‘Optimal relevance’ as a pragmatic criterion: the role of epistemic vigilance*

Diana Mazzarella

Abstract

According to Relevance Theory, pragmatic interpretation is guided by an expectation of (optimal) relevance. This expectation is constrained by considerations about the speaker’s mental states. In this paper I address a recent criticism against Relevance Theory put forth by Mazzone (2009, 2013). This criticism focuses on a core notion within the relevance-theoretic framework, that is, the notion of ‘optimal relevance’ and its role as a pragmatic criterion of acceptability. Mazzone suggests that the appeal to the notion of ‘optimal relevance’ is not enough to show *how* information about the speaker’s mental states (e.g., her knowledge and beliefs) affects on-line pragmatic processing. I develop a tentative line of solution to cognitively implement the notion of ‘optimal relevance’. My proposal is grounded on the interaction between the comprehension module and epistemic vigilance mechanisms, that is, those mechanisms which check the quality of incoming information and the reliability of the individual who dispenses it (Sperber et al., 2010).
Keywords: epistemic vigilance, Relevance Theory, speaker’s mental states, acceptability criterion, massive modularity

1 Introduction

The view of pragmatic interpretation as an inferential enterprise has its roots in the seminal work of Paul Grice (see Grice (1989)). More recent cognitively-oriented approaches to pragmatics (e.g., Relevance Theory) have tried to implement this view in a psychologically plausible account of utterance interpretation by substituting complex discourse reasoning processes with ‘fast and frugal’ comprehension heuristics.

This shift of perspective, from a philosophical to a psychological explanation of utterance interpretation, has involved the notion of inference itself (and how it should be construed). To claim psychological plausibility, inferential models of utterance interpretation need to be built on an equally psychologically plausible notion of inference. Thus the question of what counts as an inferential pragmatic process has become of central importance.

Wilson and Matsui (1998) suggest that the interpretative process can be generally described as involving the following steps:

(i) Candidate interpretations differ in their accessibility, and are therefore entertained in a certain order.
(ii) They are evaluated in terms of some criterion or standard of pragmatic acceptability that the resulting overall interpretation is supposed to meet.

* I am grateful to Robyn Carston for her valuable guidance and insightful discussions about the content of this paper. Thanks also to Deirdre Wilson for her comments on previous drafts of this paper. This work has benefited from discussions at the *3rd SIFA Graduate Conference “Language, Logic and Mind”*. I am grateful to Emma Borg, Carlo Penco and Dan Zeman, whose insightful comments and questions have led to an improved version of this paper. This work is supported by the Leverhulme Trust.

I suggest that what makes a process fully inferential is that it maintains a distinction between these two steps. They can be seen as corresponding to distinct stages of *hypothesis formation* and *hypothesis confirmation* (which are characteristically involved in non-demonstrative inferences). 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 an acceptability criterion in order to be inferentially warranted.

Pragmatic inference is a kind of non–demonstrative inference, that is, an *inference to the best explanation* about the verbal behaviour of the communicator. 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. It follows that a criterion of pragmatic acceptability should enable the selection of the best explanation for the speaker's verbal behaviour.

In what follows I focus on the inferential pragmatic framework proposed by Relevance Theory (Sperber & Wilson, 1986/1995; Wilson & Sperber, 2004, 2012) and on the notion of 'optimal relevance' as an acceptability criterion. I investigate the constraints it imposes on pragmatic interpretation and I sketch the direction for its cognitive implementation. In particular, I suggest that mechanisms of epistemic vigilance (Sperber et al., 2010) may play a role in assessing the acceptability of interpretative hypotheses and that their interaction with the comprehension process may explain how information about the speaker's mental states can affect pragmatic interpretation.

2 Relevance Theory

2.1 General framework

At the core of Relevance Theory is the notion of 'relevance'. This is defined as a property of inputs to cognitive processes (e.g., utterances) and is a cost-benefit notion: the greater the cognitive effects (benefits), the greater the relevance; the smaller the processing effort (cost) required to derive these effects, the greater the relevance. Sperber and Wilson (1986/1995) distinguish among three different kinds 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 an available assumption and contradiction and elimination of an available assumption.

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 (rather than anything else the addressee could have paid attention to at that time) and is 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, 1986/1995, 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, 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.

2.2 Optimal relevance as an acceptability criterion

The relevance-guided comprehension procedure, (3), is a dedicated *inferential* mechanism: 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. The expected level of relevance, thus, determines the stopping point of the relevance-guided comprehension procedure and represents the 'acceptability criterion' that an interpretative hypothesis needs to satisfy in order to be retained and attributed to the communicator.

As discussed in the previous section, the Communicative Principle of Relevance, (1), states that every utterance comes with a presumption of its own optimal relevance, which, in turn, determines the level of relevance that the addressee is *entitled* to expect. It follows that optimal relevance can be seen as the acceptability criterion which determines the stopping point of the comprehension procedure followed by the interpreter.[1]

In this section I explore the notion of optimal relevance and its relationship with considerations about the communicator's mental states. An utterance is optimally relevant on a given interpretation if it satisfies both clauses, (a) and (b), of the presumption of optimal relevance:

(2) **Presumption of optimal relevance**

[1] Wilson and Sperber (2004, p. 262) suggest that "a hearer's expectation of relevance may be more or less sophisticated" and that different expectations correspond to different interpretative strategies (i.e., 'naïve optimism', 'cautious optimism' and 'sophisticated understanding'). The discussion of these interpretative strategies goes beyond the scope of the present paper and will be marginally addressed only in the concluding section.

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.

Significantly, clause (b) brings into the picture the notions of '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* and *accidental irrelevance* (Wilson, 2000, p. 13). In the former case, the first interpretation that seems relevant enough to the addressee is not the intended one. In the latter, the information communicated is not relevant at all to the addressee. With the help of a few examples, I illustrate below how considerations about the communicator's abilities and preferences may guide the addressee towards the correct interpretation in both cases.

Let us start from two examples of 'accidental relevance', one motivated by the communicator's abilities (Carston, 2007), the other by the communicator's preferences (Mazzarella, 2011). Imagine the following scenario: Robyn is in one of her students' company. At some point during the conversation, the student, Sarah, addresses to Robyn the following utterance:

(4) Neil has broken his leg.

Suppose that Robyn knows two people called "Neil", her young son ($Neil_1$) and a colleague in the linguistic department ($Neil_2$). Suppose also that Robyn is so constantly worried about her son that when she hears (4) the first relevant interpretation to come to her mind is that $Neil_1$ has broken his leg. This interpretation is *accidentally relevant*: it is not (and could not) be intended by Sarah, who does not know that Robyn has a son called "Neil". The interpretative hypothesis that $Neil_1$ has broken his leg is not compatible with the communicator's abilities. For this reason, it is not selected as the output of a comprehension procedure driven by expectations of optimal relevance.

Now, modify this scenario by imagining that Robyn is speaking with her Italian student Sara, who is acquainted with Robyn's family, and that Robyn does not suffer from any maternal anxiety. When Sara utters (4), the first relevant interpretation to come to Robyn's mind is that $Neil_2$ has broken his leg. This interpretation is *accidentally relevant*: it is not intended by Sara, who adheres to the Italian social convention of referring to lecturers with formal titles (e.g., Dr Simpson). The interpretative hypothesis that $Neil_2$ has broken his leg is not compatible with the communicator's preferences and, as a consequence, it does not satisfy Robyn's expectations of optimal relevance.

To illustrate the notion of 'accidental irrelevance', we may consider cases in which the communicator mistakenly tells the addressee something that he already knows. Sperber and

Wilson (1986/1995, pp. 159-160) suggest the following example: Peter is passionate about Iris Murdoch's books and he usually buys them as soon as they are out. Mary, who knows about his passion, tells him:

(5) Iris Murdoch's new book is in the bookshops.

Accidentally, the first interpretation that comes to Peter's mind (i.e., *Jackson's Dilemma is in the bookshop*) is irrelevant to Peter: he knows that *Jackson's Dilemma* is available at the local bookshop (he has already bought a copy of it). However, considerations about the communicator's abilities (e.g., the assumption that Mary does not know that Peter has bought *Jackson's Dilemma*) prevent Peter from assessing further, more relevant, interpretative hypotheses (e.g., *that Iris Murdoch's new book – which he is not aware of – is in the bookshops*) and to attribute the intended interpretation to the communicator.

The three examples presented show that the expectations of relevance that determine the stopping point of the comprehension procedure are crucially constrained by considerations about the communicator's mental states (i.e., her 'abilities' and 'preferences'). These may ensure that unwanted interpretations are filtered out (as in (4)) or that correct interpretations that would be otherwise discarded are retained (as in (5)).

3 Optimal relevance at work: some criticisms

Mazzone (2009, 2013) has presented two sorts of objection against the role attributed by Relevance Theory to expectations of '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 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. I suggest that the first objection is unsound, while the second needs to be addressed by Relevance Theory.

3.1 *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 the two versions of example (4). The first interpretation that comes to mind, that is not only highly accessible but also sufficiently relevant, is discarded as incompatible with the speaker's mental states (her beliefs or desires). When such an incompatibility between the interpretative hypothesis and the communicator's 'abilities' or 'preferences' is detected, the former is dismissed and further hypotheses are tested.

This picture, which is certainly consistent with examples such as (4), seems to restrict 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 state, 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.[2] 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 the addressee may be well aware of the communicator's 'abilities' or 'preferences' (e.g., her beliefs on the topic under discussion). In these 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 constrained by it. This hypothesis would thus be 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.

3.2 Which cognitive mechanisms are involved?

Mazzone's second objection focuses on the cognitive underpinnings of the notion of 'optimal relevance', which – he argues – Relevance Theory fails to specify. According to Mazzone (2009, 2013), Relevance Theory does not provide an adequate account of the cognitive

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

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 '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 of 'optimal relevance' work as a pragmatic criterion of acceptability, but it does not offer a cognitively-specified description.

This objection raises an interesting issue, and the second part of this paper is devoted to an attempt to provide a solution to it. But before moving to this, I present 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". 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. In some circumstances, the comprehension procedure can be effect-driven: the occurrence of tightly constrained expectations about the intended implications (i.e., implicated conclusions) can affect the recovery of explicatures in such a way that the explicit content is constructed with the purpose of warranting the intended effects. Let us consider the following example from Wilson and Sperber (2004):

(6) *Peter*: Did John pay back the money he owed you?
Mary: 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. The logical form 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, relevance-theorists (Wilson & Carston, 2007; Carston, 2007) and Mazzarella (2011) 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 of 'optimal relevance'). However, I suggest two more fundamental reasons that explain why 'mutual parallel adjustment' does not (and cannot) do this. First, 'mutual parallel adjustment' is not treated as, nor was it ever intended to be, an acceptability criterion within the relevance-theoretic framework. Second, it can be argued that it does not (directly) involve assumptions about the communicator's mental states.

Interpretative hypotheses about the explicit and the implicit content are 'mutually adjusted' so that the derivation of the latter from the former is sound[3]. 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 Wilson and Sperber (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 (4), 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:

(4) 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 $Neil_1$).

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

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:

(4') Explicature: *Neil$_1$ has broken his leg.*
Implicated premises: *You should interrupt a meeting in case there is an emergency.*
The fact that Neil1 has broken his leg is an emergency.
Implicated conclusion: *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 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))[4]

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'.

I now turn to the investigation of the role played by information about the communicator's mental states in the rational reconstruction of pragmatic inferences, and in the process of mutual parallel adjustment. Mazzone (2013) claims that information about the communicator's mental states is injected into the process of mutual parallel adjustment as contextual assumptions. He suggests, then, that such an injection is not explained by Relevance Theory. I argue that Relevance Theory does not need to provide an explanation for this, precisely because this is not the case. Mazzone has misconstrued the role played by information about the communicator's mental states in the derivation of implicated conclusions (and in the mutual adjustment among premises and conclusion). I will call on some recent work by Mark Jary to support this argument.

Jary (2013) introduces a distinction between two types of implicature: material and behavioural. On the one hand, material implicatures are those implicated conclusions whose derivation can be rationally reconstructed without any appeal to premises concerning the speaker's verbal behaviour (e.g., The speaker has said that *p*) or the speaker's mental states. They can be reconstructed as following from the explicature of the utterance and its implicated premises. Behavioural implicatures, on the other hand, require both premises about the speaker's verbal behaviour and premises about her mental states (e.g., beliefs, desires, intentions).

On the basis of this distinction, we can investigate the different role played by information about the speaker's mental states in the derivation of material and behavioural

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

implicature. I show that none of these roles correspond to the role of intended contextual assumptions in the process of mutual adjustment attributed to it by Mazzone (2013). As a reminder, it is worth emphasising that mutual parallel adjustment is described by Relevance Theory as a process of reciprocal modulation between interpretative hypotheses concerning the explicatures of the utterance and its implicatures (i.e., its implicated premises and implicated conclusions). It does not involve any premise concerning the speaker's verbal behaviour (e.g., that the speaker has asserted a particular explicit content or uttered a particular sentence).

Jary (2013) suggests that information about the speaker's mental states does play a role in the derivation of material implicature but it does not figure as a premise in the derivation itself. Rather, information about the speaker's mental states may justify the selection of premises that are required to derive the implicature at issue[5]. As an example, consider (4). The rational reconstruction of the implicature that Robyn should leave her meeting can be represented as follows:

(4'') Explicature: *Neil$_2$ has broken his leg.*
Implicated premises: *You should interrupt a meeting in case there is an emergency.*
The fact that Neil$_2$ has broken his leg is an emergency.
Implicated conclusion: *Robyn should interrupt her meeting.*

In this case, the information that the speaker, Sarah, does not know Neil$_1$ but is acquainted with Neil$_2$, can be seen as justifying the selection of the first premise of the derivation (i.e., the explicature of the utterance).It follows that, even if information about the speaker's mental states (e.g., her beliefs) affects the derivation of material implicature, it does not play the role of contextual assumption assigned to it by Mazzone.

But what about behavioural implicature? As Jary (2013) states, behavioural implicatures do involve assumptions about the speaker's mental states in their rational reconstruction. To illustrate this with an example, let us consider the following Gricean example:

(7) Mr. X's command of English is excellent, and his attendance at tutorials has been regular.

In the context of a reference letter written for a philosophy job, (7) would be interpreted as implicitly communicating that the communicator thinks that Mr. X is a poor philosopher. Jary reconstructs this example along the following line:

(7') i. She has stated that Mr. X's command of English is excellent, and his attendance at tutorials has been regular.
ii. She has said nothing about Mr. X's merits as a philosopher.
iii. She knows that information about Mr. X's merits as a philosopher is what would be most relevant to my concerns.
iv. She is not opting out of the cooperative principle, for she has bothered to write.
v. Therefore there must be something she intends to communicate that she is

[5] Jary (2013) seems to confine this information to "in-built assumptions" concerning the fact that the speaker intends to convey something by her utterance and that the speaker intends to communicate the implications derived. I think this is a too restricted view of the kind of information about the speaker's mental states that affects the derivation of material implicature. I do not address this issue here, but my discussion of example (4) should shed some light on this.

unwilling to write down.

vi. This must be that Mr. X is a poor philosopher.

As this reconstruction clearly shows, considerations about the communicator's mental states (e.g., her beliefs, desires and intentions) act as premises for the derivation of the implicature at issue. However, the rational reconstruction of behavioural implicatures cannot be described in terms of a derivation from the explicit content of the utterance to its implicit content (since it requires a premise to the effect that the speaker has uttered that content). Thus, at least in its traditional sense, 'mutual parallel adjustment' does not apply to the derivation of behavioural implicatures.[6]

To sum up, Relevance Theory needs to provide an explanation of the cognitive mechanisms by which considerations about the communicator's mental states affect pragmatic interpretation. However, Mazzone's suggestion that such an explanation has to be found in the process of 'mutual parallel adjustment' described by Relevance Theory has been proven to be misguided for the following reasons: the inferential soundness guaranteed by 'mutual parallel adjustment' is a necessary but not sufficient condition for pragmatic acceptability, and information about the speaker's mental states does not affect pragmatic interpretation by figuring as implicated premises in the inference from the explicit to the implicit content of the utterance.

I argue that Relevance Theory can cope with the objection at issue by exploring a possibility that Mazzone (2013) has not taken into account and, consequently, ruled out. The investigation of such a tentative solution represents the main focus of the second part of this paper.

4 Epistemic vigilance and pragmatic interpretation

The line of solution explored in this paper appeals to a new area of research, pioneered by Sperber et al. (2010), which focuses on so-called 'epistemic vigilance'. Epistemic vigilance can be defined as alertness to the reliability of the source of information and to the believability of its content, as exercised by interlocutors in communicative settings.

My proposal relies on the hypothesised interaction between the relevance-guided comprehension procedure, on the one hand, and epistemic vigilance mechanisms, on the other. While the scope of this interaction has not been largely 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. This paper will mainly focus on the pragmatic perspective.

[6] This is Jary's view, based on the example of mutual parallel adjustment discussed so far in the relevance-theoretic literature, and I am following him on this for the purposes of the current paper. However, it has been drawn to my attention (Deirdre Wilson, personal communication) that Sperber and Wilson, in fact, intend that higher level explicatures (e.g., Mary stated that *p*, Mary believes that *p*) enter the mutual adjustment process in the same way as other explicatures do.

4.1 Epistemic vigilance: what it is and how it works

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.[7]

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., competent and trustworthy individuals).

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).

In the remaining part of this section, I explore the way in which epistemic vigilance is supposed to work and interact with the interpretative process. According to Sperber et al. (2010), epistemic vigilance mechanisms are activated by any piece of communicative behaviour. They work in parallel with those mechanisms involved in interpretation (e.g., the relevance-guided comprehension procedure, (3), within the relevance-theoretic framework) and assess the believability of the output of the interpretive process:

[7] 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.

> Comprehension involves adopting a tentative and labile stance of trust; this will lead to acceptance [i.e., believability (DM)] only if epistemic vigilance, which is triggered by the same communicative acts that trigger comprehension, does not come up with reasons to doubt. (Sperber et al., 2010, pp. 368-369)

If, during the interpretative process, the speaker is found to be unreliable by some epistemic vigilance mechanism, the interpreter will end up questioning the believability of the information communicated. Furthermore, if the interpretation delivered by the comprehension procedure is found to contradict assumptions strongly held by the interpreter, he might end up rejecting its content.[8]

4.2 Epistemic vigilance: an extended scope

The scope of the interaction hypothesised by Sperber et al. (2010) between the interpretative process, on the one hand, and epistemic vigilance, on the other, is relatively narrow. Both would be activated by the same communicative behaviour, but the only role of the epistemic vigilance system would be to assess the believability of the interpretation resulting from the comprehension process. In this paper I suggest an extension to this interaction (see Padilla Cruz (2012) for a different proposal along the same line).

My proposal is to extend the scope of this interaction to include not only the assessment of the believability of communicated information, but also the assessment of the acceptability of interpretative hypotheses. In other terms, epistemic vigilance mechanisms would be targeted at both the risk of misinformation and the *risk of misinterpretation*.

The terminological and conceptual distinction between the two notions of 'believability' and 'acceptability' is crucial to understand this suggestion. On the one hand, the notion of 'believability' concerns the extent to which an interpretation attributed to the communicator (i.e., the output of the interpreter's comprehension procedure) is allowed to enter the 'belief box' of the interpreter. The issue here is whether or not the interpreter ends up believing it or not. The notion of 'acceptability', on the other hand, concerns whether an interpretative hypothesis about the speaker's meaning is retained and attributed to the speaker as the intended interpretation. The issue here is whether or not an interpretative hypothesis ends up being the output of the comprehension procedure. It follows that the acceptability issue clearly precedes the believability issue: the interpreter needs to know what the intended interpretation is before he can decide whether to believe it or not.

In section 2, I introduced the relevance-theoretic framework, according to which the interpretative process follows a dedicated inferential procedure, the relevance-guided comprehension procedure. Its stopping point is determined by the expectations of relevance of the interpreter, which generally coincide with expectations of optimal relevance. These are tightly constrained by considerations about the communicator's mental states, i.e., her *abilities* and *preferences*. In section 3, I presented some objections raised by Mazzone (2009, 2013) against this framework. The main one was related to the following question: *how* do considerations about the communicator's mental states affect pragmatic interpretation? I now put forth my tentative answer: epistemic vigilance mechanisms, which assess the reliability of the source of information (i.e., the speaker), recruit information about her abilities and

[8] There are a variety of epistemic attitudes that might be yielded by epistemic vigilance: acceptance, belief, doubt, rejection, among others. In what follows I do not distinguish between acceptance and belief and I refer to the epistemic attitude of belief only. This is to avoid confusion between the 'acceptability' of communicated contents and the 'acceptability' of interpretative hypotheses (as defined in section 1).

preferences. Once recruited, this information interacts with the relevance-guided comprehension procedure and constrains the choice of its output.[9]

The interaction between the comprehension process and epistemic vigilance mechanisms seems to find support in the intuitive link between the notions of speaker's abilities and preferences (which are integral to the definition of 'optimal relevance'), on the one hand, and the notions of competence and benevolence, on the other. Since epistemic vigilance mechanisms targeted at the speaker's reliability check the speaker's competence and benevolence, it seems natural to assume that the very same mechanisms will contribute to the assessment of the speaker's abilities and preferences. I do not claim, however, that the two pairs of notions coincide with each other, that is, that 'abilities' correspond to 'competence', and 'preferences' to benevolence'. Indeed, there are reasons to doubt that such a parallelism stands: preferences, for instance, goes beyond the intention to share genuine information (i.e., benevolence), including other kinds of goals (e.g., compliance with social conventions). However, these mismatches should not obscure many important similarities. For the time being, I focus on their similarities and explore their implications for the hypothesised interaction between epistemic vigilance and pragmatic comprehension.

The notion of 'competence' sketched by Sperber et al. (2010) seems to capture (at least some of) the speaker's abilities that modulate the interpreter's expectations of relevance. On the one hand, Sperber et al. (2010) define a competent communicator as one who possesses genuine information, as opposed to misinformation or no information. In other terms, a competent communicator can be said to possess true beliefs, as opposed to false beliefs or no beliefs. On the other hand, the communicator's abilities are defined so as to include linguistic competence and broader epistemic states such as belief about or knowledge of the world. Both the communicator's 'competence' and (part of) the communicator's 'abilities' share the same epistemic characterisation (i.e., possession of true/false/no beliefs). This similarity should be clear by considering examples of pragmatic interpretation affected by considerations of the communicator' abilities. Let us consider again the following sentence:

(4) Neil has broken his leg.

uttered in a context in which the addressee (i.e., Robyn) knows two people called "Neil", $Neil_1$ and $Neil_2$, but the speaker (i.e., Sarah) knows just one of them, $Neil_2$. Assuming that the first accessible interpretation to come to Robyn's mind is that $Neil_1$, her son, has broken his leg, she would be able to reach the intended interpretation by (i) recognising that Sarah does not know that Robyn has a son called Neil and that, for this reason, she did not (and could not) intend to refer to him; (ii) by assessing less accessible interpretations, such as $Neil_2$ has broken his leg, and (iii) by stopping when her expectations of optimal relevance are satisfied.

According to Relevance Theory, the expectation of optimal relevance that drives the comprehension procedure would not be satisfied by the interpretation $NEIL_1$ HAS BROKEN HIS LEG because this is not compatible with the speaker's abilities: because of her abilities, the speaker could not have expected this interpretation to be relevant enough to the hearer (or even to have been accessed by the hearer).

The same scenario, however, can be aptly described by appealing to the notion of speaker's competence. If the interpreter, Robyn, recognises that Sarah *is not a fully competent speaker* on this topic (she does not know that Robyn has a son called Neil), then she will be able to discard the first interpretation that comes to her mind (i.e., $Neil_1$ has broken his leg).

[9] Information about mental states is assumed in this framework to be output by a dedicated mind-reading module, which can provide input to both comprehension and epistemic vigilance mechanisms (and can be called on by either).

This is a case in which recognising that the communicator possesses *no information* (about a particular topic), plays a crucial role in modulating the interpretative process.

In the next section, I explore epistemic vigilance mechanisms targeted at assessing the speaker's competence and I illustrate in more details how they interact with the interpretative process.

4.3 The communicator's competence

The definition of 'competence' provided by Sperber et al. (2010) is intrinsically context-dependent; and it could not be otherwise. 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 was 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.

This suggests the existence of epistemic vigilance mechanisms that can assess competence in a context-sensitive way, rather than simply relying on general impressions of competence and trustworthiness. The investigation of these mechanisms will prove to be crucial for a general understanding of epistemic vigilance, and its interaction with the comprehension module:

> In order to gain a better grasp of the mechanisms for epistemic vigilance towards the source, what is most urgently needed is not more empirical work on lie detection or general judgements of trustworthiness, but research on how trust and mistrust are calibrated to the situation, the interlocutors and the topic of communication. Here two distinct types of consideration should be taken into account: *the communicator's competence on the topic of her assertion*, and her motivation for communicating. (Sperber et al., 2010, pp. 370-371, *my emphasis* (DM)).

In what follows, I suggest a way in which epistemic vigilance mechanisms assessing the communicator's competence on the topic of conversation may affect the interpretative process. Once again, let us focus on example (4):

(4) Neil has broken his leg.

Sarah's utterance is a piece of communicative behaviour that triggers the parallel activation (in the addressee's mind) of the interpretative process (i.e., the relevance-guided comprehension procedure), on the one hand, and epistemic vigilance mechanisms, on the other. Among these, there are those mechanisms targeted at assessing the speaker's competence on the topic of conversation.

According to Relevance Theory, the addressee (i.e., Robyn) follows a path of least effort in assessing interpretative hypotheses, that is, interpretative hypothesis are tested in order of accessibility. In the scenario described above, among the possible candidates for the semantic value of "Neil" (i.e., NEIL_1, her son, and NEIL_2, her colleague), NEIL_1 is the most highly activated. As a consequence, the interpretative hypothesis that Neil_1 has broken his leg is the first to be accessed and assessed.

I suggest 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 $Neil_1$ has broken his leg provides a hypothesised topic of conversation (i.e., $Neil_1$) with regard to which epistemic vigilance mechanisms assess Sarah's competence. These mechanisms access the piece of information that Sarah does not know that Robyn has a son called "Neil".

As a consequence, an incompatibility between the interpretative hypothesis and the information that Sarah does not know that Robyn has a son called "Neil" is detected. This incompatibility can be described in terms of a conflict between the interpretative hypothesis and the speaker's abilities. Since the interpretative hypothesis that $Neil_1$ has broken his leg goes beyond the speaker's abilities, it is not optimally relevant and it does not satisfy the interpreter's expectation of relevance. It is thus discarded. This prompts the relevance-guided comprehension procedure to go further and assess less accessible interpretations.

The next interpretative hypothesis to be tested is that $Neil_2$ (i.e., Robyn's colleague) has broken his leg. Once again, accessing this interpretative hypothesis provides a hypothesised topic of conversation (i.e., $Neil_2$). The competence of the speaker with regard to the topic at issue is assessed by epistemic vigilance mechanisms. Since they do not detect any incompatibility between the interpretative hypothesis that $Neil_2$ has broken his leg and Sarah's abilities, the interpretative hypothesis is retained and attributed to the speaker as the intended interpretation.

This analysis of example (4) provides a more fine-grained picture of the interaction between the interpretative process and epistemic vigilance mechanisms. In particular, it suggests an answer to the challenging question about how considerations concerning the speaker's abilities and preference can affect the relevance-guided comprehension procedure (Mazzone, 2009, 2013). My tentative answer assigns a significant role to epistemic vigilance mechanisms geared to assessing the speaker's reliability. In the example at issue, the interpreter may reach the intended interpretation by monitoring the speaker's competence on the topic of conversation. Specifically, epistemic vigilance mechanisms may prompt the relevance-guided comprehension procedure to assess further interpretative hypothesis when the current one is incompatible with (what the interpreter takes to be) the speaker's system of beliefs.

4.4 The speaker's preferences

In this section I try to generalise the picture sketched above to include not only speaker's abilities but also speaker's *preferences*. The notion of 'preferences' is far more elusive than that of 'abilities', and it is not – strictly speaking – defined by Relevance Theory. It comprises a wide range of goals which are distinct from the fundamental communicative goal (Carston, 2005): compliance with social conventions such as "rules of etiquette or standards of ideological correctness" (Sperber & Wilson, 1986/1995, p. 268), linguistic preferences, such as preferences for formal or indirect modes of expressions, but also the desire to impress the interlocutor with learned vocabulary, and preferences concerning the kind of information to be shared with the interlocutor: "a speaker might choose not to say something, which could be for any number of reasons, including embarrassment, an intention to deceive, or an unwillingness to share particular information" (Clark, 2013, p. 111). This (incomplete) list reveals the heterogeneity of the range of preferences that are included under the same label and that a cognitive account of the notion of optimal relevance should consider.

A first tentative solution could be to appeal to the notion of 'benevolence'. As speaker's abilities are monitored by epistemic vigilance mechanisms targeted at the communicator's competence, speaker's preferences may be monitored by epistemic vigilance mechanisms targeted at the communicator's benevolence. While this suggestion may be apt for some preferences, it is not difficult to see that it can hardly cover the complete set of preferences

sketched above. In particular, it seems to easily apply to preferences which concern the kind and amount of information that the speaker is willing to share with her interlocutors, but not to linguistic or social preferences. Sperber et al. (2010) define the notion of benevolence as the willingness to share genuine information with the interlocutor (as opposed to making an assertion that the speaker does not regard as true, through either indifference or malevolence). This notion is certainly linked to the intention to deceive the interlocutor and the unwillingness to share particular information that Clark (2013) lists among the speakers' 'preferences'.

I adopt this tentative (and partial) solution while introducing a significant change: the notion of 'preference' cannot include, *pace* Clark, the intention to deceive the interlocutor. The reason is that this would contradict the Communicative Principle of Relevance, that is, that every utterance comes with a presumption of optimal relevance. Let us start from this remark by Sperber and Wilson (1986/1995, p. 271, *my emphasis*): "We claim that a presumption of optimal relevance is *communicated* by any act of ostensive communication" This means that by the very act of uttering something, the communicator intends to make manifest both clauses of the 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.

If clause (b) of the presumption is intentionally made manifest by every act of ostensive communication, the notion of preferences cannot include the intention to deceive the interlocutor. This would, in fact, contradict the assumption that the communicator is a rational agent: a rational communicator must intend her utterance to appear relevant enough to the interlocutor to attract his attention and make him willing to spend the effort required for comprehension. How could a rational communicator expect the addressee to be willing to invest this effort if she communicates that her ostensive stimulus may be produced with the intention of deceiving the addressee?

In light of these considerations, I suggest that epistemic vigilance mechanisms targeted at the communicator's benevolence may play a role in monitoring only a very limited range of speaker preferences, that is, those preferences that are related to the desire to withhold some information from the interlocutor when her motivation does not entail deceptive intentions (e.g., when some information is highly relevant but incriminating). This, however, leaves the following question totally unaddressed: what about linguistic and social preferences?[10]

In the remaining part of this section, I offer some speculations about the possibility that some dedicated vigilance mechanisms may have evolved to monitor the speaker's social preferences. In the field of evolutionary psychology, the idea that humans have evolved a "constellation of cognitive mechanisms for social life" has been strongly defended since the seminal work of Cosmides and Tooby (1992). These mechanisms are supposed to guide thought and behaviour in order to enable us to deal with recurrent problems posed by our social world:

[10] It is worth noticing that most of the linguistic preferences, if not idiosyncratic, have some social motivation. For instance, communicators may prefer a roundabout way of speaking when there is a risk of offending the interlocutor. Linguistic and social preferences should be considered as distinct, but highly overlapping, sets of preferences.

> To behave adaptively, they [i.e., our ancestors (DM)] not only needed to construct a spatial map of the objects disclosed to them by their retinas, but a social map of the persons, relationship, motives, interactions, emotions, and intentions that made up their social world. (Cosmides & Tooby, 1992, p. 163).

Most of the factors enumerated in this passage have undergone radical transformations with the development of more and more sophisticated forms of society. The complexity of the emerging societies and the new variety of relationships among their members has progressively changed our experience of the social world. In particular, social conventions have significantly shaped and constrained the ways in which we interact with each other and they have often defined our identity as part of a group. For instance, more sophisticated societies are generally characterised by the development of several hierarchical structures of increased complexity. Within each structure, the relationships among different members are shaped by the place each of them occupies. Different social conventions may govern the relationships between one member of the hierarchy and the individuals that occupy higher or lower positions in the same structure. Interestingly, the very possibility of being a member of a social structure is often conditional on respecting its social conventions. Conversely, breaking a social convention may lead to the risk of being excluded from the group.

It seems plausible, then, to speculate that the ever increasing importance of social conventions and, consequently, of the risk derived from disrespecting them, has prompted the development of dedicated mechanisms targeted at this risk. These mechanisms would monitor the social conventions at issue in a particular context and guide our behaviour in conformity with such conventions. In particular, we may find dedicated mechanisms guiding our verbal behaviour under these social constraints. These would be targeted at the risk of misinterpretation derived by ignoring the social preferences displayed by our interlocutors.

With this picture in mind, let us consider again example (4) in its preference-based interpretation form, in which Robyn is speaking with an Italian student who is acquainted with her family.

(4) Neil has broken his leg.

In this case, Robyn is able to reach the intended interpretation (i.e., $Neil_1$, her son, has broken his leg) because the most accessible interpretation that $Neil_2$, her colleague, has broken his leg is not compatible with the speaker's linguistic and social preferences. Recognising that the speaker generally obeys one of the most ingrained social conventions in the Italian academic context, that is, the convention of referring to lecturers by their formal title, Robyn would correctly identify the student's intention to refer to her son with the expression "Neil".

These speculations, bold as they may be, can be seen as a tentative attempt to mirror the evolutionary account that Sperber et al. (2010) provide for the emergence of epistemic vigilance.

4.5 Is epistemic vigilance (only) a 'filter'?

So far, I have explored the idea that epistemic vigilance affects the interpretative process by acting as a filter on unintended interpretation. In particular, epistemic vigilance mechanisms filter out interpretative hypotheses that are not compatible with the speaker's abilities and preferences (and that, as a consequence, are not *optimally* relevant). They allow the interpreter to dismiss otherwise relevant interpretations that the speakers would have not been willing or able to convey.

This picture, however, may not exhaustively describe how epistemic vigilance affects the interpretative process. The tentative hypothesis that I would like to suggest is that epistemic vigilance mechanisms may play a role not only in the assessment but also in the *construction* of interpretative hypotheses.

In section 3.1 I suggested that in some circumstances the interpreter cannot help but take account of information about the speaker's epistemic state on his first processing pass through the utterance. Now, what bearing might this have on the operation of epistemic vigilance mechanisms?

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 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. I cannot explore this further here, but I point to the phenomenon of 'scalar implicatures' as a fertile ground for the application of this idea. For instance, 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. 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. Can epistemic vigilance be overwhelmed?

5.1 The cost of epistemic vigilance

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 abilities and preferences illustrated above may require a great deal of processing

effort. This opens 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 paper.

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, Simmons and Todorov (2005), who experimentally showed that increasing the relevance of a piece of communicated information modulates its believability. Hasson et al. ran a modified version of Daniel Gilbert's experiments on automatic belief of communicated information (Gilbert, Krull & Malone, 1990; Gilbert, Tafarodi & Malone, 1993). Gilbert's experiments were intended to show that communicated information is automatically assumed to be true (i.e., automatically enters our 'belief box') before being examined and possibly rejected. The participants were presented with sentences about the meaning of Hopi words such as "A Monishna is a star", followed by the signals "True" or "False" to indicate their truth-values. In a subsequent recognition task participants had to assign a truth-value to the sentences presented to them (recollecting the truth-value signal associated with each sentence in the previous task). In the critical condition, some of the truth-value signals were produced while participants were distracted (by being required to respond as quickly as possible to the sound of a tone). Gilbert et al. predicted that if participants had automatically accepted the sentences, the distraction accompanying "False" signals would have been likely to affect the acceptance rate of false statements, leading participants to remember false statements as true. This is what Gilbert et al. (1990, 1993) found.

Hasson et al. (2005) repeated Gilbert's experiments modifying the relevance of the statements presented to the participants. While a statement such as "A Monishna is star" is unlikely to be relevant to the participants in the experiment, statements whose falsity carries stereotypical implications (e.g., "George owns a television" may carry the implications that he is atypical, he is the bookish type, etc.) or contradicts strongly held beliefs are generally much more relevant to participants. By modifying the material along the relevance dimension, Hasson et al. (2005) did not find automatic acceptance of false statements under cognitive load.

Hasson et al.'s results seem to support the idea that 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 go back to the scenario described in 3.2. 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:

(4) Neil has broken his leg.

As 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.[11]

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 generalizable 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)).

5.2 Relevance and epistemic vigilance

[11] 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 paper focuses on, that it, the role that they play through epistemic vigilance mechanisms in *on-line* pragmatic interpretation.

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 abilities and preferences. The rationale behind this is that speakers cannot be expected to go beyond their abilities or against their interests. I suggested a way to cognitively implement this constraint on utterance interpretation by appealing to the interaction between the comprehension procedure and epistemic vigilance mechanisms. The core idea is that an interpretative hypothesis is tested against the information retrieved by epistemic vigilance 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), monitored by epistemic vigilance.

At this point a subtle but crucial remark is needed: the set of information against which an interpretative hypotheses 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.

I start by introducing an example from Sperber et al. (2010, p. 368) aimed at showing that interpretation is guided by an expectation of relevance, rather than by a presumption of truth. Then, I modify the example at issue in order to show how the interpreter's expectations of relevance can overwhelm epistemic vigilance mechanisms.

Imagine that Barbara has asked Joan to buy a bottle of champagne for her birthday party. After reporting this to Andy, the following exchange takes place:

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

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 sensible interpretative hypothesis 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.[12]

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.

6 Conclusions

The inferential model of utterance interpretation proposed by Relevance Theory assigns a significant role to considerations about the speaker's mental states. The interpreter follows a relevance-guided comprehension procedure, (3), that is driven by expectations of relevance and he stops when his expectations are satisfied. These expectations generally coincide with a presumption of *optimal* relevance: the communicator is expected to try to achieve the highest level of relevance that is compatible with her *abilities* and *preferences*. When an

[12] 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.

interpretative hypothesis satisfies the interpreter's expectations of optimal relevance (i.e., when it satisfies the acceptability criterion proposed by Relevance Theory), it is retained and attributed to the speaker as the intended interpretation.

Within this picture, the assessment of the acceptability of interpretative hypotheses involves the active monitoring of the speaker's abilities and preferences. This corresponds to a claim that the speaker's mental states (e.g., her beliefs and desires) need to be taken into consideration in order to arrive at the intended interpretation. Mazzone (2009, 2013) has argued that Relevance Theory does not offer an adequate account of how this active monitoring of the speaker's mental states is supposed to work. How is this information recruited? How does it affect the interpretative process? These questions – Mazzone argues – have not received enough attention within the relevance-theoretic framework and are in need of answers.

This paper represents an attempt to take on Mazzone's challenge and to implement this aspect of Relevance Theory. I suggest that the answer to Mazzone's questions is to be found by looking at the interaction between the relevance-guided comprehension procedure and epistemic vigilance mechanisms. 'Epistemic vigilance' has been described by Sperber et al. (2010) as the critical alertness to the risk of being misinformed by interlocutors. It subsumes several cognitive mechanisms targeted at the assessment of the speaker's reliability (i.e., her competence and benevolence) and at the believability of the communicated content. These mechanisms modulate the addressee's epistemic trust and may prevent him from believing a piece of communicated information.

I have proposed that the scope of epistemic vigilance be extended from the risk of misinformation to the risk of misinterpretation. Not only do epistemic vigilance mechanisms affect the believability of a piece of communicated information, but they also contribute to the assessment of the acceptability of an interpretative hypothesis. On the one hand, epistemic vigilance mechanisms targeted at assessing the communicator's reliability may contribute to filter out interpretative hypotheses that are not compatible with the speaker's abilities and preferences (e.g., cases of accidental relevance). On the other hand, they may retain interpretative hypotheses that are irrelevant to the interpreter but compatible with the speaker's abilities and preferences (e.g., cases of accidental irrelevance).

To conclude, I sketch some directions for future research. Sperber (1994) has proposed that the relevance-guided comprehension procedure can be driven by more or less sophisticated expectations of relevance. He suggests the existence of three different versions of the interpretative strategy: 'naïve optimism', 'cautious optimism' and 'sophisticated understanding'. Interestingly, the first strategy, naïve optimism, is characterised by the assumption that the communicator is both competent and benevolent, whereas cautious optimism and sophisticated understanding drop, respectively, the assumption of competence and the assumption of benevolence of the communicator. Padilla Cruz (2012) has proposed to consider epistemic vigilance as the trigger for a shift in interpretative strategies. For instance, if epistemic vigilance detects that the interlocutor is not a very competent language user, it may trigger a shift from naïve optimism to cautious optimism. I believe that the relationship between epistemic vigilance and the three interpretative strategies suggested by Sperber (1994) deserves further investigation. Future research should address the following question: are Sperber's (1994) interpretative strategies encompassed by the more recent work on epistemic vigilance? The way in which the work on epistemic vigilance by Sperber et al. (2010) reflects on previous work within the relevance-theoretic perspective is far from definitely settled. I believe that once epistemic vigilance is brought into the picture, the three interpretative strategies may be 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.

Finally, the relationship between Sperber's (1994) interpretative strategies and epistemic vigilance may be fruitfully explored within a developmental perspective. Naïve optimism, cautious optimism and sophisticated understanding are said to correspond to different developmental stages. Naïve optimism would be adopted in early childhood, while cautious optimism and sophisticated understanding would emerge later on. If these three developmental stages are underpinned by the development of different epistemic vigilance mechanisms (e.g., mechanisms assessing the communicator's competence for the cautious optimistic stage and mechanisms assessing the communicator's benevolence for the sophisticated understanding stage), the developmental trajectory followed by epistemic vigilance should map onto the development of pragmatic competence. While different studies have addressed the former (e.g., Clément, Koenig and Harris (2004), Mascaro and Sperber (2009)), an explicit comparison between these two developmental trajectories still remains to be carried out.

References

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

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

Carston, R. (2005). A note on pragmatic principles of least effort. *UCL Working Papers in Linguistics, 17*, 271-278.

Carston, R. (2007). How many pragmatic systems are there? In M. J. Frapolli (Ed.), S*aying, meaning, referring. Essays on the philosophy of François Recanati* (pp. 18-48). New York: Palgrave Macmillan.

Clark, B. (2013). *Relevance Theory*. Cambridge: Cambridge University Press.

Clément, F., Koenig, M. A., & Harris, P. (2004). The ontogeny of trust. *Mind & Language, 19*, 360-379.

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

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*, 601–13.

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*, 221–33.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: on the possibility of suspending belief. *Psychological Science, 16*, 566–71.

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

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

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

Mazzarella, D. (2011). Accessibility and relevance: A fork in the road. *UCL Working papers in Linguistics, 23*, 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*, 2148-2159.

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

Padilla Cruz, M. (2012). Epistemic vigilance, cautious optimism and sophisticated understanding. *Research in Language, 10*(4), 365-386.

Recanati, F. (2002). Does linguistic communication rest on inference? *Mind & Language, 17*(1&2), 105-126.

Recanati, F. (2004). *Literal meaning*. Cambridge: Cambridge University Press.

Recanati, F. (2007). 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.

Sperber, D. (1994). Understanding verbal understanding. In J. Khalfa (Ed.), *What is intelligence?* (pp. 179-198). Cambridge: Cambridge University Press.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and Cognition* (2nd ed.). Oxford: Blackwell.

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., Clément, F., Heintz, C., Mascaro, O., Mercier, U., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 24*(4), 359-393.

Wilson, D. (2000). Metarepresentation in linguistic communication. In D. Sperber (Ed.), *Metarepresentation: A multidisciplinary perspective* (pp. 411-448). Oxford: Oxford University Press.

Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (Palgrave Advances in Linguistics) (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. (2002). Truthfulness and relevance. *Mind, 111*, 583-632.

Wilson, D., & Sperber, D. (2004). Relevance Theory. In L. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 607-632). Oxford: Blackwell.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge: Cambridge University Press.