}_S(\psi)$ ('no belief'), the weak implicature corresponds to the following conjunction: $\neg \text{BEL}_S(\psi)$ & $\neg(\text{BEL}_S(\neg\psi)$ ('no opinion'). The latter is stronger that the former (Geurts, 2011, pp. 39-40).

ignorant about the more informative alternative (e.g. all the students did well on the test), the addressee will usually take her utterance to implicate that she does not know whether the more informative alternative holds – thus, the 'don't know' implicature (Carston, 1998). In relevance-theoretic terms, if it is mutually manifest that the teacher would like to be more specific, then her utterance (38a), together with the presumption that her utterance would be the most relevant one compatible with her abilities and preference, will imply that she is *unable* to be more specific, that is, that the teacher does not know whether or not all the students did well on the test.

Much less attention has been devoted to the role played by non-epistemic mental states. In this respect, Relevance Theory represents a significant exception. The relevance-guided interpretation process is said to be constrained by both the speaker's 'abilities' and the speaker's 'preferences'. The latter correspond to a wide range of goals in addition to the fundamental communicative goal (e.g. the desire to withhold some information from the interlocutor). In virtue of this, Relevance Theory is able to offer an integrated account of the different kinds of implicature which may arise from an utterance of the form 'Some of the X…'. Among these are the class of 'don't want to say' implicatures, like the one generated by B's utterance (in an appropriate context) in the following exchange:

(39) A: Which of your colleagues support the strike?
B: Some of them do.

(Carston, 1998, p. 271)

If it is mutually manifest that B could have been more specific (e.g. if it is mutually manifest that B knows who among her colleagues supports the strike), then her answer, together with the presumption that her utterance would be the most relevant one compatible with her abilities and preference, will imply that she is *unwilling* to be more specific, that is, that she doesn't want to say which colleagues support the strike.

To sum up, according to Relevance Theory, an utterance containing the scalar expression 'some' does not necessarily give rise to a scalar implicature of the form 'not all…'. The scalar implicature is derived if the context satisfies the following

constraints: it is upper-bound, that is, the relevance of the stronger alternative is mutually manifest to the interlocutors, and the speaker is both willing and able to provide the hearer with information about the stronger alternative.

6.4.3 Relevance Theory and face-threatening contexts

I now turn to Relevance Theory and its predictions regarding the effect of politeness on scalar inferences. In the previous section, I emphasised that the derivation of genuine scalar implicatures crucially depends on contextual premises as to whether the stronger alternative is entertained as a relevant possibility. In other terms, the scalar implicature '[The speaker believes that] not all X-ed' is derived only if the alternative 'All *X*-ed' is entertained as a relevant possibility in the context at issue. In the contexts set up by Bonnefon et al. (2009) in the three experiments discussed above, we may assume that it is mutually manifest among the interlocutors that the alternative 'All *X*-ed' (e.g. all members hated your poem) would be more relevant to the addressee (who, in all the three experiments, coincides with the participant himself/herself). According to Relevance Theory, this leads to three different possibilities: (a) the derivation of the scalar implicature '[The speaker believes that] not all X-ed', (b) the derivation of the 'don't know' implicature (i.e. 'The speaker does not know whether all *X*-ed'), and (c) the derivation of the 'don't want to say' implicature (i.e. 'The speaker is reluctant to say that all *X*-ed'). The interpretative route is chosen among (a)-(c) on the basis of context-specific considerations about the speaker's 'abilities' and 'preferences'. Consider again the following scenario (Experiment 1):

(32) Imagine that you have joined a poetry club, which consists of five members in addition to you. Each week one member writes a poem, and the five other members discuss the poem in the absence of its author. This week, it is your turn to write a poem and to let the others discuss it. After the discussion, one fellow member confides to you that "Some people hated your poem".
(Bonnefon et al., 2009, p. 251)

According to Relevance Theory, if it is mutually manifest that the speaker has complete knowledge, the addressee would either infer that the speaker is unwilling to be more specific or that she believes that the stronger alternative 'All of the people hated your poem' is false. If either of these possibilities is manifest and relevant, it

will be implicated. In the former case, the hearer would derive the 'don't want to say' implicature (i.e. *The speaker is reluctant to say that all the people hated the poem*); in the latter he would derive the 'scalar implicature' (i.e. *[The speaker believes that] Not all of the people hated the poem*). The possibility that the speaker is unwilling to be more specific could be manifest to the hearer in circumstances in which the speaker is clearly trying to be elusive and vague.[41] So, in a face-threatening context, a speaker attempting to save the face of the hearer in such a way may increase the manifestness of the possibility that she is unwilling to be more specific. As a consequence, the hearer may derive the 'don't want to say' implicature. This is, however, a somewhat peculiar scenario since it requires the speaker's politeness/face-saving concerns to be mutually manifest to the speaker and the hearer.

In all other circumstances, it seems plausible to assume that the possibility that the speaker believes that not all *X*-ed would be manifest enough to make it mutually manifest that the speaker intended it to be manifest, since it increases the relevance of her utterance and is compatible with her manifest preferences. As suggested in section 6.3.1, however, the addressee may end up rejecting this communicated content on the basis of considerations about the speaker's 'preferences'. He may fail to accept the scalar inference as true because he believes that the speaker's preference for politeness rather than strict honesty underlies her communicating to him that not all of the people hated his poem. As a consequence, he may not rule out the possibility, incompatible with the scalar inference, that all of the people hated his speech (and that the speaker believes this to be so). In this case, considerations about the speaker politeness 'preferences' do not affect the interpretation of the utterance, but rather the assessment of its believability.

In conclusion, when the speaker is assumed to be in a position to know whether the more relevant alternative 'All *X*-ed' holds, the possible effects of politeness on the

[41] For instance, the speaker may rely on prosodic clues such as the use of a B-accent on 'some' (Jackendoff, 1972), which usually indicates that the speaker is unwilling to utter (at least) one alternative (for reasons that need to be further contextually specified).

derivation of scalar inferences *Not all X-ed* are diverse.[42] If it is mutually manifest that the speaker has an opinion as to whether 'All *X*-ed' is true (whose content is not manifest to the addressee), the addressee may derive either the 'don't want to say' implicature, or the scalar implicature *[The speaker believes that] not all people hated my poem*. In the former case, which is presumably less common, politeness triggers the derivation of the 'don't want to say' implicature; in the latter, it puts under discussion the believability of the communicated scalar inference.

Summing up: while Bonnefon and colleagues present a very interesting set of data, whose explanation needs to be integrated within theoretical pragmatics frameworks, the contexts adopted in their studies do not provide the means for discriminating Relevance Theory's different predictions on the effect of politeness on the derivation of scalar implicatures. Furthermore, Relevance Theory does not predict that face-threatening utterances of the form 'Some *X*-ed' would be interpreted as conveying *All X-ed* in any context. However, this is the only interpretation that Bonnefon et al. (2009) try to account for in light of the relevance-theoretic framework. It comes as no surprise, then, that the result falls short of theoretical adequacy.

Recall that one of the aims of Bonnefon et al. (2009) is to broaden the empirical scope of the debate between 'defaultist' and 'context-sensitive' accounts of scalar inferences by identifying new contexts which could tell apart the predictions of the two accounts. With this in mind, it is interesting to notice that Relevance Theory seems to predict that in the face-threatening contexts set up by Bonnefon et al. the scalar utterance always gives rise to an implicature (be it a 'don't want to say' implicature or a scalar implicature). If this is the case, the contextual manipulation investigated here may not provide an appropriate means to test alternative theories of scalar inferences ('default' accounts vs. 'context-sensitive' accounts). The problem here is that, even if it may (occasionally) make the scalar inference inappropriate, the expected speed of interpretation of 'some' may not vary across the two models. While default accounts predict that the time taken to derive the interpretation would

[42] If it is mutually manifest that the speaker does not have an opinion as to whether 'All *X*-ed' holds, the addressee is likely to derive the 'don't know' implicature (e.g. *Denise does not know whether all of the people hated your speech*). In this case, considerations about politeness do not enter the picture.

increase in virtue of the extra cognitive effort involved in cancelling the inappropriate scalar inference (which has arisen automatically), relevance theorists predict that the time taken would (equally?) increase because of the cognitive effort involved in deriving the 'don't want to say' implicature.[43]

6.5 Scalar inferences and pro/contra-attitudinal contexts

Heyman, Schaeken and Pipijn (2012) investigate the effect of further contextual manipulations on the derivation of scalar inferences. Specifically, they target the following 'subjective manipulation': contexts can be either pro-attitudinal or contra-attitudinal (or neutral), depending on whether some available assumption (e.g., the content of an utterance) is in line with the feelings, motives and beliefs of the addressee. If it is, then the context is pro-attitudinal; if it is not, the context is contra-attitudinal. They suggest that scalar inferences are less likely to be derived in pro-attitudinal contexts, that is, contexts in which the scalar utterance is in line with the beliefs and attitudes of the addressee. For instance, in (40), an addressee who is in favour of the death penalty would be less likely to derive the scalar inference (*Not in all countries where the death penalty is applied, crime has decreased*) than he would in neutral and contra-attitudinal contexts.

(40) a. In some countries where the death penalty is applied, crime has decreased.
b. In some but not all countries where the death penalty is applied, crime has decreased.

In order to test this hypothesis, Heyman and colleagues conducted the following study, whose experimental design is adapted from Bonnefon et al. (2009). Participants were presented with a questionnaire to gather their attitude towards several issues (e.g., the death penalty, legalisation of soft drugs, etc.). They were asked to indicate on a 5-point scale anchored at *strong in favour* (1) and *strong in disfavour* (5) how they felt about each issue. Furthermore, they indicated how certain they felt about their answers (1 = *not sure*, 5 = *very sure*). The critical ratings were

[43] More generally, this problem seems to extend to context-sensitive accounts other than Relevance Theory. It is usually assumed that upper bound contexts in which the derivation of the scalar implicature is blocked give rise to implicatures of other sorts (e.g. non-epistemic quantity implicatures, see Geurts (2011, p.p. 35-36)).

those concerning the death penalty. In a second phase of the experiment, participants read the following story:

(41) A research group of sociologists and criminologists did a large-scale study into the relationship between death penalty and crime. From this research has become clear that in some countries where the death penalty is applied, crime has decreased/increased.

(Heyman et al., 2012, p. 458)

Half of the participants read the 'decreased' version of the story, while the other half read the 'increased' version of it. Finally, participants where asked the following yes/no question: "Do you think then that it is possible that in all countries where the death penalty is applied, crime has decreased/increased?".

Participants were divided into three groups depending on their responses to the attitude questionnaire: in favour of the death penalty (1 or 2 on the scale), neutral to the death penalty (3), and against the death penalty (4 or 5). Participants whose certainty ratings were equal or lower than 3 were excluded from the analysis.

In line with Heyman et al.'s (2012) prediction, the results show that the percentage of 'No' answers was 65.5% in the pro-attitudinal condition, 82.2% in the neutral condition, and 91.8% in the contra-attitudinal condition. While discussing these results, Heyman et al. (2012, p. 459) conclude that "[d]epending on one's attitude, the scalar is more likely to be interpreted pragmatically, in contra-attitudinal utterances, or logically, in pro-attitudinal utterances, compared with a neutral condition".

In light of the discussion on the methodological impications of the cognitive distinction between comprehension and acceptance (see section 6.3.2), it should be clear that the same methodological concerns which apply to Bonnefon et al.'s (2009) studies carry over to Heyman et al.'s (2012) investigation. Once again, what the results tell us is simply that in pro-attitudinal contexts, participants are less likely to rule out the possibility that 'All X-ed' (i.e. *In all countries where the death penalty is applied, crime has decreased*) when told that 'Some X-ed' (i.e. *In some countries where the death penalty is applied, crime has decreased*). This pattern of results is

compatible with two alternative lines of explanation. The first, which is put forth by Heyman et al. (2012), is that scalar inferences are less likely to be derived in pro-attitudinal contexts. The second is that scalar inferences are less likely to be accepted as true in pro-attitudinal contexts.

Interestingly, the acceptance rate of the scalar inference in pro-attitudinal contexts would be modulated by epistemic vigilance mechanisms towards the content. When a piece of incoming information (e.g., the scalar inference *In some but not all countries where the death penalty is applied, crime has decrease*) contradicts some background belief of the addressee (e.g., *where the death penalty is applied, crime typically decreases*), the addressee may be less likely to accept it as true. This is particularly true when the background belief at issue is held with much conviction. As emphasised by Sperber et al. (2010), belief revision depends on the strength of the beliefs involved (as well as on the degree of confidence in the reliability of the source), That is, in case of contradiction, coherence is achieved via either mistrusting the source (when the incoming piece of information contradicts strongly held background beliefs) or by revising weak background beliefs (when the source is regarded as quite authoritative). In all other cases, coherence checking may involve a conscious decision in order to resolve the issue.

6.6 Future research

6.6.1 The state of the art of the experimental investigation

The literature in experimental pragmatics has fruitfully focused on scalar inferences and investigated how several contextual factors modulate their derivation. Among these, politeness and subjective attitude have been recently explored and face-threatening and pro-attitudinal contexts have been claimed to block the derivation of scalar inferences (similarly to lower-bound contexts). To my knowledge, this conclusion has been accepted as uncontroversial (e.g., Cummins, 2013).

In the previous sections, I have raised methodological and theoretical issues with regard to the studies on politeness and scalar inferences proposed by Bonnefon and colleagues (Bonnefon et al., 2009; Feeney & Bonnefon, 2012) and adapted by

Heyman et al. (2012) to the study of pro-/contra-attitudinal contexts. As argued in section 6.3.2, the experimental evidence provided so far falls short of demonstrating that scalar inferences are less likely to be derived in face-threatening contexts. In fact, it is compatible with an alternative explanation: politeness may affect the believability (rather than the derivation) of the scalar inference. In face-threatening contexts, addressees may be less likely to accept the scalar inference as true because they recognise that the speaker is being polite. Furthermore, I suggested that Relevance Theory makes predictions that are compatible with this line of explanation (but not with the one proposed by Bonnefon et al. (2009)).

The crucial distinction to be drawn here is the one between the speaker's intended meaning (i.e. the communicated content of the utterance) and the beliefs about the world that the hearer adopts on the basis of that communicated content. Significantly, the hearer may attribute an interpretation to the speaker without accepting it as true. Even more, he may end up forming a belief that is incompatible with the content communicated by his interlocutor. When told that 'some' of the people hated his poem, the hearer may end up believing that *all* of them did, even if he attributes to the speaker the intention to convey that *some but not all* the people hated the poem.

The very same distinction can be used to undermine the experimental evidence proposed to support the hypothesis that (self-rated) honesty positively correlates with the derivation of scalar inferences. Far from showing the effect of individual differences in the *interpretation* of scalar expressions, Feeney and Bonnefon's results are (at least) compatible with a very different explanation: more honest people may be more likely to accept the scalar inference as true because of their tendency to grant epistemic trust to the interlocutor.

Similarly, Heyman et al. (2012) results could be interpreted along the following line: addressees are less likely to accept the scalar inference when its content contradicts some of their background beliefs. That is, the subjective contextual manipulation investigated by Heyman and colleagues (pro- vs. contra-attitudinal contexts) may modulate the acceptance rate (i.e. the believability) of the scalar inference rather than its derivation.

To sum up, the focus on the *derivation* of scalar inferences (as part and parcel of the comprehension process) has obscured the fact that scalar inferences may be derived as part of the speaker's intended meaning and yet fail to be *accepted* as true by the addressee. Crucially, behavioural responses in the experimental settings of Bonnefon and colleagues, as well as in those of Heyman and colleagues, fail to distinguish between (and consequently amalgamate) cases in which the scalar inference is not derived by the addressee and cases in which the scalar inference is indeed derived as part of the speaker's meaning, but it is not accepted as true. This potential for confusion is what beckons the distinction between comprehension and acceptance, a distinction that can be addressed through both theoretical insight and empirical research.

In this section, I suggest some lines of research for experimentally investigating the distinction between comprehension and acceptance of scalar inferences discussed above. The aim is to shed light on how epistemic vigilance mechanisms modulate the acceptance of scalar inferences. This breaks down into the investigation of the following hypotheses:

(i) Epistemic vigilance mechanisms *towards the source* may affect the believability of scalar inferences in face-threatening contexts. In face-threatening contexts, as in (31), participants recognise that the speaker is trying to be nice and polite (by allowing the listener to generate a reading that can be glossed as *Some but not all the guests thought A drank too much*); however, participants also recognise that it is probable that B's comment is not entirely honest and, consequently, they judge part of what she communicates as not true, so they do not accept it. In other words, participants are more likely to conclude that (probably) *all the guests thought A drank too much*.

(ii) Epistemic vigilance mechanisms *towards the content* may affect the believability of scalar inferences in pro-attitudinal contexts. When the scalar inference contradicts some background belief(s) of the participants, they may derive it as a component of speaker meaning but be less likely to accept it as true. For example, in (41), pro-death penalty participants may derive the scalar inference that *In some but not all countries where the death penalty is applied, crime has decreased*, but resist

accepting it as true (and integrating it into their belief-system). As a consequence, participants are more likely to conclude that *it is not the case that not in all countries where the death penalty is applied, crime has decreased*.

6.6.2 The believability of scalar inferences in face-threatening contexts[44]

Study 1

The aim is to conduct a self-paced reading study to measure the time course of processing scalar utterances in face-boosting contexts, face-threatening contexts and neutral/lower-bound contexts, that is, contexts in which there is no explicit or tacit question as to whether 'All X-ed' is true. The hypotheses to be tested are the following:

H_1 (from Bonnefon et al. (2009)): scalar inferences are less likely to be derived in face-threatening contexts.
H_0 scalar inferences are *derived* equally across face-boosting and face-threatening contexts but are less likely to be *accepted* as true in the latter.

The neutral/lower-bound context provides a baseline against which to compare the time course of processing scalar utterances in upper-bound face-boosting and face-threatening contexts (where the scalar inference is informative and relevant). Table 1 displays a sample of the experimental conditions.

H_0 predicts significantly different reading times for scalar utterances in neutral contexts, on the one hand, and scalar utterances in face-boosting and face-threatening contexts, on the other, given that the scalar inference would be licensed in the latter two (irrespectively of the face-boost/face-threat) but not in the former: reading times are thus expected to be longer for face-boosting and face-threatening scalar utterances than for scalar utterances in neutral contexts. In contrast, H_1 predicts that processing a scalar utterance in a face-boosting context would take significantly longer than processing it in a face-threatening or neutral context, given that face-threatening contexts would block its derivation.

[44] I am hugely indebted to Ira Noveck for valuable discussions on the design of the experiments presented in this section.

Imagine you gave a speech at a small political rally. You are discussing your speech with Denise, who was in the audience. You are considering whether to give this same speech to another audience and would like to know the audience's reaction.		
Neutral context	**Face-threatening context**	**Face-boosting context**
Hearing this, Denise tells you: "I missed most of it. I found the sound-system in the room to be very poor and I saw that **/some people felt the same/**"	Hearing this, Denise tells you: "I'm afraid the party leader didn't like it and I saw that /**some people felt the same/**"	Hearing this, Denise tells you: "I think the party leader loved it and I saw that /**some people felt the same/**"

Table 1 Sample of conditions

Study 2

Using the more fine-grained ERP technique it is possible to locate the actual moment of scalar enrichment (Noveck & Posada, 2003). However, it has been shown that individual differences between participants must be taken into account: Nieuwland, Ditman and Kuperberg (2010) reported steeper N400's (which reflect depth of semantic processing) for the final word in underinformative clauses, e.g. 'lungs' in (42a), when compared to the final word in control clauses, e.g. 'pets' in (42b), among socially inclined participants only.

(42) a. Some people have lungs, which require good care.
b. Some people have pets, which require good care.

Social inclinations were determined by participants' *Communication Score* in Baron-Cohen's Autism Quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). The materials proposed in Table 1 transfer well to an ERP study. Plus, with regard to Nieuwland et al.'s study, the proposed paradigm has the added advantage of providing greater context and *identical* test-sentences.

In other words, Study 2 is similar to Study 1 but with EEG. Besides collecting social measures to determine whether individual differences are associated with pragmatic

enrichment, it can also determine whether there are differences among the conditions. It is also possible that these materials – which are more engaging than stand-alone sentences – will produce differences among all participants (regardless of their AQ score).

6.6.3 The believability of scalar inferences in pro-attitudinal contexts

Study 3

The aim is to conduct self-paced reading to measure the time course of processing pro-attitudinal scalar utterances. The hypotheses to be tested are the following:

H_1 (from Heyman et al. (2012)): scalar inferences are less likely to be derived in pro-attitudinal contexts.
H_0 scalar inferences are *derived* but are less likely to be *accepted* as true in pro-attitudinal contexts.

Based on their responses to an attitude questionnaire (as in Heyman et al. (2012)), participants are divided into three groups, for example: A) in favour of the death penalty, B) against the death penalty and C) neutral towards the death penalty. The critical group for the current study is the one in favour of the death penalty (A). The group is presented with a self-paced reading task in which pro-attitudinal scalar utterances are displayed in three conditions: upper-bound context, lower-bound context, 'only some' context. Scalar utterances are followed by a second sentence, which always starts with the target phrase 'In the rest' (see Table 2). As shown by Breheny et al. (2006) (see section 6.1), the derivation of the scalar inference facilitates the interpretation of the anaphoric expression and results in shorter reading time of the target phrase.

H_0 predicts significantly different reading times for the target phrase ('In the rest') in the upper-bound and 'only some' contexts, on the one hand, and lower-bound context, on the other hand, given that the scalar inference would be licensed in the former two but not in the latter: reading times for the target phrase are thus expected to be longer in the lower-bound context. In contrast, H_1 predicts that interpreting the target phrase in both upper-bound and lower bound contexts would take significantly

longer than in the 'only some' context given that the pro-attitudinal content of the scalar utterance would block the derivation of the scalar inference in the upper-bound context.

A research group of sociologists and criminologists did a large-scale study into the relationship between the death penalty and crime to investigate the hypothesis that the enforcement of the death penalty reduces rates of criminality.		
Upper-bound context	**'Only-some context'**	**Lower-bound context**
Data collection confirmed the hypothesis in some countries. **In the rest** no supporting evidence was gathered from the crime records.	Data collection confirmed the hypothesis in only some countries. **In the rest** no supporting evidence was gathered from the crime records.	Researchers worked in cooperation with police officers in some countries. **In the rest** access to the crime records proved to be difficult.

Table 2 Sample of conditions

Summary and conclusion

Current research on linguistic communication is grounded on the well-established assumption that speakers typically communicate more than they linguistically encode (Carston, 2002; Grice, 1989; Sperber & Wilson, 1986/1995). Pragmatics is the study of how addressees bridge the gap between linguistic meaning and communicated meaning (or 'speaker's meaning'). As suggested by Neale (2013), pragmatic investigation aims at addressing the following question:

PQ What sources of information and what types of cognitive operations drive the (typically spontaneous) formation of A's interpretative hypotheses about what S means by uttering X on a given occasion?

With this question in mind, I have focused on the following topics of investigation: (i) the nature of pragmatic processing (i.e. the *type* of cognitive operations involved), and (ii) its interfaces with other cognitive systems, in particular 'epistemic vigilance', that is, mechanisms which check the quality of the incoming information and the reliability of the individuals who provide it (Sperber et al., 2010).

The first part of my thesis addressed the question of whether pragmatic processes are *inferential* in nature. Gricean and post-Gricean pragmatic theories (e.g. Relevance Theory) maintain that linguistic communication is essentially inferential. Pragmatic processing is conceived of as a kind of non-demonstrative inference – an *inference to the best explanation* – where addresses form a hypothesis about the speaker's communicated meaning on the basis of linguistic evidence and other contextual cues, which is evaluated in light of general principles of communication which speakers are expected to observe.

This 'inferential view' has been recently challenged by 'accessibility-based' accounts (Mazzone, 2011, 2013; Recanati, 2002a, 2004). According to these, pragmatic processing is underpinned by a low-level, brute-causal process whose output is simply the most accessible (i.e. the most highly activated) interpretative hypothesis (with no appeal to pragmatic principles/maxims).

The debate between inferential and accessibility-based approaches to pragmatics has been affected by terminological and conceptual confusions about the notion of 'inference' at issue. For this reason, Chapter 1 addressed the preliminary question of what counts as an 'inferential' cognitive process.

I reviewed and assessed criteria for distinguishing between inferential and non-inferential cognitive processes which have been proposed in the literature. I formulated a tripartite classification of the levels of analysis at which the distinction between inferential and non-inferential approaches can emerge. The aim of this terminological/conceptual clarification was two-fold: on the one hand, to develop a more fine-grained account of the role of inference in utterance interpretation, and, on the other hand, to provide the background to assess the debate between inferential and non-inferential approaches to pragmatics.

I proposed the following criterion to distinguish between inferential and non-inferential approaches to pragmatics. While the former conceive of pragmatic interpretation as a two-stage process involving 'hypothesis formation' and 'hypothesis confirmation', accessibility-based approaches collapse these two stages: a simple dynamics of differential activation of concepts in associative networks can account for how an interpretation is selected, with no confirmatory stage being involved.

In Chapter 2, I reviewed and assessed some recent experimental literature focusing on the debate between inferential and accessibility-based approaches to pragmatics (Rubio-Fernández, 2007, 2013). Specifically, I addressed the question of whether non-demonstrative inferential processes implemented in a fast and automatic system can be empirically disentangled from 'smart' associative systems. I concluded that the distinction between these two kinds of system (smart associative and automatic inferential) is too fine-grained to be tested by currently available experimental techniques. So a theoretical investigation of the conceptual tools employed by alternative frameworks is essential in providing complementary methods for evaluating these alternative approaches. This investigation is the task I set myself in Chapter 3.

In Chapter 3, I argued for the indispensability of a criterion of pragmatic acceptability based on consideration of the speaker's mental states (i.e. her beliefs and intentions): a 'rational justification' of the interpretation derived requires this kind of evaluation (and confirmation) before it can be attributed to the speaker as her intended meaning. That is, pragmatic interpretation requires a stage of 'hypothesis confirmation' in which hypotheses about the speaker's meaning are evaluated in light of considerations of the speaker's mental states. Importantly, I showed that accessibility-based approaches to pragmatics cannot accommodate speaker's beliefs and other mental states even when these crucially affect the interpretation of an utterance. Furthermore, accessibility-based accounts claim that the level of accessibility of an interpretative hypothesis determines its success in the interpretative process. This predicts that information that is inconsistent with the interpretative hypothesis under construction would be deactivated. However, as discussed in section 3.8, this very same information may play a significant role in the process of evaluating the believability of the selected interpretation. It follows that accessibility cannot be the decisive factor in the selection of the intended interpretation.

The relationship between comprehension and acceptance (belief) was thoroughly explored in Chapter 4. Two main models of belief-formation via testimony have been discussed in the philosophical and psychological literature: the 'Spinozan' model and the 'Cartesian' model. While the Spinozan model suggests that comprehension involves the automatic acceptance of the communicated content (which can be 'unaccepted' at a later, optional, stage), the Cartesian model conceives of comprehension and epistemic assessment as two distinct and sequential processes. Interestingly, according to both the Spinozan and the Cartesian model, epistemic assessment is an optional and resource-demanding stage that follows comprehension (whether this involves acceptance or suspension of belief). In Chapter 4 I reviewed the current experimental evidence on the time course of epistemic assessment and the effect of cognitive load on its operations. This evidence suggest that comprehension and epistemic assessment are parallel processes and that some form of spontaneous epistemic assessment is not prevented by extra-cognitive load but rather modulated by expectations of relevance. In line with this, I introduced and discussed Sperber et

al.'s (2010) 'pragmatic model' of the relationship between comprehension and acceptance. According to Sperber et al. (2010), comprehension and epistemic assessment are triggered by any act of ostensive communication: while comprehension follows a relevance guided procedure which selects the interpretative hypothesis that would be relevant to the hearer if he accepted it, epistemic vigilance mechanisms assess the speaker's reliability and the factual plausibility of the communicated content and establish whether the selected interpretation is indeed worth being accepted as true: "comprehension, the search for relevance, and epistemic assessment are interconnected aspects of a single process whose goal is to make the best of communicated information" (Sperber et al., 2010, p. 376).

In Chapter 5, I argued that, not only do epistemic vigilance mechanism assess the believability of the output of the comprehension system, but they also play a crucial role in determining that output. As discussed above, Chapter 3 established that pragmatic interpretation involves a stage of 'hypothesis confirmation' where interpretative hypotheses about the speaker's meaning are evaluated in light of (what the addressee takes to be) the speaker's mental states (i.e. her beliefs and intentions). This raised the question of what cognitive system(s) implements this checking process. I suggested that the answer to this question is to be found in the interaction between the operations of the relevance-guided comprehension procedure and epistemic vigilance mechanisms. Specifically, epistemic vigilance mechanisms towards the source modulate the expectations of relevance that drive the comprehension procedure. If they detect that the speaker is not competent, they downgrade the initial expectation of actual optimal relevance to *attempted* optimal relevance. Similarly, if epistemic vigilance mechanisms towards deception detect that the speaker is not benevolent, they set the expectations of relevance at an even lower grade, that is, *purported* optimal relevance. As a consequence, epistemic vigilance mechanisms may filter out interpretative hypotheses that are incompatible with assumptions about the speaker's epistemic state or retain interpretative hypotheses that are irrelevant to the addressee (e.g. because he knows them to be false) but compatible with the speaker's deceptive intention. This has important consequences for the relationship between comprehension and acceptance discussed in Chapter 4. Crucially, epistemic assessment taking place *after* comprehension can

be made redundant whenever the selection of the intended interpretative hypothesis is grounded on the recognition that the speaker has a deceptive intention. If it is by realising that the speaker has the intention to mislead him that the addressee recognises what the speaker intended to communicate, there is no need to further assess the believability of the misleading pieces of communicated information. However, when the role of epistemic assessment in comprehension is that of warranting an interpretation that is compatible with the speaker's epistemic state, it remains an open question whether the communicated content is worth being accepted as true. In this latter circumstance, further epistemic assessment would typically follow comprehension.

Finally, Chapter 5 discussed the implications of the proposed interaction between comprehension and epistemic vigilance mechanisms for the development of pragmatic abilities. Building on Sperber's (1994b) hypothesis that children deploy increasingly sophisticated interpretative strategies which are based on the recognition that the speaker may not be *competent* or *benevolent*, I proposed that such strategies emerge thanks to the unfolding of epistemic vigilance capacities in the child's cognitive development. This sheds a new light on the relation between communicative abilities and other types of metarepresentational capacity (in particular, epistemic vigilance), which is open to further empirical investigation.

The experimental research in pragmatics has not yet focused on the distinction between comprehension and epistemic assessment or on the role played by epistemic vigilance in determining the output of the comprehension system, as well as in assessing its believability. As discussed in Chapter 6, this has important methodological implications. Recent behavioural studies have investigated the effect of contextual manipulations on the derivation of scalar inferences. Two of these manipulations were relevant for our discussion: face-threatening vs. face-boosting contexts (Bonnefon et al., 2009; Feeney & Bonnefon, 2012), and pro-attitudinal vs. contra-attitudinal contexts (Heyman et al., 2012). The results of these studies are interpreted as suggesting that face-threatening and pro-attitudinal contexts may block the derivation of the scalar inference as participants were found to be more likely to accept the possibility that 'All X-ed' when told that 'Some X-ed' (and the scalar

utterance was face-threatening/pro-attitudinal). For instance, they were more likely to believe that it was possible that all the people hated their poem when told that some did. Crucially, the off-line paradigm adopted in these studies amalgamates cases in which the addressee does not derive the scalar inference, and cases in which the scalar inference is indeed derived as part of the speaker‘s meaning, but it is not accepted as true. The suggested hypothesis is that epistemic vigilance mechanism may affect the believability of scalar inferences in the contexts at issue. Specifically, epistemic vigilance mechanisms *towards the source* may affect the believability of the scalar inference in face-threatening contexts (by recognising that the speaker may want to be polite or kind rather than strictly honest), whereas epistemic vigilance mechanisms towards the content may affect the believability of the scalar inference in pro-attitudinal contexts (by rejecting a piece of communicated information that contradicts some background beliefs of the addressee). Chapter 6 concluded by sketching directions for experimental studies which could test this hypothesis via the adoption of more fined-grained empirical measurements (e.g. reading times).

In conclusion, I developed a model of the relationship between comprehension and epistemic assessment, according to which they are more closely intertwined than assumed by traditional models of belief-formation via testimony. The account proposed is grounded in a compelling analysis of the pragmatic dimension of comprehension. This suggests a programme of future experimental research on the interaction of different metarepresentational capacities in the development and functioning of communication.

References

Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*, 82–93.

Bonnefon, J.-F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, *112*(2), 249–258.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457.

Breheny, R. E. T., Ferguson, H., & Katsos, N. (2012). Investigating the time course of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes*, *28*(4), 443–467.

Breheny, R. E. T., Ferguson, H., & Katsos, N. (2013). Taking the epistemic step: toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423–440.

Breheny, R. E. T., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Brown-Schmidt, S., & Hanna, J. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialog and Discourse*, *2*(1), 11–33.

Carpenter, M., Call, J., & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology*, *20*, 393–420.

Carston, R. (1990). Quantity maxims and generalised implicature. *UCL Working Papers in Linguistics*, *2*, 1–31. Reprinted in *Lingua*, *96*, 213-244.

Carston, R. (1997). Relevance-theoretic pragmatics and modularity. *UCL Working Papers in Linguistics*, *9*, 1–27.

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance Theory: Applications and Implications* (pp. 179–236). Amsterdam: John Benjamins.

Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Carston, R. (2007). How many pragmatic systems are there? In M. J. Frapolli (Ed.), *Saying, Meaning, Referring. Essays on the Philosophy of Francois Recanati* (pp. 18–48). New York: Palgrave Macmillan.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures, Volume 3* (pp. 39–103). Oxford: Oxford University Press.

Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, & K. von Heusinger (Eds.), *Handbook of semantics, Vol. 3* (pp. 2297–2331). Mouton de Gruyter.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 85–116). Cambridge, MA: Cambridge University Press.

Cummins, C. (2013). Modelling implicatures from modified numerals. *Lingua*, *132*, 103–114.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128–133.

Descartes, R. (1664). *Principles of Philosophy*. Dordrecht: Reidel.

Einav, S., & Robinson, E. J. (2011). When being right is not enough: Four-year-olds distinguish knowledgeable from merely accurate informants. *Psychological Science*, *22*, 1250–1253.

Feeney, A., & Bonnefon, J.-F. (2012). Politeness and Honesty Contribute Additively to the Interpretation of Scalar Expressions. *Journal of Language and Social Psychology*, *32*(2), 181–190.

Fodor, J. (1983). The modularity of mind. Cambridge, MA: MIT Press.

García-Carpintero, M. (2006). Recanati on the sematics/pragmatics distinction. *Crítica*, *38*(112), 35–68.

Geurts. (2011). *Quantity Implicatures*. Cambridge: Cambridge University Press.

Gilbert, D. T. (1991). How Mental Systems Believe. *American Psychologist*, *46*, 107–119.

Gilbert, D. T. (1993). The assent of man: Mental representation and the control of belief. In D. M. Wegner & J. W. Pennebaker (Eds.), *Handbook of mental control Century psychology series* (pp. 57–87). Englewood Cliffs: Prentice-Hall.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, *59*, 601-613.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, *65*, 221–233.

Grandy, R. (1989). On Grice on language. *Journal of Philosophy*, *86*, 514–525.

Grice, H. P. (1957). Meaning. *Philosophical Review*, *66*, 377–388.

Grice, H. P. (1968). Utterer's meaning, sentence meaning and word meaning. *Foundations of Language*, *4*, 225–242.

Grice, H. P. (1975). Logic and Conversation. In D. Davidson & G. Harman (Eds.), *The Logic of Grammar* (pp. 64–75). Encino, CA: Dickenson.

Grice, H. P. (1989). *Studies in the Way of Words*. *Philosophy* (Vol. 65). Cambridge, MA: Harvard University Press.

Grosse, G., Behne, T., Carpenter, M., & Tomasello, M. (2010). Infants communicate in order to be understood. *Developmental Psychology*, *46*(6), 1710–1722.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science (New York, N.Y.)*, *304*, 438–441.

Hall, A. (2014). "Free" enrichment and the nature of pragmatic constraints. *International Review of Pragmatics*, *6*(1), 1–28.

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA: Harvard University Press.

Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe It or Not. *Psychological Science*, *16*, 566–571.

Heyman, T., Schaeken, W., & Pipijn, K. (2012). Logical or pragmatic, as long as it suits our convenience: Scalar inferences in a pro- and contra-attitudinal context. In N. Miyake, D. Peebles, & R. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*. Austin, Tx: Cognitive Science Society.

Hirschberg, J. (1985). *A theory of scalar implicature*. PhD dissertation. University of Pennsylvania.

Horn, L. R. (1972). *On the semantic properties of logical operators in English*. PhD dissertation. University of California, LA.

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Shiffrin (Ed.), *Meaning, form, and use in context* (pp. 11–42). Washington, DC: Georgetown University Press.

Horn, L. R. (2004). Implicature. In L. R. Horn & G. L. Ward (Eds.), *The Handbook of Pragmatics* (pp. 3–28). Oxford: Blackwell Publishing.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Jary, M. (2013). Two types of implicature: material and behavioural. *Mind & Language*, *28*(5), 638–660.

Kasher, A. (1991). On the pragmatic modules: A lecture. *Journal of Pragmatics*, *16*, 381-397.

Katsos, N. (2008). The semantics/pragmatics interface from an experimental perspective: the case of scalar implicature. *Synthese*, *165*(3), 385–401.

Katsos, N., Breheny, R. E. T., & Williams, J. (2005). The interaction of structural and contextual constraints during the on-line generation of scalar inferences. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1108–1113). Mahwah, NJ: Erlbaum.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Kissine, M. (2009). In defence of direct perception through language. In J. M. Larrazabal & L. Zubeldia (Eds.), *Meaning, Content and Argument* (pp. 365–381). University of the Basque Country Press.

Kissine, M. (2013). *From Utterances to Speech Acts*. New York: Cambridge University Press.

Koenig, M. a., & Harris, P. L. (2007). The Basis of Epistemic Trust: Reliable Testimony or Reliable Sources? *Episteme*, *4*(03), 264–284.

Lee, J. J., & Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review*, *117*(3), 785–807.

Levinson, S. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature* (MIT Press.). Cambridge, MA.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, *112*(3), 367–380.

Matsui, T., Rakoczy, H., Miura, Y., & Tomasello, M. (2009). Understanding of speaker certainty and false-belief reasoning: A comparison of Japanese and German preschoolers. *Developmental Science*, *12*(4), 602–613.

Mazzarella, D. (2011). Accessibility and relevance. A fork in the road. *UCL Working Papers in Linguistics*, 11–20.

Mazzone, M. (2009). Pragmatics and Cognition: Intentions and Pattern Recognition in Context. *International Review of Pragmatics*, *1*, 321–347.

Mazzone, M. (2011). Schemata and associative processes in pragmatics. *Journal of Pragmatics*, *43*(8), 2148–2159.

Mazzone, M. (2013). Attention to the speaker. The conscious assessment of utterance interpretations in working memory. *Language & Communication*, *33*(2), 106–114.

Mercier, H., & Sperber, D. (2009). Intuitive and reflective inferential processes. In J. S. B. T. Evans & K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond* (pp. 149–170). Oxford: Oxford University Press.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: new Foundations for Realism*. MIT Press.

Millikan, R. G. (2004). *Varieties of Meaning: The 2002 Jean Nicod Lectures*. MIT Press.

Millikan, R. G. (2005). *Language: A Biological Model*. New York: Oxford University Press.

Neale, S. (2013). Determinations of meaning. In *CUNY/NYU Mind & Language Seminar* (pp. 1–52).

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*(3), 324–346.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210.

Noveck, I. A., & Reboul, A. (2008). Experimental Pragmatics: a Gricean turn in the study of language. *Trends in Cognitive Sciences*, *12*(11), 425-431.

Noveck, I. A., & Sperber, D. (2007). The why and how of experimental pragmatics: the case of "scalar inference." In N. Burton-Roberts (Ed.), *Advances in Pragmatics* (pp. 184–212). Basingstoke: Palgrave.

Noveck, I. A., & Spotorno, N. (2014). Experimental Pragmatics. In J. Verschueren (Ed.), *Handbook of Pragmatics*. Benjamins Publishing Company.

Origgi, G. (2008). Trust, authority and epistemic responsibility. *Theoria. Revista de Teoría, Historia Y Fundamentos de la Ciencia*, *23*(61), 35–44.

Padilla Cruz, M. (2012). Epistemic vigilance, cautious optimism and sophisticated understanding. *Research in Language*, *10*(4), 365–386.

Recanati, F. (1989). The pragmatics of what is said. *Mind & Language*, *4*, 295–329.

Recanati, F. (1993). *Direct Reference: From Language to Thought*. Oxford: Blackwell.

Recanati, F. (1995). The alleged priority of literal interpretation. *Cognitive Science*, *19*, 207–232.

Recanati, F. (2002a). Does linguistic communication rest on inference? *Mind & Language*, *17*, 105–126.

Recanati, F. (2002b). Unarticulated constituents. *Linguistics and Philosophy*, *25*, 299–343.

Recanati, F. (2004). *Literal Meaning*. Cambridge University Press.

Recanati, F. (2007a). *Perspectival Thought : A Plea for (Moderate) Relativism: A Plea for (Moderate) Relativism (Google eBook)*. Oxford University Press.

Recanati, F. (2007b). Reply to Carston. In M. J. Frapolli (Ed.), *Saying, Meaning, Referring. Essays on the Philosophy of François Recanati* (pp. 49–54). New York: Palgrave Macmillan.

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, *96*, 538–558.

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: a review and meta-analysis. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *9*(1), 32–47.

Robinson, E. J., & Einav, S. (2014). *Trust and Skepticism: Children's Selective Learning from Testimony*. New York: Psychology Press.

Rubio-Fernández, P. (2007). Suppression in Metaphor Interpretation: Differences between Meaning Selection and Meaning Construction. *Journal of Semantics*, *24*, 345–371.

Rubio-Fernández, P. (2013). Associative and inferential processes in pragmatic enrichment: The case of emergent properties. *Language and Cognitive Processes*, *28*(6), 723–745.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, *27*, 367–391.

Saul, J. M. (2012). *Lying, misleading and what is said: An exploration in philosophy of language and in ethics*. Oxford: Oxford University Press.

Schiffer, S. R. (1972). *Meaning*. Oxford: Clarendon Press.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*(6), 907–912.

Sperber, D. (1994a). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 36–67). Cambridge: Cambridge University Press.

Sperber, D. (1994b). Understanding verbal understanding. In J. Khalfa (Ed.), *What is intelligence?* (pp. 179–198). Cambridge: Cambridge University Press.

Sperber, D. (2001). In defence of massive modularity. In E. Dupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 47–57). Cambridge, MA: MIT Press.

Sperber, D. (2005). Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and content* (pp. 53–68). New York: Oxford University Press.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, *25*(4), 359–393.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and Cognition* (2nd ed.). Oxford: Blackwell Publishing.

Sperber, D., & Wilson, D. (1998). The mapping between the mental and the public lexicon. In P. Carruthers & J. Boucher (Eds.), *Thought and Language* (pp. 184–200). Cambridge: Cambridge University Press.

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind & Language*, *17*(1&2), 3–23.

Sperber, D., & Wilson, D. (forthcoming/2015). Beyond speaker's meaning. *Croatian Journal of Philosophy*.

Strawson, P. F. (1964a). Identifying reference and truth values. *Theoria*, *3*, 96–118.

Strawson, P. F. (1964b) Intention and convention in speech acts. *The Philosophical Review, 73*, 439–60.

Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology, 13*, 191- 204.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592–598.

Wilson, D. (2000). Metarepresentation in linguistic communication. In D. Sperber (Ed.), *Metarepresentations. A multidisciplinary perspective* (pp. 411–448). New York: Oxford University Press.

Wilson, D. (2004). Relevance and lexical pragmatics. *UCL Working Papers in Linguistics*, *16*, 343–360.

Wilson, D. (2005). New directions for research on pragmatics and modularity. *Lingua*, *115*(8), 1129–1146.

Wilson, D., & Carston, R. (2007). A Unitary Approach to Lexical Pragmatics: Relevance, Inference ad Ad Hoc Concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 230–260). Basingstoke: Palgrave Macmillan.

Wilson, D., & Matsui, T. (1998). Recent approaches to bridging: Truth, coherence, relevance. *UCL Working Papers in Linguistics*, *10*, 1–28.

Wilson, D., & Sperber, D. (1986). Pragmatics and modularity. In A. M. Farley, P. T. Farley, & K.-E. McCullough (Eds.), *Chicago Linguistic Society Parasession on Pragmatics and Grammatical Theory* (pp. 67–84). Reprinted in S. Davies (Ed.), Pragmatics: A reader (pp. 583-595). New York: Oxford University Press.

Wilson, D., & Sperber, D. (2004). Relevance Theory. In L. R. Horn & G. L. Ward (Eds.), *Handbook of Pragmatics* (pp. 607–632). Oxford: Blackwell Publishing.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge: Cambridge University Press.

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of Mind deficits and discourse interpretation in right-hemisphere brain-damaged patients. *Brain & Language, 62*, 89-106.

Winner, E., & Leekam, S. (1991). Distinguishing irony from deception. Understanding the speaker's second-order intention. *British Journal of Developmental Psychology, 9*, 257-270.

Yeh, W., & Barsalou, L. (2006). The situated nature of concepts. *The American Journal of Psychology*, *119*(3), 349–384.