CROATIAN JOURNAL OF PHILOSOPHY

Vol. XV · No. 44 · 2015

Articles

Book Review

Introduction

A day of the conference on the Philosophy of Language and Linguistics *held at the Interuniversity center (IUC) in Dubrovnik, September 7th to 11th, 2014, was dedicated to Dan Sperber and Deirdre Wilson's work. This issue of the* Croatian Journal of Philosophy *publishes most of the papers given at this conference.*

Sperber and Wilson *presented an introductory paper on their work which here bears the title "Beyond Speaker's Meaning" in which, after considering some of the difficulties raised by Grice's three-clause definition of speaker's meaning, they argue that constructing an adequate theory of communication involves going beyond Grice's notion of speaker's meaning. Thus they argue that the characterisation of ostensive communication introduced in relevance theory can provide a conceptually unified explanation of a much wider range of communicative acts than Grice was concerned with.*

Anne Bezuidenhout *in her contribution entitled "Cognitive Environments and Conversational Tailoring" explores the psychological notion of context as cognitive environment (CE) that is part of the Relevance Theory (RT) framework and describes the way in which such CEs are constrained during the course of conversation as the conversational partners engage in "conversional tailoring".*

Zsófia Zvolenszky *in her paper "Inferring Content: Metaphor and Malaproprism" looks into Sperber and Wilson's reasons for holding the view that metaphorical utterances occupy one end of a continuum that includes literal, loose and hyperbolic utterances with no sharp boundaries in between them. She labels this the* continuum argument about interpreting metaphors. *Zvolenszky aims to show that this continuum argument doesn't work. For if it were to work, it would have an unwanted consequence: it could be converted into a continuum argument about interpreting linguistic errors, including slips of the tongue, of which malaprops are a special case.*

Diana Mazzarella *in her paper "Pragmatics and Epistemic Vigilance: The Deployment of Sophisticated Interpretative Strategies" looks into Sperber and Wilson's suggestion that competent hearers can deploy sophisticated interpretative strategies in order to cope with deliberate deception or to avoid misunderstandings due to speaker's incompetence. She investigates the cognitive underpinnings of sophisticated interpretative strategies and suggests that they emerge from the interaction between a relevance-guided comprehension procedure and epistemic vigilance mechanisms.*

Nenad Miščević *in his contribution "Pejorative and Relevance: Synchronic and Diachronic Issues" considers a possible relevantist treatment, in the spirit of Wilson and Sperber's work, of pejoratives and argues for several claims concerning them: 1. The negative content of pejorative is the normal part of their lexical meaning. 2. He argues for an evaluative-content approach for the relevantist, in contrast to its neutral-content alternative. 3. He sees an interesting parallel between the echoing-cum-reversal processes Wilson and Sperber propose for irony and the repeating-and-reversing process typical of appropriation of pejoratives. 4. A brief application of the relevantist understanding of metaphor is proposed as a tool for understanding the genealogy of pejoratives of figurative origin.*

Nenad Smokrović *in his paper under the title "Argumentation as a means for Extending Knowledge" bases his claim on two focal points: 1. Reasoning is designed for argumentation, and 2. Argumentation process is an exceptionally successful media that provokes usage of methods reliable for the extension of knowledge. He relies on Sperber and Mercier's evolutionary psychological approach to argumentation. Taking this ground as a departing point, the goal of the paper is to broaden that particular approach with epistemological insights based on Williamson's safety theory of knowledge.*

The last paper by Dunja Jutronić *is a contribution of a sociolinguist based on her own field research. Her paper under the title of "Cognitive Pragmatics and Variational Pragmatics: Possible Interaction?" looks into a possible way in which cognitive pragmatics can help out variational studies in explaining the processes of language change. After broadly setting the scene this article proceeds by giving basic information about variational pragmatics and then concentrates on Sperber and Wilson's relevance theory and its possible interaction with social sciences, namely its application in sociolinguistics. The hope is that such discussions can bring closer cognitivists, i. e. relevantists, to sociolinguists, i. e. variationists.*

The last two papers by Nenad Smokrović and Dunja Jutronić were not presented at the above mentioned conference but were given at Dan Sperber's symposium held in March 18th 2013 in Rijeka. The symposium was organized by the Philosophy Department in Rijeka together with the Croatian Society for Analytic Philosophy. We thought it appropriate to include it in this selection since they also discuss Sperber and Wilson's work. We are grateful for Dan Sperber's presence at Rijeka symposium and both Sperber and Wilson's presence at Dubrovnik conference. This issue is dedicated to them both.

Rijeka, September 2015. Dunja Jutronić

Croatian Journal of Philosophy
Vol. XV, No. 44, 2015

Beyond Speaker's Meaning

DAN SPERBER
Central European University, Budapest, Hungary
Centre National de la Recherche Scientifique, Paris, France

DEIRDRE WILSON
University College London, London, United Kingdom
Centre for the Study of Mind in Nature, Oslo, Norway

Our main aim in this paper is to show that constructing an adequate theory of communication involves going beyond Grice's notion of speaker's meaning. After considering some of the diffi culties raised by Grice's three-clause defi nition of speaker's meaning, we argue that the characterisation of ostensive communication introduced in relevance theory can provide a conceptually unifi ed explanation of a much wider range of communicative acts than Grice was concerned with, including cases of both 'showing that' and 'telling that', and with both determinate and indeterminate import.

Keywords: Communication, showing vs telling, paraphrasability, manifestness, ostension.

1. *Introduction*

In *Relevance: Communication and Cognition* (Sperber and Wilson 1986; revised edition 1995) we put forward a number of novel ideas, several of which have been influential and others more controversial. However, there is one idea that we feel did not get the discussion it deserved. We proposed a characterisation of communication which, although inspired by Grice's definition of speaker's meaning, implied that speaker's meaning does not have the degree of unity or autonomy needed to make it the proper object of a philosophical definition or a scientific theory. Communication, on the other hand, or rather the kind of 'ostensive' communication that humans engage in, is such a proper object of inquiry. We argued that our account of communication does a better job of explaining how utterances are interpreted than a standard Gricean approach, and also makes good sense of our fuzzy intuitions about speaker's meaning without giving the notion an unduly important theoretical role. Here, we take up the issues again. In the first part of the paper, we discuss some difficulties with Grice's notion of speaker's meaning, and in the second part, we consider how to resolve them.

2. *Difficulties with Grice's notion of speaker's meaning*

2.1 *The continuum between meaning and showing*

Grice was aware of two particular problems with the notion of speaker's meaning. The first, linked to his sharp distinction between natural and non-natural meaning, arises when one tries to separate 'meaning that' from 'displaying direct evidence that'[1] in cases like the following (Grice 1989: 109):

> (a) Herod, showing Salome the head of St. John the Baptist, cannot, I think, be said to have meant that St. John the Baptist was dead.
> (b) Displaying a bandaged leg (in response to a squash invitation)
> In (b) the displayer could mean (1) that he cannot play squash
> Or (dubiously) (2) that he had a bad leg (the bandage might be fake)
> But not (3) that the leg is bandaged.

In discussing case (a), Grice comments:

> Herod intended to make Salome believe that St. John the Baptist was dead and no doubt also intended Salome to recognise that he intended her to believe that St. John the Baptist was dead… Yet I certainly do not think that we should want to say that we have here [a case] of meaning$_{NN}$ … What we want to find is the difference between, for example, 'deliberately and openly letting someone know' and 'telling' ... (1989: 218).

Grice's solution was to add a third clause to his definition of utterer's meaning. In order to mean something by an utterance, the utterer must intend the addressee

> (1) to produce a particular response *r*
> (2) to think (recognise) that the utterer intends (1)
> (3) to fulfil (1) on the basis of his fulfilment of (2),

where (3) is understood as stipulating that the addressee's recognition of the utterer's intention in (1) must be "at least part of his reason for producing r, and not merely the cause of his producing *r*" (Grice 1989: 92). Despite some debate in the literature about whether this third clause was needed (Schiffer 1972; Vlach 1981; Recanati 1986; Bach 1987; Neale 1992; Wharton 2009), it remained central to Grice's later discussions of meaning and his distinction between natural and non-natural meaning (Grice 1989: 290–97, 349–59).

[1] Of course, if a reliable speaker both says that *P* and means it, this is evidence that *P*. However, it is indirect evidence (in the sense in which we want to make the direct/indirect distinction here) in that the content of the evidence depends on the interpretation of the communicator's meaning, and the force of the evidence depends on the communicator's trustworthiness.

In characterising ostensive communication, we built on the first two clauses of Grice's definition and dropped the third. This was not because we were willing to broaden the definition of utterer's meaning—we agreed with Grice that talk of 'meaning' is awkward in certain cases—but because it seemed obvious that there is a continuum of cases between 'meaning that' (typically achieved by the use of language) and 'displaying evidence that' (in other words, showing), and we wanted our account of communication to cover both (Sperber and Wilson 1986: 46–54). Grice's example of the bandaged leg suggests how the continuum of cases can be constructed. The communicator wants the addressee to come to believe that *P*. In pure cases of showing, as in (b3), what is being shown provides sufficient evidence for the addressee to believe that *P*, and the fact that the communicator intended him to believe that *P* does not even strengthen that evidence. In pure cases of meaning, as in an ordinary linguistic assertion, all the evidence that *P* is provided by the communicator's giving overt evidence of her intention that the addressee should believe that *P* (and it is good evidence provided that the addressee trusts her competence and honesty in the matter). But of course, the evidence can come both from whatever is displayed (either shown or uttered) and from what the communicator's communicative behaviour indicates of her intention, as in (b1) and (b2).

When a piece of evidence is shown to an addressee, it is typically interpreted in the light of the fact that it is being shown. In Grice's example (b), the bandage on the leg may in itself be only weak evidence that the communicator cannot play squash: as he puts it, the bandage may be "fake", and the condition it covers may be quite compatible with playing squash. However, the fact that it is being shown in answer to the question suggests that the condition it conceals makes playing squash impossible, or at least undesirable. Thus, (b1) is neither a pure case of meaning nor a pure case of producing direct evidence. Make the evidence stronger, say by showing a cast instead of a bandage, and the addressee will arrive at the intended conclusion mostly, if not wholly, on the ground of what is shown.

We described (b3) above as a pure case of showing. Could showing a bandaged leg ever 'mean' that the communicator has a bandaged leg (as some commentators have been willing to accept)? Suppose the showing was in response to the question, "Is your leg bandaged?", how does this differ from simply answering "Yes"? One difference is that the addressee has to trust the communicator in the 'yes' case and not in the showing case. Suppose, then, that the communicator pulls up her long skirt just enough to show what could be the bottom of a large bandage, giving weak, inconclusive evidence that her leg is bandaged. Since some trust is needed to accept the intended conclusion, would this now be a case of meaning? There is, of course, a continuum of positions to which the communicator could pull up her skirt, exposing a little more of the bandage each time, until the fact that the leg is bandaged is perceptually beyond doubt. At each point, less trust would be needed, exemplifying the continuum between meaning and displaying direct evidence.

Suppose Salome has never seen John the Baptist before, and is unable to recognise him. Then for her, seeing a severed head would not be compelling evidence that John is dead. When the head is shown to her by Herod, the evidence is stronger, because it is combined with recognition of his intention. This revised scenario seems to involve both direct and indirect evidence (i.e. both showing and meaning): Herod showed Salome that the person whose head he is displaying was dead, and meant that this person was John. Of course, Herod also overtly intended Salome to think that he was responsible for John's death, and that he had had John killed to satisfy her wishes; since these were not wholly evidenced by John's severed head, they must have been meant.

Perhaps Grice could have said that as long as recognition of the communicator's intention plays a role—however small—in the addressee's coming to the intended conclusion, the case is one of meaning. This seems to fit with his stipulation, in the third clause of his definition, that the audience's recognition of the utterer's intention should be "at least part" of their reason for producing the intended response. It would follow that any case of 'showing that' in which the evidence for the intended conclusion was less than decisive would have to be reclassified as a case of 'meaning that'. But surely, if the part played by recognition of the utterer's intention can vary from 100% to less than 1%, then many, if not most, cases of showing a piece of evidence seem to involve meaning, and the common sense understanding of meaning, and of the distinction between showing that and meaning that, is lost. A more sensible response would be to study the whole continuum—characterised by its two end points of pure meaning and pure showing—as such, and get rid of the third clause. However, this amounts either to extending the notion of speaker's meaning way beyond what is intuitively recognisable as such, or to demoting it from its central theoretical role to a loosely descriptive use that may nonetheless be adequate when dealing with fairly standard cases of linguistic communication.

2.2 *The continuum between determinate and indeterminate 'meaning'*

A second difficulty Grice was aware of with the notion of speaker's meaning arises when one tries to complete a description of the form: "The speaker meant that ___". As Grice recognised, it is not uncommon for at least part of the intended meaning to be less than fully determinate, so that the best rendering of it may be an open disjunction of propositions, and hence not itself a proposition. As Grice put it (1989: 39–40),

> Since, to calculate a conversational implicature is to calculate what has to be supposed in order to preserve the supposition that the Cooperative Principle is being observed, and since there may be various possible explanations, a list of which may be open, the conversational implicature in such

> cases will be an open disjunction of such specific explanations, and if the list of these is open, the implicatum will have just the kind of indeterminacy that many actual implicata do in fact seem to possess.

But this is tantamount to saying that there are some cases of speaker's meaning where "The speaker meant that ___" cannot be properly completed, not because the speaker failed to communicate a meaning, but because that meaning is not a proposition.[2]

We argued in *Relevance* that there is a continuum of cases from those where the communicator's meaning is a proposition, or can be paraphrased as such, to those where it is not paraphrasable at all. At one end of the continuum are utterances such as the railway official's reply to the passenger's inquiry below:

> *Passenger:* What time is the next train to Oxford?
> *Railway official:* 12.48.

Assuming that he has spoken in a neutral tone of voice and with an impersonal facial expression, his meaning could be paraphrased as the proposition that the next train to Oxford leaves at 12.48, and nothing more. Add an urgent tone of voice or a warning look, and although his assertion would remain the same, part of the intended import would be rather less determinate: he might be implicating, for instance, that the train is about to leave, that the seats are filling up fast, that the platform is further away than the passenger might have thought, that the passenger's estimated walking speed may not be enough to get her there on time, and so on. In that case, his meaning would be partly precise and partly vague.

With a hyperbole such as "I could kill for a glass of water", some of the speaker's words (e.g. "kill", "glass") are loosely used and no determinate proposition is asserted. Despite this element of indeterminacy, it is easy to see roughly what she is implicating: for instance, that she is very thirsty, that she has an urgent need or desire for water, and/or that getting hold of some water is a top priority for her. Although her meaning is less than fully determinate, identifying it is unlikely to give ordinary addressees much pause for thought.

With a poetic metaphor such as "Juliet is the sun", the speaker's meaning comes closer to the 'indeterminate' end of the continuum, and has the type of vagueness Grice saw as best rendered by an open disjunction of propositions. As Stanley Cavell comments (1965/1976: 78),

[2] In the 'Retrospective Epilogue', Grice described non-natural meanings as "conceptions or complexes which involve conceptions", and suggested that it would be legitimate to ask "how conceptions enter the picture and whether what enters the picture is the conceptions themselves or their justifiability" (Grice 1989: 350). This is very different from the picture normally presented in philosophy of language and linguistics.

> Romeo means that Juliet is the warmth of his world; that his day begins with her; that only in her nourishment can he grow. And his declaration suggests that the moon, which other lovers use as emblem of their love, is merely her reflected light, and dead in comparison, and so on. … The 'and so on' which ends my example of paraphrase is significant. It registers what William Empson calls 'the pregnancy of metaphors', the burgeoning of meaning in them.

Vaguer still are non-verbal cases such as the following, taken from *Relevance* (Sperber and Wilson 1986/1995: 55):

> Mary and Peter are newly arrived at the seaside. She opens the window overlooking the sea and sniffs appreciatively and ostensively. When Peter follows suit, there is no one particular good thing that comes to his attention: the air smells fresh, fresher than it did in town, it reminds him of their previous holidays, he can smell the sea, seaweed, ozone, fish; all sorts of pleasant things come to mind, and while, because her sniff was appreciative, he is reasonably safe in assuming that she must have intended him to notice at least some of them, he is unlikely to be able to pin down her intentions any further.

We went on to comment,

> Is there any reason to assume that her intentions were more specific? Is there a plausible answer, in the form of an explicit linguistic paraphrase, to the question, what does she mean? Could she have achieved the same communicative effect by speaking? Clearly not. (Sperber and Wilson 1986/1995: 55–6).

If asked what she intended to convey in this case, one of the best answers Mary could give is that she wanted to share an *impression* with Peter. Thus, at one end of the paraphrasability continuum are cases where the speaker's meaning is fully determinate, and at the other are those involving the communication of impressions, where the communicator's meaning cannot be paraphrased without loss.

2.3 *The two continua combined*

In raising these two issues, in arguing that there is a continuum between meaning and producing direct evidence and that paraphrasability is a matter of degree, we were not just being finicky: we were not making the trivial and boring point that there may be unclear, mixed or borderline cases along both dimensions. If that was the problem, a good theory of meaning and/or a good theory of showing could be used to arrive at theoretically-grounded decisions in unclear cases; and indeed, while awaiting the development of such good theories, one could ignore or idealize away fuzzy or borderline cases and investigate speaker's meaning by focusing on prototypical cases: that is, one could go on with philosophical business as usual. But our point was that to do this would be to idealise away essential features of communication, raising questions about the appositeness of the resulting theories.

Let us call the overtly intended cognitive effect of a communicative act its *intended import*. We want to argue that the two dimensions of intended import we are considering—meaning/showing and determinate/indeterminate—are orthogonal. As we have seen, completing the description "X meant that ___" with a proposition is sometimes unproblematic, sometimes impossible, and there is a continuum of cases in between. We will shortly demonstrate that the same point applies to "W showed that ____". The two continua interact to yield a two dimensional-space, with intended imports occurring anywhere in this space. This is illustrated in Figure 1 and the examples immediately below:

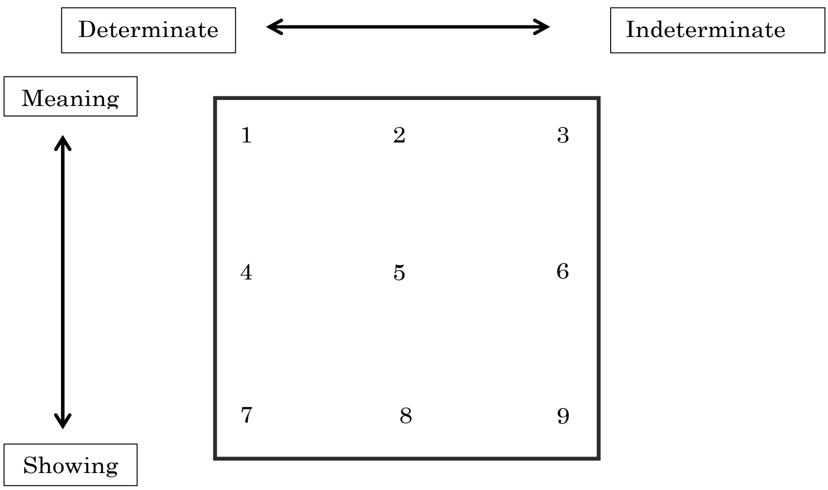


Figure 1

Determinate meaning (vicinity of 1). An example would be the railway official's reply "12.48" to the passenger's question about the time of the next train to Oxford, spoken in a neutral tone with an impersonal facial expression. This is a case of pure meaning, since all the evidence for the intended import comes from the speaker's intentions, and the meaning is determinate, since it is paraphrasable as a proposition. Most discussions of meaning in philosophy of language and linguistics focus exclusively on this type of case.

Semi-determinate meaning (vicinity of 2). An example would be the hyperbole "I could kill for a glass of water," where the intended import is vaguer than with the railway official's reply "12.48", but it is easy to see roughly what type of conclusions the addressee was intended to derive. This is a case of meaning, since all the evidence for the intended conclusions is indirect, but the meaning is less than fully determinate.

Indeterminate meaning (vicinity of 3). With a poetic metaphor such as "Juliet is the sun", the intended import is still vaguer, and is not paraphrasable as a proposition at all. This is again a case of meaning, since all the evidence for the intended import is indirect, but it is closer to the 'indeterminate' end of the paraphrasability continuum.

Determinate meaning/showing (vicinity of 4). When asked who is the tallest pupil in the class, the teacher points to an individual who at first sight is the tallest in the class (although some pupils might be absent) and says, "He is." She both means that the pupil she is pointing at is the tallest (since some of the evidence for the intended conclusion comes from her intentions), and displays direct evidence that he is the tallest. In both cases, the intended import is paraphrasable as a proposition.

Semi-determinate meaning/showing (vicinity of 5). On a tourist trip, Mary points to the view and says "What a view!" Here, the linguistic meaning of her utterance (combined with her tone of voice, facial expression etc.) indicates roughly what type of conclusions she expects the addressee to derive, but does not pin them down precisely, so the utterance falls towards the middle of the 'determinate/indeterminate' continuum. Moreover, the evidence for the intended conclusions comes both from Mary's intentions and from what she has pointed out, so the utterance also falls towards the middle of the 'meaning/showing' continuum.

Indeterminate meaning/showing (vicinity of 6). On a tourist trip, Mary points to the view and says "Wow!" This time, the linguistic meaning of her utterance (to the extent that it has one) gives no more than a rough indication of the type of conclusions the addressee is being encouraged to derive, and the intended import is not paraphrasable as a proposition at all. In Grice's terms, what Mary communicates is an open disjunction of propositions; in our terms (to be discussed further below), what she communicates is an impression.

Determinate showing (vicinity of 7). When asked for the time, Mary points to a clock showing the time as five o'clock. Here, the intended import is as determinate as if she had said "It's five o'clock". However, the case is one of showing rather than meaning, since all the evidence for the intended conclusion comes from the clock itself, rather than from the fact that it has been pointed out.

Semi-determinate showing (vicinity of 8). Peter and Mary are out for a walk when she points to menacing clouds on the horizon. Here it is easy to see roughly what she intends to convey—that it may rain soon, that they should reassess their plans and maybe think about curtailing their walk—but the intended import is less than fully determinate.

Indeterminate showing (vicinity of 9). An example might be showing pictures of one's children. Here, there is no proposition that would complete the description "The communicator showed that ___", and the intended import cannot be rendered as a proposition at all.

Notice that it is possible to mean and show the same thing, as when the teacher, asked who is the tallest pupil in the class, points to the tallest individual in the room and says "He is". This allows us to handle a type of example that seems to be incompatible with the third clause of Grice's definition of meaning, and led him to contemplate dropping this clause entirely. When the communicator is producing a logical argument, she typically intends her audience to accept the conclusion of this argument not on her authority, but because it follows from the premises (this type of case is insightfully discussed by Schiffer 1972):

> *Conclusion of argument: p, q, therefore r (from already stated premises):* While *U[tterer]* intends that *A[ddressee]* should think that *r*, he does not expect (and so intend) *A* to reach a belief that *r* on the basis of *U*'s intention that he should reach it. The premises, not trust in *U*, are supposed to do the work. (Grice 1989: 107).

Since the third clause was crucial to maintaining Grice's distinction between 'meaning that' and 'displaying direct evidence that', he was reluctant to drop it. We would analyse this type of example as a case of determinate meaning/showing. The speaker provides both direct and indirect evidence that the conclusion follows from the premises: that is, she both means it and shows it.

There is another type of case that Grice did not discuss, but that does raise a serious problem for his distinction. Many utterances contain deictic elements whose function is not just to specify a referent but also to specify conceptual content (the referent being a token that contributes to the interpretation a type to which it belongs). Compare, for instance:

> To open a champagne bottle, you can do this (*demonstrates how to open a champagne bottle by opening one*)
>
> To open a champagne bottle, you should do this (*demonstrates how to open a champagne bottle by opening one*)

The first communicative act (comprising both the utterance and the demonstration) does not satisfy the third clause of Grice's definition of speaker's meaning, since the demonstration provides sufficient evidence of the fact that a bottle of champagne *can* be opened in the way demonstrated. By contrast, the second communicative act (where the only difference is that 'can' has been replaced by 'should') is a perfect case of Gricean speaker's meaning, since the demonstration is not sufficient evidence for the normative claim. But of course the two acts are very similar in their communicative import, and should be analysed in very similar ways.

This example also illustrates the fact that deixis which helps to specify conceptual content is quite commonly a source of indeterminacy. In a communicative act of this type, the demonstrative behaviour is merely indicative of what the speaker intends to convey. Some of its features should be replicated in future performances, and others not. As in most 'how-to' demonstrations, the movements are individually highlighted in a way that is useful to the demonstration, but not to the opening of a champagne bottle. This highlighting is not to be replicated in future performances. Moreover, the demonstrator has her own idiosyncrasies—she may be left-handed, for instance—that need not be replicated either. To comprehend such a communicative act involves inferring the type of action of which the demonstration was an ostensive, and hence untypical, token, and understanding that the 'this' denotes the type. Quite commonly, such demonstrations are used because no perspicuous verbal description is available—in which case, the content communicated is not paraphrasable.

This example shows that not only meaning and showing but also determinate and indeterminate aspects of the intended import can coexist in a communicative act. The co-occurrence of precise and vague import is also common in ordinary verbal communication, as when the railway official's reply "12.48" is accompanied by a warning tone or look.

In the next section, we consider the cognitive background against which communication takes place, reviewing several distinctions among types of mental state that are relevant to our discussion. In the following two sections, we introduce a theoretical notion, *manifestness***,** which helps to clarify the relations among these various types of mental state. Finally, we will use the notion of manifestness to characterise ostensive communication, and apply the resulting framework to some of the examples discussed in sections 2 and 3.

To keep the discussion brief, we will consider only declarative (as opposed to directive) acts, for instance, saying "It's five o'clock", or pointing to a clock showing the time as five o'clock. In *Relevance* and elsewhere, we have suggested how to extend the analysis to other kinds of illocutionary act (Sperber and Wilson 1986/1995, chap. 4, section 7; Wilson and Sperber 1988/2012).

3. *Types of mental state: beliefs and impressions*

In his earliest work on meaning, Grice assumed that declarative acts were intended to induce a belief in the audience (e.g. the addressee of the utterance "It's five o'clock" would be intended to form the belief that it was five o'clock). As he later recognised, however, reminders and recapitulations present problems for this approach (Grice 1989: 106–7):

> *Reminding*: Q: "Let me see, what was that girl's name?"
> A: "Rose" (or produces a rose).
>
> The questioner is here already presumed to believe that the girl's name is Rose (at least in a dispositional sense); it has just slipped his mind. The intended effect seems to be that A should have it in mind that her name is Rose.
>
> *Review of facts:* Both speaker and hearer are supposed already to believe that *p* (*q*, and so forth). The intended effect again seems to be that *A* (and perhaps *U* also) should have "the facts" in mind (altogether).

In response, Grice (1989: 109) suggested that declarative acts might be intended to induce not just a belief but an 'activated belief' (in his terms, a belief that the addressee not only has, but "has in mind"). An addressee might fall short of having an activated belief in one of three ways. He might:

(1) neither believe that *p* nor have it in mind that *p*
(2) believe that *p* but not have it in mind that *p*
(3) not believe that *p* but have it in mind that *p*

Ordinary assertions, reminders and recapitulations could then be seen as addressing different types of shortfall, inducing activated beliefs by different routes.

Grice's suggestion raises a more general question about distinctions among types of belief (roughly captured by contrasts such as 'activated/latent', 'occurrent/dispositional', 'explicit/implicit'), and the extent to which declarative acts are intended to induce some specific type of belief in the audience. We will argue that the typologies of belief-states common in philosophical psychology may not be adequate to answer this question.

3.1 *Occurrent, dispositional and implicit beliefs*

Beliefs are commonly seen as playing a central causal role in the explanation of thought and behaviour. In the kind of accounts we are concerned with here, a belief is a representation that has to be occurrent or activated in the mind in order to play such a causal role. Occurrent or activated beliefs contrast with inactive or dispositional beliefs, which are also understood as being 'in the mind' (in a different sense from Grice's), although not immediately available for use as premises in theoretical or practical inferences.

A somewhat psychologically richer way of describing this distinction might be to say that activated beliefs are in working memory, whereas latent beliefs are in long-term memory and have to be retrieved in order to play a causal role. Or, to borrow a metaphor from Robert Audi (1994: 420),

> What is dispositionally as opposed to occurrently believed is analogous to what is in a computer's memory but not on its screen: the former need only be brought to the screen by scrolling—a simple retrieval process—in order to be used, whereas the latter is before one's eyes. Compare a dispositionally believed proposition's needing to be "called in," as in answering a request to be reminded of what one said last week, with an occurrently believed proposition's being focally in mind, roughly in the sense that one attends to it, as where one has just formulated it to offer as one's thesis.

Activated and latent (or occurrent and dispositional) beliefs are seen as representations 'in the mind'. Both are also described as 'explicit', and contrast with contents that an agent may be said to believe even though they are not represented in her mind; these are sometimes called 'implicit', or 'tacit', beliefs.

Here is how Eric Schwitzgebel (2006) presents the distinction between explicit and implicit beliefs in the *Stanford Encyclopedia of Philosophy*:

> One believes *P explicitly* if a representation with that content is actually present in the mind in the right sort of way—for example, if a sentence with

> that content is inscribed in the "belief box"... One believes *P implicitly* (or *tacitly*) if one believes *P*, but the mind does not possess, in a belief-like way, a representation with that content.[3]

This raises the question of how to distinguish implicit beliefs from contents that are not themselves believed, although they follow logically from one's beliefs. Schwitzgebel suggests that there may be no clear cut-off point between the two:

> Perhaps all that's required to implicitly believe something is that the relevant content be swiftly derivable from something one explicitly believes … Thus, in the planets case, we may say that you believe explicitly that the number of planets is 9 and only implicitly that the number of planets is less than 10, less than 11, etc. Of course, if swift derivability is the criterion, then although there may be a sharp line between explicit and implicit beliefs (depending on whether the representation is stored or not), there will not be a sharp line between what one believes implicitly and what, though derivable from one's beliefs, one does not actually believe, since swiftness is a matter of degree.

In keeping with this suggestion, Robert Audi (1994: 419) argues that what are generally called implicit beliefs are better viewed not as beliefs at all, but as *dispositions to believe*:

> Do you believe that this sentence has more than two words? And do you believe that 98.124 is larger than 98? It would be natural to answer affirmatively. And surely, for most readers considering these questions, that would be answering truly. […] Antecedent belief may be the readiest explanation of our spontaneous answers, but it is not the best explanation. I contend that, here, what may seem to be antecedently held but as yet unarticulated dispositional beliefs are really something quite different: dispositions to believe.

Integrating these 'dispositions to believe' into his computer screen metaphor, Audi writes:

> By contrast with both of these cases of actual belief [i.e., occurrent and dispositional beliefs], propositions we are only disposed to believe are more like those a computer would display only upon doing a calculation, say addition: the raw materials, which often include inferential principles, are present, but the proposition is not yet in the memory bank or on the screen. The suggested difference between a dispositional belief and a disposition to believe is in part that between accessibility of a proposition by a retrieval process that draws on memory and its accessibility only through a belief formation process.

As a result of these considerations, one might divide beliefs into three categories, as in Figure 2:

[3] Schwitzgebel notes that some philosophers use the term '*dispositional*' for what he is calling implicit beliefs. However, he reserves the term for latent (as opposed to occurrent) beliefs, and we will follow him on this.

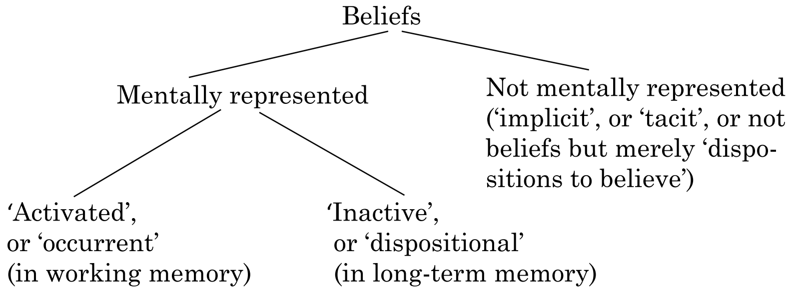


Figure 2

The problem with this proposal is that what actually exists, at least in the case of dispositional beliefs and dispositions to believe, are not distinct categories but a continuum of cases, Arguably, we do actually have activated representations in working memory. However, there is a long-standing consensus in the psychology of memory that what we have in long-term memory is not a repertoire of representations that we can simply move to working memory (or scroll down to), but traces or bits of information from which actual representations are reconstructed for use in working memory. This is not to deny that there are also likely to be some full-fledged representations which can simply be activated (as when you remember the Pythagoras theorem you learned by heart at school). But when you remember facts about the last departmental meeting—say, that John spoke after Jean, and that she seemed irritated—you are not simply moving these representations from long-term to short-term memory: the chances are that they were not stored there as distinct representations in the first place. What you have in long-term memory is information from which these representations can be accurately constructed—as opposed to just being pulled out. With some pieces of memorized information, retrieval—a misleading term—involves more inferential reconstruction than with others. But the point is that, since retrieval from long-term memory typically involves some inference, it is not possible to distinguish implicit beliefs from latent/dispositional beliefs on the ground that one is derived via inference and the other is not.

Here is a continuum between dispositional and implicit beliefs in the spirit of Audi's argument that we merely have a disposition to believe that 98.124 is larger than 98. If the only requirement for a belief of yours to be dispositional is that you have memorized this information at some point in the past and are now remembering it, then your belief that 9 is larger than 8 is surely dispositional: the sequence of numbers from 0 to some small number is permanently represented in your mind.[4] Your belief that 99 is larger than 98 might also be disposi-

[4] This is not the same as having the relation between any two successive numbers in that sequence itself stored as a distinct representation, but we will ignore this for the time being.

tional, since you have probably also entertained it (in some form), for instance when counting to 100. What about your belief that 7899 is larger than 7898? Well… For some *n*, your belief that *n* is larger than *n–1* has been previously represented in your mind, whereas your belief that *n+1* is larger than *n* has not (and of course you don't know the value of *n*). However, there is no interesting epistemic or psychological difference between beliefs about successive numbers lower than *n* and beliefs about successive numbers higher than *n*. Moreover, reaction time studies show that such beliefs are constructed on the basis of a mental number line, in a way that goes quite against the idea that previously held beliefs are merely reactivated (Dehaene 1997). These studies provide evidence that, for any two numbers *m* and *n*, it takes more time to answer the question, "Is *m* larger than *n*?" when *n* and *m* are closer than when they are further apart. Thus, it takes more time to answer 'Yes' to the question "Is 29 larger than 28?" than to the question "Is 69 larger than 28?" Yet when *n* and *m* are close, and even more so when they are adjacent, it is much more likely that the proposition *m is larger than n* has already been entertained than when *n* and *m* are distant. So if activating a belief was a matter of retrieving past representations, the answer should be faster for adjacent numbers.

With numerical examples of this type, the propositions we are disposed to believe follow logically from what we explicitly believe. But we are also disposed to believe propositions that follow non-demonstratively from our mentally represented beliefs. So, for instance, we implicitly believe (or are disposed to believe, and in any case would assent to the claim) that the weather in New York will be warmer next July than next January, that more people were born in 1992 than in 1932, that Helsinki is east of Naples, and so on.

How readily we assent to some statement that does not express an activated belief of ours depends on two factors, one epistemic, and the other cognitive. On the epistemic side, we don't simply believe or not believe a proposition: we believe it more or less strongly. The less strongly we believe it, the less willing we may be to assent to it when expressed, and the less appropriate it is to describe our attitude to it as one of belief. The point being that here, too, there is a continuum, between propositions we believe and propositions we neither believe nor disbelieve, with no cut-off point or even a bimodal distribution of instances. On the cognitive-processing side, a given dispositional or implicit belief of ours may be more or less salient, more or less easy to reconstruct or infer. The salience of a belief is not just a function of its epistemic strength: some logical entailments of what we strongly believe may not be salient at all, whereas a merely probable implication of what we weakly believe may be highly salient. For instance, when someone tells us that her sister Jane is in town, the implication that Jane or the Pope is in town is unlikely to become very salient, although it follows logically from what we have been told (and may quite strongly believe).

By contrast, when someone tells us that a neighbour of theirs may have Alzheimer's, the probable implication that the neighbour is old is likely to be quite salient, although our evidence for this is not conclusive. So dispositional and implicit beliefs are on a continuum with propositions we do not believe in any sense (activated, dispositional or implicit), either because they are too hard for us to infer, or because we are not disposed to give them enough credence.

There are also propositions that we are disposed to believe although they don't follow either logically or non-demonstratively from what we already believe, but follow instead from what we already believe together with what we perceive. Consider Audi's example: "Do you believe that this sentence has more than two words?" It is conceivable (just) that on reading this sentence, you formed the mentally-represented belief that you were reading the sentence, 'Do you believe that this sentence has more than two words?' and are able to infer from that belief, together with your ability to inspect it in memory, that the sentence has more than two words. What about, "Do you believe that there are more than 50 words in this paragraph?" You will probably answer 'Yes', but the way you arrive at this answer will be by looking at the paragraph rather than consulting your memory. Here again, there is a continuum of cases between those where you have a disposition to believe immediately, and those where, in order to answer the question, you have to attend perceptually to more than you were attending to already. In these latter cases, the disposition to believe is not (so to speak) wholly inside you, but also involves the environment. However, the fact that the environment is involved does not stop it being a disposition. This environment-dependent disposition to believe may in fact be stronger than a purely internal disposition to believe.

With all this in mind, let's return to Grice's suggestion that a declarative act might be intended to induce not just a belief but an 'activated belief'. In the case of reminders, this suggestion seems quite plausible; but how far does it generalise? Consider the following dialogue:

(*John is offering drinks to his guests; some have already taken whisky, vodka, cognac or orange juice)*

John (to Rita): Do you want some whisky?

Rita: I don't drink alcohol.

Rita is explicitly communicating that she doesn't drink alcohol and implicitly communicating several further propositions:

(a) She doesn't want whisky
(b) Her reason for refusing his offer of whisky is that she doesn't drink alcohol
(c) She doesn't want cognac
(d) She doesn't want vodka
(e) She might accept orange juice

But does Rita intend John to entertain all these implicatures as activated beliefs? Certainly, she intends him to form the activated belief that she doesn't want whisky (implicature (a)). Of the other implicatures, though, she would typically intend him to activate only those that may turn out to be relevant to his actions. Suppose that, instead of forming the activated belief that her reason for refusing his offer is that she doesn't drink alcohol (implicature (b)), he merely holds it dispositionally, to await activation should the need arise. Even so, it would be quite wrong to say that communication has failed: a belief can be communicated without being activated in the addressee. In appropriate circumstances—for instance, if it crosses his mind to offer her vodka instead of whisky—he may form the activated belief that she doesn't want vodka (implicature (d)); otherwise, the fact that he is disposed to form this belief as a result of her utterance should be enough for communication to succeed.

This description of the different kinds of credal disposition an utterance may cause in an audience is sensitive to distinctions Grice does not envisage; however, it is still not fine-grained enough. In many cases of verbal or non-verbal communication, what the communicator wants to do is not to induce a specific belief or set of beliefs in the audience, but to cause what might be roughly described as an impression, giving rise to a range of non-paraphrasable effects. Grice's suggestion that what is conveyed in this type of case might be analysed as an "open disjunction" of propositions is not really helpful; we will try to improve on it here.

3.2 *Impressions*

What is an impression? In section 2, we used the example of Mary, newly arrived on holiday, sniffing appreciatively and ostensively at the fresh seaside air in order to share an impression with Peter. Here are two more examples:

> Robert, working at his desk, is wondering whether to take a break and go for a walk. He gets up, opens the window: the sky is grey; the air is chilly; clouds, some of them rather dark, are moving fast. The impression he forms of the conditions outside make him change his mind. He will stay at home.

> John has told Julia—who believed him—that the artist is a pretentious and rather conventional painter. However, he has to go to the exhibition, and he begs her to come too. As she walks through the gallery, she is pleasantly surprised by several of the paintings. Although she couldn't have pinpointed what she likes about them, she finds them arresting and somehow insightful. The impression she forms makes her change her mind. What John has told her is false. The painter, she comes to believe, is original and talented.

As these examples show, the formation of an impression, just like the formation of a belief, can bring about a theoretical or practical change of mind.

Bringing about a change of mind in one's audience is a typical goal of communication: indeed, with a suitably extended understanding of 'change of mind', it is the goal of any act of communication. The goals of communication, then, can sometimes be achieved by conveying an impression. For instance, Robert says to Susan, "Let's go for a walk." She might answer by opening the window and showing him the poor weather outside. She would thereby be causing him to form an impression about the weather, and doing so ostensively (i.e. in such a way as to let him know that she is trying to communicate with him). Robert could then infer from his impression that it is not a good idea to go for a walk, and moreover that she intended him to form that impression and come to that conclusion, both of which they now share. Or as they walk through the exhibition, Julia might say to John: "I don't know how to put it… not what I expected… these paintings have something…, I'm sure you too must…Thanks for making me come!" As a result of her utterance, John might form an impression of Julia's impression, which in turn might help him share it, or at least revise his first impression in the direction of hers. In many circumstances, this might be quite enough for Julia's communication to succeed.

But what exactly is an impression? Does this common-sense notion pick out some mental state that we should be able to describe in cognitive terms, or is it irrelevant to scientific psychology? In any case, what is the process commonly described as forming an impression, and how does it achieve its cognitive impact? How can an impression be what a communicator wants to convey? How is it triggered and exploited in communication?

4. *Manifestness*

What occurrent, dispositional and implicit 'beliefs' have in common is that there is some proposition that you are likely to some positive degree to entertain and accept as true. Following our proposal in *Relevance,* we will say that this proposition is *manifest* to you. Manifestness depends on two factors mentioned in section 3.1 above: strength of belief and salience.[5] These factors are quite different—one is epistemic and the other cognitive—and for some purposes it would be unsound to lump them together. However, we need to consider their joint effect in order to explain or predict the causal role of a piece of information in the mental processes of an individual. The greater the degree of manifestness (i.e. the resultant or vector sum of these two factors, epistemic strength and salience) of some piece of information to an individual, the greater the causal role of that information in the individual's thought and behaviour.

Here, then, is a definition of manifestness that differs marginally in formulation, though not in import, from the one given in *Relevance* (Sperber and Wilson 1986/1995: 38–41), followed by some clarificatory comments:

[5] In *Relevance*, we used the term 'accessibility' rather than 'salience' to refer to the same property.

Manifestness

A proposition is manifest to an individual at a given time to the extent that he is likely to some positive degree to entertain it and accept it as true.

Comments:

- In *Relevance*, we talked not of propositions but of assumptions as being manifest. Either term will do, and so would 'pieces of information'. What we are talking about are things that can be true or false and that, when they are true, are facts.
- The time referred to is a not a time point but a time span within which an inferential process may take place. Since such processes can be more or less extended in time, the time span may vary from a fraction of a second to a much longer period in which a protracted inference process such as a scientific discovery takes place. What makes a proposition manifest in such a time span is not only the mind-brain state of the individual during that time, but also the environment of the individual and the information it provides him with via perception and communication. Of course, if we are talking about what is manifest to Robert at the moment he opens the windows, or what is manifest to Julia as she walks through the exhibition, or what was manifest to Eratosthenes while he was calculating the circumference of the earth, the situations are quite different. Depending on the time span, what is manifest to the individual may involve information provided by the environment, or communicated by others, to a lesser or greater degree. Another way of putting this is that the mental processes involved may be more or less extended not only in time, but also in physical and social space, and include processes characteristic of what is described in the literature as embodied, situated, or distributed cognition. We see this not as a problem with our definition but as an advantage. Here, though, we will be concerned with the kind of short time span and organism–environment interactions involved in understanding an utterance.
- In *Relevance*, we described the two factors that contribute to the manifestness of a proposition: its salience ('accessibility') and the degree to which it is accepted as true. These two factors affect the probability that a proposition will influence an individual's beliefs or decisions: the higher the probability that it will be accessed, the higher, ceteris paribus, the probability that it will have some influence, and the higher the degree to which it is accepted as true, the stronger that influence. However, manifestness is this ceteris paribus probability of influence, rather than the factors that contribute to it. The same point can be made on the basis of the Cognitive

> Principle of Relevance ('Human cognition tends to be geared to the maximisation of relevance'): Given a belief which has a cognitive effect and which is therefore relevant in a categorical sense, its relevance will be comparatively greater 1) to the extent that its processing is less costly because it is more salient, and 2) to the extent that its epistemic strength is greater. If a proposition is relevant at all, then the greater its manifestness, the greater its relevance.

In *Relevance*, based on this notion of manifestness, we introduced several further notions:

A ***cognitive environment*** of an individual at a time is a set of assumptions/propositions that are manifest to that individual at that time. A cognitive environment can be shared between two or more individuals if it is a cognitive environment of each of them. Among the propositions manifest in a ***shared cognitive environment,*** some may enumerate or identify the people who share that environment. In that case, this shared environment is also a ***mutual cognitive environment***, and all the propositions in it are mutually manifest. In *Relevance,* we spend some time showing that the notion of ***mutual manifestness*** is more realistic, more psychologically relevant, and at least as cogent as the notions of mutual knowledge, common knowledge or common ground (Lewis 1969; Clark and Marshall 1981; Sperber and Wilson 1986/1995: chapter 1, sections 3, 8; 1990; Stalnaker 2002) used in particular to explicate the Grice–Strawson insight that communicative intentions are 'overt'. We will not discuss these uses of the notion of manifestness here.

We also used the notion of manifestness to redefine the content of communicative intentions so as to provide an explicit and unitary account of cases involving the communication of single propositions, on the one hand, and of what Grice describes as "open disjunctions", on the other. This is the use of 'manifestness' that we will now elaborate and discuss.

We will begin our account in the next section by reassessing the types of mental state discussed in section 3 in the light of this characterisation of manifestness. We will argue, first, that the rather artificial categories of occurrent/dispositional/implicit beliefs are on a continuum of degrees of manifestness, and second, that we can give a better account of what it is to have an impression using the notion of manifestness.

5. *Manifestness, beliefs and impressions*

5.1 *Beliefs and manifestness*

Occurrent beliefs are not all equally manifest. In the first place, they are not all equally salient. Suppose you are asked to recommend your two favourite restaurants. You activate a number of beliefs about each, but some of them stand out more than others: for instance, your belief that the desserts at Pierre's are exceptional may be the most salient of all your beliefs about either restaurant, whereas your belief that it is hard to park at La Cantina, although activated, may not be much attended. In the second place, occurrent beliefs may be epistemically stronger or weaker: you might think of both restaurants as having an excellent wine list, but be more convinced of this in the case of La Cantina than of Pierre's. Ceteris paribus, the activated beliefs that are more salient, and those that are more firmly held, are more likely to inform your conclusions.

At a given moment, there may be a genuine functional difference between occurrent beliefs in working memory and dispositional beliefs in long-term memory. However, since the contents of working memory are constantly changing, with some elements being added and others dropping out, a belief that is about to be recovered from long term memory may well play a greater role in a given inference than a belief that is still in working memory (though not for long) because of the role it played in a preceding inference. If we look at the mental status of beliefs dynamically (as we should), it should be clear that an occurrent belief is not necessarily more manifest than a dispositional one. Dispositional beliefs of course vary along the dimensions of salience and strength, and hence of manifestness. Some of our dispositional beliefs may well be less manifest—indeed, much less manifest—than implicit beliefs we have never entertained before but for which there is strong and salient evidence, either in memory or in the environment. For instance, you may be able to answer the question, "Would the children prefer Pierre's or La Cantina?" immediately and with confidence, even though you had never previously entertained the answer you now give; by contrast, it may take you some time to answer the question, "What is the name of the chef at La Cantina?" and your answer may not be entirely confident, even though you have heard the name before (*hum, I remember commenting that it was not an Italian name but a Hungarian one, a famous name actually, Kadar? Molnar? Lukács? Yes, Lukács, like the philosopher, I think*).

To sum up, the division of beliefs or potential beliefs into three categories is too coarse. A gradient in terms of manifestness is more helpful.

5.2 *Impressions and manifestness*

Using the notion of manifestness, we can also give a more precise account of impressions. When Robert, intending to go for a walk, opens the windows to see what the weather is like and alters his plans, what happens to, and in, his mind? We might be tempted to say that, on the basis of his perceptions, he has formed new beliefs and used them as premises in a practical inference. Which new beliefs? Well, maybe the belief that the sky is grey and the air is quite cold, that it is therefore likely to rain, and that the weather is not right for taking a walk. Many combinations of similar and related beliefs would indeed warrant the practical conclusion he has arrived at. Although we cannot know exactly which beliefs Robert formed and used as premises, we might assume that only such a determinate set of beliefs could have caused Robert to change his mind. It might suit our theoretical agenda to think that there are facts of the matter, and that Robert knows them, or may have known them, however fleetingly, at the time.

But we have all been in Robert's position, and we should envisage the possibility that he might have come to his decision without ever being aware of clear and distinct premises, or of deriving his decision from those premises. Here is an alternative description. When Robert opened the window, an array of propositions became manifest or more manifest to him, in the sense characterised above: they became more likely to be attended to, and more likely to be taken as true, than they had been before, and were therefore more likely to influence his decision. He may have been aware of this increase in the manifestness of an array of propositions, and of their general drift, without entertaining all of them, and maybe even without entertaining any of them as a distinct proposition, except for the practical conclusion that he would not go for a walk.

Not all inferences involve step by logical step derivations of explicit conclusions from explicit premises. Arguably, the vast majority of inferences made by humans and other animals do not involve such derivations. What happened in Robert's brain when he opened the window might be better described as changes in patterns of activation, none of which would properly speaking amount to the fixation of a distinct credal representation, but the totality of which would correspond to the formation of an impression. These changes would then jointly inhibit what may have been a distinct volitional representation, his desire to take a walk. Thus, rather than a step-by-step derivation of an explicit conclusion from explicit premises, the inferential process might have consisted in John's impression of the weather undermining his desire to go for a walk. More generally, many (if not all) inferences can be described not as more or less standard logical derivations but as competitions between alternative conclusions (it will rain/it won't rain, let's go for a walk/let's not, and so on). The winner of such competitions is determined by activation or inhibition caused by brain states that represent information in all kind of ways (from consciously entertained propositions to unconscious weightings of features—where 'represent' is broadly understood as meaning *fulfil the function of making some information available for processing*). If the mental mechanisms which decide the outcome of such competitions tend to favour warranted conclusions, then although the process is quite different from a sequence of good old syllogisms, it would still be genuinely inferential. We are not arguing for this view of human inference here, but merely arguing against tying our understanding of the role of inference in communication to an old, narrow and questionable view of what such inference must be like.

We are suggesting, then, that an impression is a change in the manifestness of an array of propositions which all bear on our understanding the same phenomenon, answering the same question, or deciding on the same issue. This is not intended as an analysis of the ordinary use of the word 'impression'. If someone told us that this account of impressions does not capture the ordinary usage of the term, we would say that we are proposing to use 'impression' as a technical term, to denote a psychologically relevant category, whether or not this category is recognised as such in common-sense or philosophical psychology. In fact, though, we believe our technical use corresponds fairly closely to—and, if anything, sharpens—the typically vague common-sense meaning of 'impression.' It may be worth briefly showing this by considering a possible objection. 'To have the impression that' can be construed as expressing a propositional attitude that takes a proposition as its complement. For instance, Robert might say "I had the impression that it would rain". Does this use of 'impression' differ in meaning from the one we were trying to capture above? One might be tempted to say that 'to have the impression that' denotes a weak credal state, not quite a belief that *P*, but a belief that *probably P*. However, this is demonstrably wrong. Suppose you enter a classroom and say, "I have the impression that there are more than fifty and fewer than a hundred people in this room". This is an appropriate use of the phrase, and corresponds to a situation where you have formed an impression in the sense we described earlier, from which a conclusion follows. That is, an array of propositions have become manifest to you, and although you are not aware of them individually, this overall change in your cognitive environment warrants the inference that there are probably more than fifty and fewer than a hundred people in the room. Now suppose you were to say instead, "I have the impression that the number of people in the room is not a multiple of 11," drawing on your knowledge of the fact that the chance of a random integer being a multiple of 11 is one in eleven. In this case, your use of the phrase would not be appropriate, even though you would be expressing the attitude of taking the proposition embedded in your utterance to be probably true. Thus, 'impression' does not simply pick out a weak credal attitude; it picks out a certain type of vague information basis for such an attitude.

We now have all we need to address Grice's worries about the kind of effects a declarative act is intended to induce in the audience, and to provide a framework in which the full range of possible effects—including prototypical cases of speaker's meaning—can be treated in a conceptually unified way.

6. *Manifestness and communicative intentions*

In *Relevance,* we began by proposing an informal and incomplete definition of the two intentions involved in ostensive communication (corresponding to the first two clauses of Grice's definition of meaning) (Sperber and Wilson 1986/1995: 29):

> *Informative intention*: to inform the audience of something;
> *Communicative intention*: to inform the audience of one's informative intention.

Using the notion of manifestness, we then gave a more precise and fuller version which we reformulate slightly here (Sperber and Wilson 1986/1995: chapter 1, section 10):

> *Informative intention*: to make manifest or more manifest to the audience an array of propositions **I**.
> *Communicative intention*: to make it mutually manifest to audience and communicator that the communicator has this informative intention.

We won't discuss here the role of mutual manifestness in the communicative intention. Instead, we want to highlight and develop the claim that in all cases of communication, wherever they fall on the meaning/showing continuum or the determinate/indeterminate continuum, the intended import is achieved in the same way: by making mutually manifest one's intention to make an array of propositions manifest or more manifest to the audience.

Consider the version of Robert's story where he says to Susan, "Let's go for a walk" and she responds by opening the window and showing him the poor weather outside. By responding in this way, she makes manifest to him her intention to make manifest or more manifest an array of propositions which are relevant to his proposal to go for a walk, and which have become perceptually salient as a result of her opening the window. Of course, these are not the only propositions that her behaviour has made more manifest: it has become more manifest, for instance, that the window can be easily opened, that the street is noisy, and that she doesn't want to go for a walk because of the bad weather. However, out of all these propositions that her behaviour has made more manifest, Robert is able to identify the array of propositions that she manifestly intended to make more manifest. How? They are the propositions whose increase in manifestness makes her communicative behaviour relevant in a way she may have intended and expected.

In another version, Susan responds to Robert's proposal by saying, "The weather is really awful!" By replying in this way, she makes manifest to him her intention to make manifest the proposition that the weather is really awful. Her behaviour also makes manifest a variety of other propositions: for instance, that she has a sore throat, and that she doesn't want to go for a walk because of the bad weather. Again, Robert is able to identify the array of propositions she has intentionally made manifest: as before, they are the propositions whose increase in manifestness makes her communicative behaviour relevant in a way she may have intended and expected.

When Susan's response is verbal, the array of propositions she communicates can be partly characterised by enumerating some of its members. These include the explicature(s) of her utterance (in this case, that the weather is awful) plus any implicatures that she made it wholly manifest that she intended to communicate (in this case, that she doesn't want to go for a walk). However, the full array of propositions communicated by Susan cannot be enumerated by listing (or providing a procedure for listing) all its members. Of course, enumerating all the members of an array is not the only way to identify it. For instance, it can also be identified by description. The array of propositions communicated by Susan contains the explicature and the one clear implicature of her utterance, plus all those implications of her utterance that were to some degree manifest to her and that she expected and intended her utterance to make more manifest to Robert in a way that would make her utterance optimally relevant.

When Susan responds non-verbally, by opening the window, the array of propositions she communicates can be characterised purely by a description: she intends to make more manifest to him those propositions which have become more manifest both to her and to Robert—i.e. which have become part of their mutual cognitive environment—as a result of her opening the window, and which are relevant because they imply an answer to Robert's proposal; the gist of this answer being that she doesn't want to go for a walk because of the bad weather.

Enumeration and description are not the only two ways in which an addressee may identify the array of propositions that a communicator manifestly intended to make (more) manifest to him. For instance, as a result of the communicator's behaviour, the addressee may experience a certain change in his cognitive environment, and identify this change, or part of it, as something the communicator intended to cause in him and to have him recognise as what she intended to communicate. In this case, what is needed to identify the array is neither enumeration nor description, but merely metacognitive acquaintance.

Note that in talking of metacognitive acquaintance, we are not bringing onto the scene a new and unheard of kind of psychological awareness of the effects of other minds on our own. On the contrary, our awareness of the psychological effects that others have on us is a quite unremarkable aspect of our interactions with one another. We know it when our understanding of what others have in mind pleases us, angers us, shames us, makes us feel proud, and—less emotionally—makes us see things in a new light, makes us like or dislike things, makes us rethink the past and anticipate the future differently. We are often aware of the fact that a change of mind (whether or not we could spell out its exact content) was brought about by what we understood of the minds of others. What people do when they communicate is precisely to overtly reveal something of their own mind in order to bring about such changes of mind in their audience.

A central claim of *Relevance* and later elaborations of the theory is that, because of the Communicative Principle of Relevance, the addressee can intuitively identify the array of propositions communicated by an act of ostensive communication by following a task-specific 'modular' inferential procedure (Sperber and Wilson 2002). The same procedure applies whether the act is one of 'meaning that' or of 'showing that', and whether the intended import is determinate or indeterminate.

There are two specific propositions that any communicator, using any form of ostensive communication, makes it mutually manifest that she intends to make manifest: these serve as premises for the interpretation of her communicative act, rather than being part of that interpretation. The first proposition describes the particulars of the communicative behaviour of the communicator: for instance, that Susan opened the window in response to Robert's proposal, or that Susan said "The weather is really awful!" in response to Robert's proposal. Any theory of comprehension assumes that some such information is represented in the comprehender's mind. Relevance theory claims that for any act of ostensive communication, there is a second immediately identifiable proposition that the communicator makes it mutually manifest that she intends to make manifest: the presumption that this act of communication is optimally relevant to the addressee (in a precise sense of 'optimal' defined in *Relevance:* 266–71).

Interpreting an utterance involves using these two propositions as premises (together with contextual information) in order to identify the array **I**. It is this identification that constitutes the interpretation of the communicative act. A central claim of relevance theory is that this array is identified by following a path of least effort, and stopping when the resulting interpretation is such that the communicator could have expected it to satisfy the presumption of relevance automatically conveyed by that communicative act. On this approach, the intended import of a communicative act is not inferred on the basis of general maxims or principles, but on the basis of a presumption of the relevance of that specific act, which is communicated by the act itself without being part of its interpretation.

The addressee can identify the array **I** of propositions that an act of ostensive communication makes manifest in a variety of ways. We will consider cases of verbal communication in the next section. Here are some examples of non-verbal communication.

Peter asks Mary: "Did you bring your cell phone?" She answers by showing him her cell phone. Here, the array of propositions she makes it mutually manifest that she intends to make manifest to Peter may be a singleton: the proposition that she did bring her cell phone.

If Peter had said, "I'm sure you forgot to bring your cell phone," Mary's act of showing him her cell phone would have made manifest an array containing, on the one hand, two distinct propositions—that she did bring her cell phone and that Peter was wrong to be sure she had forgotten to bring it—and on the other hand a vague sub-array of propositions whose increase in manifestness would amount to shaming Peter for his lack of trust in Mary.

In a case of non-verbal communication with indeterminate import, as when Mary sniffs appreciatively and ostensively at the seaside air, the whole array **I** is vaguely identified as that array of propositions which makes (or is expected to make) the communicator's behaviour optimally relevant to Peter. Suppose, for instance, that Peter was already appreciating the seaside air on his own: then her act may achieve relevance just by making mutually manifest that what is becoming manifest to each of them is mutually manifest.

7. *What, if anything, remains of speaker's meaning?*

One of our aims in this paper was to show that building an adequate theory of communication involves going beyond Grice's notion of speaker's meaning. Another was to provide a conceptually unified explanation of how a wider variety of declarative acts than Grice was concerned with—including both cases of 'showing that' and 'telling that'—are understood. We will end by considering where the resulting account leaves the notion of speaker's meaning we began with, and what light, if any, it sheds on the rather fuzzy intuitions that Grice's definition was designed to capture. In this last section, we focus on linguistic cases.

One intuitive distinction that Grice originally wanted his definition of speaker's meaning to capture was between 'deliberately and openly letting someone know' (by displaying direct evidence for the intended conclusion) and 'telling' (where all the evidence would be indirect). As discussed in section 2, the existence of a continuum of mixed cases involving both direct and indirect evidence (in different proportions and combinations) presents problems for this approach. In cases of 'showing that', either the evidence for the intended conclusion is not only direct but conclusive, or else some of the evidence (or at least some strengthening of the evidence) has to come from the communicator's intentions, and this cannot but lead to over-attributions of 'meaning' as defined by Grice. On the other hand, under-attribution of 'meaning' should occur with cases of 'telling that' where no evidence from the communicator's intentions is needed in order to accept the message, either because the logical structure of the utterance makes it self-confirming (as in Grice's example of a logical argument), or because the utterance refers to an object or event that provides conclusive evidence for the truth of the message (as in the utterance, "To open a champagne bottle, you can do this" where "this" refers to a demonstration of how to open a champagne bottle).

For philosophers of language and linguists who only want to use Gricean notions to discuss linguistic cases, one rather convenient, though un-Gricean, way to go to is to forget about the third clause of his definition of utterer's meaning, forget about his extended sense of 'utterance', and forget about his concern with providing a principled distinction between cases of meaning_{NN} and cases of mere showing. Linguistic pragmatics is based on the distinction between what is linguistically encoded and what is verbally communicated. Grice illuminated our understanding of what is verbally communicated, i.e. of 'speaker's meaning' in the ordinary sense; so let's forget his attempt at a more ambitious theoretical definition, treat the first two clauses of his definition as necessary conditions characteristic of speaker's meaning, and make it a tacit rule only to study cases of verbal communication (or gestures like nodding that stand for verbal behaviour) for which this two-clause definition of speaker's meaning seems to pick out the intended phenomena.

We have given two reasons not to go that way. The first is that Grice was right to characterise, in a novel and ground-breaking way, a type of communicative behaviour—what we have called ostensive-inferential communication—that encompasses, but is not restricted to, verbal communication. The first two clauses of his definition of speaker's meaning (or better, the informative and communicative intentions proposed in relevance theory and inspired by Grice) do identify a fundamental form of human interaction, in the context of which, inter alia, verbal communication can be better understood and studied. The second reason is that one and the same conceptual twist—starting from the assumption that the aim in all cases of human ostensive-inferential communication is to make an array of propositions (more) manifest —makes it possible to handle both the meaning/showing continuum and the determinate/indeterminate continuum in a unified way. Focusing solely on the study of verbal communication does not resolve the problem that this second continuum raises for the Gricean approach.

In verbal communication, as in non-verbal communication, we find that addressees exploit the full range of methods for identifying the array of propositions that the communicator intends to make (more) manifest.

In a case of 'telling that' with no implicatures, where the information the speaker intends to communicate is a single clear, paraphrasable proposition, the array is a singleton. An example would be the railway official's reply "12:48" to the question, "When is the next train to Oxford?" In relevance theory, we define a notion of strength of communication which applies to individual members of a communicated array. A proposition is strongly communicated to the extent that it is strongly mutually manifest that the communicator intends to make this specific proposition manifest to the addressee. Any strongly communicated proposition falls unproblematically under the description 'speaker's meaning'. Thus, the proposition that the next train to Oxford leaves at 12:48 is strongly communicated by the railway official and is also a clear case of speaker's meaning. Whenever the array of propositions **I** is a singleton, its single member is strongly communicated and is a prototypical case of speaker's meaning.

In a case of 'telling that' with clearly identifiable explicatures and implicatures, the array can be identified by enumerating its members. Consider, for instance, the following exchange between new acquaintances chatting at a dinner party in London:

Rob: Do you live in London?
Jen: I live in Chelsea

Here, the relevance of Jen's utterance depends on two clearly identifiable propositions: the explicature that she lives in Chelsea, and the implicature that she lives in London. These two are strongly communicated and are good cases of speaker's meaning. Given that the exchange takes place in London, Jen may take the precision of the phrase "in Chelsea" to be relevant to Rob and not to carry any further implicatures. However, in appropriate circumstances it might carry such implicatures, and her tone of voice, looks, and so on (although not indispensable) might help to achieve these further effects. Since Chelsea is a distinctly posh neighbourhood, Rob's social status would be a highly relevant contextual factor in interpreting Jen's utterance. If he were of a similar social status, her utterance might make weakly manifest that she is willing to share with him more personal information than he has requested, that she is not unwilling to see him again, and so on. The gist is clear, but no single proposition is strongly communicated or could be confidently described as part of Jen's meaning. On the other hand, if Rob were, say, a poor academic, then Jen's utterance, depending on the tone of voice, might weakly implicate that she belongs to a different and superior milieu and is not eager to deepen their acquaintance, or, especially if her tone of voice is apologetic, that she is aware of living in privileged circumstances and wishes he would not hold it against her, and so on. In all these cases, if Rob is at all savvy, he will correctly understand that something more than an answer to his question has been subtly communicated. The array **I** in this case is identified by its two strongly communicated members plus an awareness of some further cognitive effects that Jen was overtly intending to achieve by answering in the way she did.

It is all too easy for pragmaticists simply to ignore these weak effects and implicatures, and concentrate on strong implicatures that fit straightforwardly with the notion of speaker's meaning. However, as we have been arguing and will illustrate again below, there is a continuum of cases which should signal that the research is not quite on the right track.

The problem raised by such effects becomes harder to ignore when it affects not only implicatures but also explicatures. As we have argued (Sperber and Wilson 1998, 2008; Wilson and Sperber 2002), the content of implicatures and explicatures is inferred through a process of mutual adjustment whose goal is to produce an overall interpretation consistent with the addressee's expectations of relevance, and where the implicatures are warranted by the explicatures. The linguistically encoded sense of an utterance serves as a piece of evidence of the intended meaning, and provides a point of departure for constructing a contextually appropriate meaning that may be narrower or broader than the encoded meaning, or overlap with it, or even be identical with it, this latter outcome being neither preferred nor arrived at 'by default'.

Recent work in lexical pragmatics confirms that most encoded concepts are adjusted, or modulated, in the course of the interpretation process (Carston 1997; 2002) Sperber and Wilson 1998, 2008; Wilson and Carston 2007). Here is an example where the explicature contains an ad hoc concept constructed for the purposes of that particular interpretation:

Mark: We can't afford La Cantina.
Pamela: I've got money.

Pamela's utterance "I've got money" would be literally true if she had 50c in her pocket; but it would not be optimally relevant. In fact, if she only had 50c, it would have been more relevant to produce the literally false utterance, "I have no money". What Pamela's reply communicates to Mark is the relevant information that she is both able and willing to pay for a meal at La Cantina. This implicature would not be warranted by the literal interpretation of her utterance, hence the construction of an ad hoc concept, MONEY*, whose presence in the explicature warrants the implicature. But what exactly does she communicate explicitly when she says "I've got money"? The explicature, although not very vague, is not that easy to spell out. Pamela is asserting more than simply that she has money: if it turned out that she has 50c, Mark could justifiably complain that she had not just misled him but lied to him. Nor does she *explicitly* communicate that she has enough money to go to La Cantina: this is a consequence, rather than a rendering, of what she explicitly conveys. Basically, what she is referring to is an amount of money such that she is willing and able to pay for a meal at La Cantina, an amount which cannot be less than what the bill is likely to come to, but which may be quite a bit more. Mark might know her well enough to figure out roughly what amount she means, or he might just defer to her as the 'expert' (without this being a case of deferential meaning!). But in any case, there is no word or expression in English, or in any meta-language used by semanticists, which denotes what Pamela has in mind, and which she succeeds more or less in communicating. In other words, Pamela's explicit meaning in this case has a certain vagueness both for her addressee and for the analyst, and while this does not compromise communication between her and Mark in any way, it does compromise a standard account of her speaker's meaning.

Elsewhere, we and others working in the framework have given many examples and analyses of ad hoc concepts: some, like "money" in the preceding example, involve narrowing the linguistically specified meaning, while others involving broadening it, and others overlap with it. Note that we have described the case of "money" as one of narrowing, but in most situations, it could be a case of both narrowing and broadening, and hence of overlap. For instance, Pamela could very well say "I have money" when she has only a few dollars in her wallet, but also a credit card. In all these cases, pinpointing a proposition that would constitute the speaker's meaning is difficult or impossible. On the other hand, all these cases can be easily described as making more manifest an array of propositions any of which would warrant the implicature. The speaker is encouraging the hearer to accept any proposition from this array as quite probable, while not committing to one of them in particular. This is not quite what is understood by 'speaker's meaning', then, but perhaps it might be close enough for some people to be tempted to idealise away the complexity and ignore our proposal about how to treat it.

These were cases of fairly strong communication. But as we have emphasised, there is a continuum between these and cases of very weak communication, where any conceivable paraphrase of the speaker's meaning would be quite defective. This happens when both implicatures and explicatures are weak, as is typically the case with metaphors.

Suppose that in an idle chat among friends, someone tells you, "Freddy is a waste of space". The idea that a person could be a waste of space has no clear literal sense, and you will have to construct an ad hoc concept WASTE OF SPACE* in the course of the mutual adjustment process in order to arrive at an array of implicatures that satisfies your expectations of relevance (which in this case are themselves likely to be rather vague and unconstraining). What you will probably end up with is a general impression of Freddy, based on the explicature that Freddy is a WASTE OF SPACE*, an ad hoc concept derived by adjusting the linguistically encoded meaning in the light of whatever information is available to you about Freddy, your friend and the relations between them. In this example, both the explicature and the implicatures are weak: neither is easily paraphrasable, and although your friend has succeeded in communicating with you, you may find it hard to say exactly what she meant. On the other hand, the description of what is communicated in terms of increasing the manifestness of an array of propositions can be developed without idealising anything away. It won't inject into the description the kind of precision and crispness that some would feel more comfortable with, but the phenomenon itself lacks both precision and crispness.

Or, to conclude with the classic example "Juliet is the sun", the explicature (one might say) is Juliet is the SUN* where, SUN* is an ad hoc concept whose meaning is (vaguely) specified by mutually adjusting explicatures and implicatures in order to satisfy expectations of relevance: the explicature that Juliet is the SUN* must carry an array of implicatures which makes the utterance relevant as expected, and the sense of SUN* must be such that the explicature does indeed contextually imply these implicatures. These implicatures are weak, and cannot be enumerated. Hence, the explicature that warrants these implicatures is itself weak. There is no paraphrase in an adequate metalanguage—or even in English used as such a metalanguage—that provides a plausible analysis or rendering of the speaker's explicit meaning. Even adding starred concepts to the metalanguage (as someone might suggest) would not allow us to identify a proposition as the speaker's explicit meaning, since what a starred concept does in this context is to vaguely indicate a range of possible interpretations that are all made more manifest (i.e. more probable and salient) without any one of them being THE correct interpretation. Just as Romeo need not have intended any one of these propositions to be taken as his exact meaning, so the audience need not, indeed should not, aim to attribute any exact meaning to him.

So the intended import of "Juliet is the sun", as of so many creative metaphors, is best described as an array that the audience identifies not by enumeration but by metacognitive acquaintance, by attributing to the communicator's intention what they mentally experience. In general, what is needed for successful communication is that the addressee's mind be changed in the way overtly intended by the communicator, i.e. that the addressee be now disposed or more disposed to draw the kind of inferences the communicator intended (or at least that the addressee should understand the communicator's intention, even if he does not fulfil it). The communicator need not intend the addressee to make this or that specific inference; her intentions may concern only the general drift of the addressee's inferences and remain quite vague, and so may the addressee's understanding, without this amounting to a failure of comprehension. What is aimed at in such cases of weak communication is a degree of cognitive alignment, not a duplication of precise contents.

What, then, remains of speaker's meaning? Cases in one corner of the bidimensional continuum we have described; cases that have held the attention of linguists and philosophers of language at a time when pragmatics was non-existent, underdeveloped or, more recently, ignored; cases that we have tried to show do not have the kind of unity and autonomy needed to constitute a proper object of theorising. Like the proverbial drunkard in the night looking for his glasses under the lamppost not because of any strong reason to believe that they were there, but because at least he could see there, students of language have stayed close to the lampposts of semantics and logic. The drunkard's strategy need not be irrational. But after a while… especially if there are glimmers of light around…

Acknowledgements

We are grateful to Dunja Jutronić and all the participants in the conference on Philosophy of Linguistics and Language, University of Dubrovnik, September 2014 for insightful comments and discussion during our presentation of this paper. We would also like to thank Nicholas Allott, Mark Jary, Stephen Neale and Tim Wharton for valuable comments on an earlier version.

References

Audi, R. 1994. "Dispositional beliefs and dispositions to believe". *Noûs* 28: 419–34.

Bach, K. 1987. "On communicative intentions: A reply to Recanati". *Mind & Language* 2: 141–54.

Carston, R. 1997. "Enrichment and loosening: Complementary processes in deriving the proposition expressed?" *Linguistische Berichte* 8: 103–27.

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

Cavell, S. 1965. "Aesthetic problems of modern philosophy". In *Philosophy in America,* ed. Max Black. Ithaca, NY: Cornell University Press. (Reprinted in S. Cavell (1976) *Must we Mean what we Say?* Cambridge: Cambridge University Press: 73–96.)

Clark, H.H. and Marshall, C. 1981. "Definite reference and mutual knowledge." In *Elements of Discourse Understanding*, eds. A. Joshi, B. Webber and I. Sag. Cambridge: Cambridge University Press: 10–63.

Dehaene, S. 1997. *The Number Sense.* New York: Oxford University Press.

Grice, H.P. 1989. *Studies in the Way of Words.* Cambridge, MA: Harvard University Press.

Lewis, D. 1969. *Convention: A Philosophical Study.* Cambridge, MA: Harvard University Press.

Neale, S. 1992. "Paul Grice and the philosophy of language". *Linguistics and Philosophy* 15: 509–59.

Recanati, F. 1986. "On defining communicative intentions". *Mind & Language* 1: 213–42.

Schiffer, S. 1972. *Meaning*. Oxford: Clarendon Press.

Schwitzgebel, E. 2006. "Belief". In *Stanford Encyclopedia of Philosophy* (*Winter 2011 Edition*), ed. E.N. Zalta. Available at: http://plato.stanford.edu/archives/win2011/entries/belief/

Sperber, D. and Wilson, D. 1986. *Relevance: Communication and Cognition*. (2nd edition with a new Postface, 1995). Oxford: Blackwell.

Sperber, D. and Wilson, D. 1990. "Spontaneous deduction and mutual knowledge". *Behavioral and Brain Sciences* 13: 179–84.

Sperber, D. and Wilson, D. 1998. "The mapping between the mental and the public lexicon". In *Language and Thought: Interdisciplinary Themes*, eds. Peter Carruthers and Jill Boucher. Cambridge: Cambridge University Press: 184–200. (Reprinted in Wilson and Sperber 2012: 31–46.)

Sperber, D. and Wilson, D. 2002. "Pragmatics, modularity and mindreading". *Mind & Language* 17: 3–23. (Reprinted in Wilson and Sperber 2012: 261–78.)

Sperber, D. and Wilson, D. 2008. "A deflationary account of metaphors". In *The Cambridge Handbook of Metaphor and Thought,* ed. R. W. Gibbs. Cambridge: Cambridge University Press: 84–105. (Reprinted in Wilson and Sperber 2012: 97–122.)

Stalnaker, R. 2002. "Common ground". *Linguistics and Philosophy* 25: 701–21.

Vlach, F. 1981. "Speaker's meaning". *Linguistics and Philosophy* 4: 359–91.

Wharton, T. 2009. *Pragmatics and Non-Verbal Communication.* Cambridge: Cambridge University Press.

Wilson, D. and Carston, R. 2007. "A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts". In *Pragmatics*, ed. N. Burton-Roberts. Basingstoke: Palgrave: 230–59.

Wilson, D. and Sperber, D. 1988. "Mood and the analysis of non-declarative sentences". In *Human Agency: Language, Duty and Value*, eds. J. Dancy, J. Moravcsik, and C.C.W. Taylor. Palo Alto, CA: Stanford University Press: 77–101. (Reprinted in Wilson and Sperber 2012: 210–29.)

Wilson, D. and Sperber, D. 2002. "Truthfulness and relevance". *Mind* 111: 583–632. (Reprinted in Wilson and Sperber 2012: 47–83.)

Wilson, D., and Sperber, D. 2012. *Meaning and Relevance.* Cambridge: Cambridge University Press.

Croatian Journal of Philosophy
Vol. XV, No. 44, 2015

Cognitive Environments and Conversational Tailoring

ANNE BEZUIDENHOUT
University of South Carolina, Columbia, USA

This paper explores the psychological notion of context as cognitive environment (CE) that is part of the Relevance Theory (RT) framework and describes the way in which such CEs are constrained during the course of conversation as the conversational partners engage in "conversional tailoring".

Keywords: Relevance theory, psychological notion of cognitive environment, conversational tailoring, procedural meaning, framing, centering.

1. *Introduction*

This brief paper is an elaboration on one section of Bezuidenhout (forthcoming).[1] In the longer paper, I explore the relation between two conceptions of context, namely the psychologistic notion of a cognitive environment (CE) that belongs to the Relevance Theory (RT) framework and the formal notion of an index that belongs to the traditional Twentieth-century philosophy of language framework. The purpose of the longer paper is to argue that the RT notion is better able than the formal notion to handle some problem cases involving the use of indexicals that have been thought to put pressure on the formal notion of context. In order to make my case, I had to address the question as to the nature of and the constraints on CEs—that is, contexts in the RT sense. It is this section of the longer paper that I will focus on here, in the hopes that I can say a little more about the notion of "conversational tailoring" that is central to my account of how CEs are constrained.

[1] The longer paper, Bezuidenhout (forthcoming), was presented at a conference on Philosophy of Language and Linguistics, held at the Interuniversity Center, Dubrovnik, Croatia, September 8–12, 2014. I thank the conference participants, and especially Deirdre Wilson and Karen Lewis, for comments on that longer paper, some of which are relevant to this shorter paper, too.

2. *What are CEs and how are they constrained?*

A person's CE consists of all the information that is manifest to that individual at a time. Information is manifest if it is either currently explicitly represented in the individual's working memory (WM) or could be so represented, *under appropriate conditions*. Sperber and Wilson (1986: 39) add that, at the time of representation, the person must accept the information as true or probably true. In a more recent account, Sperber and Wilson (2014) say that information in a CE is more or less manifest depending on the degree of salience of that information and its degree of probability. A CE can thus be pictured as a field of represented or representable information where the manifestness of that information increases along two orthogonal dimensions, as illustrated in Figure 1 below:

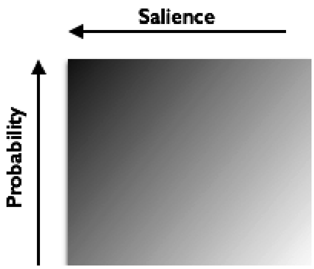


Figure 1: *Degrees of manifestness of information in a CE*

Note that there are two distinct ways in which the term 'manifestness' is being used above. In the first sense, the manifestness of a piece of information is a matter of being currently represented or representable in the WM of some individual communicator. In this sense manifestness is an all-or-nothing matter. Either information can be so represented in WM or it cannot. In the second sense, the manifestness of a piece of information is a matter of degree, depending as it does on the degree of salience that information has for an individual communicator and on the degree of probability that the communicator assigns to that information. The notion of manifestness in the sense of 'being salient' or 'being probable' is a graded notion, as information can be more or less probable/salient. For example, the firmer one's conviction or the better evidence one has, the more manifest (because the more probable) that information will be to you. Clark (2013: 114–116) compares hearing the sounds of rain to actually seeing the rain and says that in the latter circumstance the information that it is raining will be more manifest to you.[2]

[2] Note that Clark is comparing the degrees of manifestness of bits of information across times or across possible situations, assuming that a bit of information (e.g., that it is raining) can be more manifest than that same information is to you at an earlier/later time or than it would be under different conditions. This entails comparing

It might seem strange to have a term 'manifestness' that can be understood as referring both to an all-or-nothing concept and to a concept that admits of degrees. I agree it would be most perspicuous to have two separate terms for these two notions. So I try below to use the term 'representability' for the all-or-nothing concept and reserve 'manifestness' for the graded notion and as a cover term for the dual dimensions of salience and probability. However, I do want to note that there are other terms in English that have a similar ambiguity. The term 'visible' is such a term. For example, a road sign partially obscured by trees and bushes at the side of the road is less visible than that sign after the trees and bushes have been trimmed back. Similarly, the features of someone's face are less visible from a distance than up close. But if these facial features are indeed visible to you, then you can perceptually represent those features, and this applies in both the close-up and the distant situations. In this latter sense, visibility is an all or nothing matter. You either can or cannot perceptually represent the features. What *is* a matter of degree is the fidelity and vividness of your representation. In the close-up situation your percept is likely to be more faithful to the original than in the distant situation and have more phenomenal details filled in.

As noted above, information is manifest (in the all-or-nothing sense) if it is either currently represented in WM or could be so represented, *under appropriate conditions*. The issue now is what should be allowed to count as an "appropriate" condition. If we allow these conditions to be unconstrained, then almost any piece of information would turn out to be manifest to you, given that there is undoubtedly some condition or another under which you would be able to represent that information. For instance, you could explicitly represent arcane facts about quasars, if you were first to undergo graduate training in astrophysics. Clearly, we do not want to allow conditions like this, as we do not want to say that these arcane facts are currently manifest to you.

What is manifest to you now is constrained by what you actually already know—by your *current* store of encyclopedic knowledge. Something that you already know in this sense could be activated in WM and thus is a part of your current CE. We also want to allow that information in your current *physical* environment is manifest if, via the redirection of attentional resources and the (normal) operation of one or

a bit of information represented in an individual's CE with a bit of information represented in another (earlier, later, counterfactual, or possible) CE of that same individual. In a more full-fledged account of CEs, it would be important to add such a dynamic component into the overall account. However, note that, in Figure 1, bits of representable information are pictured as points in a two-dimensional "field" that is the CE of an individual *at a particular time and in a particular situation*. The degree of manifestness of a bit of information is represented as distance from the probability/salience maxima. Figure 1 is too schematic in another way as well; we presumably would want to allow that two distinct bits of information that are represented in an individual's CE can have exactly the same degree of manifestness.

more of your sensory organs, you would become sensorily aware of this information. For example, the hum of the refrigerator in your apartment is manifest to you right now because even though you aren't paying attention to that hum, you would become aware of it if you directed your hearing towards the part of the room where your fridge is located. In addition, information about your interlocutors' informative and communicative intentions, their preferences and abilities, as well as sociocultural information about them (e.g., facts about their social status) will be manifest to you in virtue of your "mind-reading" skills. Finally, interlocutors generally need to keep track of what has transpired so far in the conversation, and thus linguistic information at various levels (e.g., phonological, semantic, and pragmatic) will be manifest to you (and of course it will need to be *mutually* manifest to all the interlocutors in the particular conversational situation if communication in that situation is to succeed). In sum, what is manifest includes encyclopedic, situational, social and linguistic information.

However, even this is still too broad, as we do not want to say that *everything* you know at a particular time is a part of the CE used to understand an utterance by one of your conversational partners at that time. In the 1970s, when Chomsky, Fodor and others popularized the idea of a modular mind, they argued that modules controlling perception and the language module (what Hauser et al. (2002) call the language faculty narrowly conceived) are encapsulated, in the sense that the operations of such modules are sensitive only to domain specific information written in a proprietary code that is manipulated in accordance with module-internal rules, but that the central processor is unencapsulated, in the sense that any piece of knowledge could become relevant for its purposes. This modularist tradition treats pragmatic reasoning as a central cognitive process *par excellence*.

RT in its earliest iterations embraced the idea of pragmatics as part of the central processing system, and of pragmatic inferences as unencapsulated. Where Sperber and Wilson (1986) parted company with Fodor was that, while Fodor believed there could be no scientific account of central processes, Sperber and Wilson thought it was possible to give a scientific account of pragmatics, their RT framework being just such an account. Sperber and Wilson resisted the idea of a "pragmatics module", I believe, because accounts of pragmatics modules on offer in the 1980s appealed to ideas from then existing pragmatic theories (e.g., speech act theory, theories of politeness, and theories of turn-taking in conversation) and simply assumed that the pragmatic processing module (or its sub-modules) operated according to internalized versions of these theories.[3] RT on the other hand was intended to be a single overarching framework to explain all pragmatic processing and as an alternative to theories such as Grice's theory of conversational implicature, Brown and Levinson's politeness theory, and Searle's speech act

[3] See Kasher (1984).

theory. According to Sperber and Wilson, pragmatic reasoning operates according to cognitive and communicative principles of relevance. These principles are conveniently summarized in Wilson and Sperber (2012: 6–7):

> *Cognitive principle of relevance*: Human cognition tends to be geared to the maximization of relevance.
> *Communicative principle of relevance*: Every act of overt communication conveys a presumption of its own optimal relevance.
> *Presumption of optimal relevance*: (a) The utterance is relevant enough to be worth processing; (b) it is the most relevant one compatible with the communicator's abilities and preferences.

The relevance of an ostensive stimulus—and utterances are paradigm examples of such stimuli—is a matter of degree. The greater are the *cognitive effects* of processing a stimulus and the smaller is the *cognitive effort* needed to process it, the more relevant will that stimulus be (Wilson and Sperber 2012: 88). Cognitive effects are modifications of an individual's CE. The information carried by the stimulus will either combine with existing assumptions to yield a contextual implication or will combine with existing information to alter the probability or saliency of that information or will interact so as to eliminate or cancel existing assumptions (Wilson and Sperber 2012: 176, 200). With respect to the effort factor, Wilson and Sperber mention two things that impact it, namely the form in which the stimulus is presented (its audibility, legibility, dialect, register, syntactic form, familiarity, etc.) and the degree to which it taxes the resources of memory and imagination (Wilson and Sperber 2012: 176).

As I said, initially Sperber and Wilson resisted the idea of a pragmatics module. In later iterations of RT, however, they became more receptive to the idea of a pragmatics module, and suggested it is a specialized sub-module of the Theory of Mind module (ToMM) that has been proposed to explain the human capacity for "mind reading"—in the sense of being readily and automatically able to "read" a person's purposive (ostensive) behavior and see it as a manifestation of that person's underlying beliefs, intentions and desires (Wilson and Sperber 2012: 261–278). Since we want assumptions about interlocutors' informative, communicative and referential intentions to be a part of the interlocutors' CEs and hence part of what is *mutually* manifest to them, we want the output from the ToMM to interact with other information in a person's CE. However, precisely because we need more than ToMM output to produce and understand verbal utterances, I would resist the idea of a pragmatics module as a sub-module of the ToMM.

Rather than accounting for the constraints on CEs by arguing that pragmatics is modular and hence that the information that feeds into pragmatic processes is constrained in virtue of being domain specific or encapsulated, the key to understanding how CEs are constrained is to understand the principles according to which CEs are modified in the course of conversational exchanges. We saw above that, according to

RT, a speaker's utterance, which is an ostensive stimulus, is intended to modify the CEs of the conversational partners. Understanding the principles that govern such modifications is important for understanding CEs and their role in utterance production and comprehension. The attempt to shape one's interlocutors' CEs is what I call "conversational tailoring". I have described some aspects of such tailoring in the context of a discussion of generalized conversational implicatures in Bezuidenhout (2015) and in the context of a discussion of presuppositions in Bezuidenhout (2010, 2014).

Not everything that you know is a part of your current CE and hence currently manifest to you. (Moreover, not everything in your current CE is currently activated in the "spotlight of attention". I return to this second issue in the following section). What carves out your current CE from the totality of what you know? This is just the old "Frame problem" raised in early AI research in the 1960s and 1970s. The issue initially arose in attempts to account for commonsense reasoning and conscious deliberation (including scientific theorizing), which modularists such as Fodor believed could never be explained scientifically. However, many of the same issues arise in the context of language performance studies, given the crucial role of pragmatic inferences in language performance. Such pragmatic processing is equally the target of those who believe that non-modular processes are not scientifically tractable.

Given how old the frame problem is, I cannot hope to summarize all the attempts to answer it here. For current purposes, I need merely to invoke the well-entrenched notions of mental frames, scripts, or schemas, which are representations stored in long term memory and that contain default information about typical scenarios we have encountered in the past (e.g., we may have stored a "restaurant" script that contains information about the sorts of things that typically transpire in a restaurant). When one of these scripts is triggered by the current conversational situation, this will already to a large degree shape how you will interpret the actors, actions and events that transpire. Similarly, it has been argued that we operate with various sorts of mental schemas (or heuristics), such as a causal schema that accounts for our tendency to see co-occurring events as causally connected, or a purposive schema that accounts for our tendency to see events (even natural events) as the result of underlying agency or purpose.

These ideas of frames, scripts and schemas certainly explain why not everything you know will be invoked in a particular conversational situation and how information in your CE might be limited to situation-specific information. However, they don't explain what invokes a particular script or frame in the first place. It is on this issue that I think RT has some valuable insights to offer. I will briefly describe two ways in which ideas from RT can explain how particular frames are invoked in the course of a conversation and thus explain how interlocutors are able to shape one another's CEs.

Firstly, the very idea that the inferential phase of language comprehension is a relevance driven process will partially explain why a particular frame is accessed. As we saw above, relevance is a matter of balancing contextual effects and processing effort. Degree of effort is a factor in determining relevance. Thus the more effort it takes to retrieve a frame, the less relevant the stimulus that is interpreted relative to that frame would be (other things being equal). I assume that the processing effort involved in accessing a frame is affected by such factors as recency of use, frequency of use, ease of access (e.g., cases of self-deception may involve blocking access to certain frames; a mother who refuses to read the signs of her son's drug addiction is mentally blocking a way of framing the situation), and so on. And on the flip side, the more contextual effects that are yielded by invoking a particular frame, the more relevant the stimulus that is interpreted relative to that frame would be (other things being equal).

Suppose you are dining in a restaurant with a married couple. As the meal draws to a close, the husband, gesturing towards his wife, suddenly comes out with an utterance of (1):

1. She will pay for it.

You take him to have said that his wife will pay for the meal. After all, you are in a restaurant and maybe you have been thinking for a while about the touchy issue of who will pick up the tab for the meal. So the restaurant script will have been recently activated and be easily accessible. Moreover, the utterance interpreted relative to this frame answers a question that was on your mind and, you reasonably assume, on the minds of the other diners. That is, interpreted in this way, the husband's utterance is highly relevant, and the relevance driven comprehension procedure is likely to halt at this point, without seeking alternative possible interpretations.

Of course, it is always possible that the husband has been silently fuming all evening about his wife's infidelity, which he found out about just before the start of the dinner when the private detective he had hired to follow his wife slipped him a set of photographs of his wife *in flagrante delicto*. Unable to contain himself anymore, he gives vent to his desire for revenge against his wife by uttering (1). In other words, in such a situation, interpreting (1) as the claim that the wife will pay for the meal would involve some sort of misunderstanding, although not one that can be blamed on the interpreters.

This raises an interesting issue that is tangential to my main point, namely whether in this situation the husband's intention to express his desire for revenge is the one that fixes the context for interpretation. I will not pursue this issue here. Suffice it to say that I follow Bianchi (2013), who argues that context-fixing intentions must be ones that the speaker *makes available* to the hearer, in the sense that in the normal course of things a hearer could reasonably be expected to discern these intentions. Intentions that are hidden or opaque are not communicative intentions. The husband had no communicative intentions behind his utterance of (1). If the misunderstanding is detected, some conversational repair can be done and the correct frame for interpreting the husband's utterance of (1) can be made accessible to the interlocutors. Or the husband might exploit the misunderstanding to cover up his embarrassment at showing his emotions in public and let the interpretation that his wife will pay for the dinner stand.

The second RT idea that can help to explain how and when frames, scripts or schemas will be invoked is the idea of *procedural meaning*. The idea is that some linguistic expressions encode procedures rather than concepts. There may also be expressions that encode both concepts and procedures (see Wilson and Sperber 1993, 2012). The notion of procedural meaning "arises from the observation that there are linguistic expressions which can help hearers to follow the [speaker's] intended inferential path" (Clark 2013: 312). Expressions that encode procedures are called procedural markers. One important sub-class of such markers is the class of discourse connectives, such as 'but', 'however', 'nevertheless', inferential 'so', 'because', 'after all', 'moreover', and so on. Many other lexical items and linguistic constructions have been said to encode procedural meanings. For example, prosody, syntactic structures (e.g., it-clefts), interjections (e.g., 'wow' and 'oh'), illocutionary force indicators and hearsay particles have all been taken to encode procedures.[4]

Blakemore (1987: 2002) has discussed discourse markers from an RT perspective, treating them as expressions that guide the interferential phase of interpretation. She argues that they encode rules that indicate the type of CE in which the utterances of which they are a part are to be processed. In this way they guide the hearer towards intended contextual effects and hence reduce the overall effort required to process the discourse. As Clark (2013: 310) puts it, procedural expressions "guide the hearer by making one of the possible inferential connections more salient than others". Such expressions "indicate the way in which the utterance might be relevant and so make particular ways of processing more salient" (Clark 2013: 326).

Consider for example a situation in which a speaker utters the sequence of sentences in (2) below:

2. He's a heavy drinker. His wife left him.

There are several possible ways in which these two sentences could hang together to create a coherent discourse segment. The speaker could intend to convey that the wife's leaving was a consequence of the man's heavy drinking or that it was a reason for his heavy drinking. Or perhaps his heavy drinking and his wife's leaving are just two of the ways in which the man's life is falling apart. There are undoubtedly other possibilities too. Assumptions in the mutual cognitive environ-

[4] For further discussion, see Andersen and Fretheim (2000) and Escandell-Vidal et al. (2011).

ment (including the topic currently under discussion, the identity of the man, what his relation to the interlocutors is, and so on) may make it easy to figure out which of these possible relations between the two events is the one that the speaker intends to be talking about. However, there are ways for the speaker to make it easier on the listeners by using a discourse connective, such as in (3)–(5) below:

3. He's a heavy drinker, so his wife left him.
4. He's a heavy drinker because his wife left him.
5. He's a heavy drinker. Moreover, his wife left him.

In the case of (3), one is constrained to interpret the man's drinking as the reason for the wife's leaving rather than the other way about. In the case of (4), the inferential connection is just the reverse. In both these cases, certain types of scripts or frames are likely to be invoked. In the case of (3), a script about how alcoholic men behave in domestic situations might lead one to infer that the man physically abused his wife while under the influence of alcohol. In the case of (4), a script about how a man scorned by his love would act might lead one to infer that the man sought solace in his local pub night after night and became addicted to alcohol. In the case of (5), a script about people down and out on their luck might lead one to infer that the man has experienced further woes beyond his alcoholism and having been deserted by his spouse, such as having huge debts, and so on.

There is a lot more that could be said about the notion of procedural meaning, about procedural markers, and about discourse markers in particular. There is an extensive literature devoted to all these issues. For current purposes, the main point I wish to emphasize is that procedural markers can help us understand how particular frames, scripts and schemas can be invoked in the course of an unfolding conversation. Procedural markers are used as a way to shape the CEs of one's interlocutors. They are devices for conversational tailoring.

3. *CEs and the center of attention*

In the previous section I invoked the well-entrenched notion of a frame, script or schema as a way of explaining how CEs are constrained and then invoked notions from RT (in particular the ideas of relevance-driven processing and of procedural meaning) to answer the antecedent question as to how particular frames/scripts/schemas are invoked in particular situations. I suggested that by constraining one another's CEs in these ways, interlocutors are engaging in conversational tailoring. In this section I address a second way in which people engage in conversational tailoring. Once having invoked a particular frame, and thereby having made certain information manifest, there is a further question as to which bits of information in the frame are currently activated and at the center of attention. This too requires conversational tailoring.

Although the idea of a bit of information coming to be at the center of attention may seem a simple process, I believe that it is in fact a rather complex matter with multiple dimensions. One way to characterize the process is to think of it as a matter of foregrounding some bit of information in one's CE, which of course presupposes that other information remains backgrounded. The backgrounded information functions partly in an identificatory role, helping to pick out what is at issue (just as something stands out as a figure only relative to a particular sort of ground). Much of what has been treated in the literature as presupposed information could be re-construed as backgounded information. Here I follow Abbott (2000). Thus, the means available to interlocutors to background certain information and, by contrast, to put other information at the center of attention, are just the sorts of devices that have been much discussed in the literature of presupposition. I can't summarize that huge literature here. Bezuidenhout (2010, 2014) makes a start at showing how presupposition triggers can be thought of as devices for separating the foreground from the background.

I will give just one very simplistic illustration here. In the presupposition literature, definite descriptions are frequently treated as presupposition triggers. Thus an utterance of a sentence containing a description of the form 'The F' is said to presuppose the existence of a salient and identifiable F.[5] So consider (6) below:

6. The man in the red shirt looks like he might be able to help us.

Suppose I utter this to you at the entrance to a nightclub in Seoul where we have stopped because we have lost our way. The man at the entrance is checking people's IDs as they enter the club and let us suppose that he is in fact wearing a white shirt. I do not realize that he is standing under a red light that makes his shirt look red from my perspective. The presupposed information that there is a salient and identifiable man in a red shirt is intended to help you identify the man who could help us. Since this information is part of the background that is intended to bring a particular individual to the fore, it does not matter that the information is incorrect, so long as it plays the intended role.[6] It will do this if you are subject to the same visual illusion that I am subject to or if you can see that that from my point of view what is in fact white would look red and make adjustments accordingly.

The use of definite descriptions for this sort of identificatory role is one of the most straightforward ways in which backgrounded (= pre-

[5] Here I follow Beaver and Geurts (2014). For instance, they say and I agree that an utterance of 'It was the knave who stole the tarts' presupposes at least the following: There is a (salient and identifiable) knave. There were (salient and identifiable) tarts. Somebody stole the tarts. The it-cleft construction is another type of presupposition trigger that I do not discuss in the text above.

[6] Note that it is not the flesh and blood individual that is "brought to the fore" or "placed at the center of attention". Rather it is a *representation* of that individual that is centered. What is centered is an element of a person's cognitive environment (CE) and a CE is a mental construct.

supposed) information helps put something at the center of attention. In a more extended treatment I would hope to show that this applies to many other sorts of presupposition triggers too, including grammatical devices such as it-clefts and right- and left-dislocations, phonological devices such as pitch accents and information contours, and many other such linguistic and paralinguistic means that we have for information structuring purposes.

4. *Conclusions*

I have laid out some of the details of the psychological notion of context assumed within the RT framework, namely the notion of context as the mutual cognitive environments (CEs) of the conversational participants. I have also tried to show how the ideas of relevance-driven processing and of procedural meaning can help us to explain the ways in which CEs are constrained in the course of language production and understanding, because they help to explain how in the course of language production and comprehension particular frames, scripts, or schemas are invoked. I then very briefly mentioned some ideas about how information in a CE gets to be at the center of attention. These ideas of framing and centering help us to see some of the ways in which interlocutors are continually engaged in a process of conversational tailoring as a conversation unfolds.

References

Abbott, B. 2000. “Presuppositions as nonassertions”. *Journal of Pragmatics* 32: 1419–1437.

Andersen, G. and Fretheim, T. (eds.). 2000. *Pragmatic Markers and Propositional Attitude*. Amsterdam: John Benjamins.

Beaver, D. and Geurts, B. 2014. “Presupposition”. In *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/win2014/entries/presupposition/

Bezuidenhout, A. 2010. “Grice on presupposition”. In K. Petrus (ed.), *Meaning and Analysis: Themes from H. Paul Grice*. Basingstoke: Palgrave McMillan: 75–102.

Bezuidenhout, A. 2014. “Presupposition failure and the assertive enterprise”. *Topoi*,. Online first http://link.springer.com/article/10.1007/s11245-014-9265-4

Bezuidenhout, A. 2015. “The implicit dimension of meaning: Ways of “filling in” and “filling out” content”. *Erkenntnis* 80 (1): 89–109.

Bezuidenhout, A. Forthcoming. “The role of context in semantics: A Relevance Theory perspective”. In S-J. Conrad and K. Petrus (eds.). *Meaning, Context, and Methodology*. Mouton Series in Pragmatics, Berlin: de Gruyter.

Bianchi, C. 2013. “Illocutions in context”. In C. Penco and F. Domaneschi (eds.). *What is Said and What is Not*. Stanford: CSLI Press: 131–139.

Blakemore, D. 1987. *Semantic Constraints on Relevance*. Oxford: Blackwell.

Blakemore, D. 2002. *Relevance and Linguistic Meaning*. Cambridge: Cambridge University Press.

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

Escandell-Vidal, V., Leonetti, M. and Ahern, A. (eds.) 2011. *Procedural Meaning: Problems and Perspectives*. Bingley, UK: Emerald Group Publishing.

Hauser, M., Chomsky, N. and Fitch, W. T. 2002. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" *Science* 298: 1569–1579.

Kasher, A. 1984. "Pragmatics and the modularity of mind". *Journal of Pragmatics* 8: 539–557.

Sperber, D. and Wilson, D. 1986. *Relevance: Cognition and Communication*. Oxford: Blackwell.

Sperber, D. and Wilson, D. 2014. "What is the role of meaning in communication?" Paper presented at the Philosophy of Language and Linguistics conference, Dubrovnik, Croatia, September 8–12, 2014.

Wilson, D. and Sperber, D. 1993. "Linguistic form and relevance". *Lingua* 90: 1–25.

Wilson, D. and Sperber, D. 2012. *Meaning and Relevance*. Cambridge: Cambridge University Press.

Croatian Journal of Philosophy
Vol. XV, No. 44, 2015

Inferring Content: Metaphor and Malapropism

ZSÓFIA ZVOLENSZKY
Department of Logic, Institute of Philosophy, Eötvös University, Budapest, Hungary

It is traditionally thought that metaphorical utterances constitute a special—nonliteral—kind of departure from lexical constraints on meaning. Dan Sperber and Deirdre Wilson have been forcefully arguing against this: according to them, relevance theory's comprehension / interpretation procedure for metaphorical utterances does not require details specific to metaphor (or nonliteral discourse); instead, the same type of comprehension procedure as that in place for literal utterances covers metaphors as well. One of Sperber and Wilson's central reasons for holding this is that metaphorical utterances occupy one end of a continuum that includes literal, loose and hyperbolic utterances with no sharp boundaries in between them. Call this the continuum argument about interpreting metaphors. My aim is to show that this continuum argument doesn't work. For if it were to work, it would have an unwanted consequence: it could be converted into a continuum argument about interpreting linguistic errors, including slips of the tongue, of which malaprops are a special case. In particular, based on the premise that the literal–loose–metaphorical continuum extends to malaprops also, we could conclude that the relevance-theoretic comprehension procedure for malaprops does not require details specific to linguistic errors, that is, details beyond those already in place for interpreting literal utterances. Given that we have good reason to reject this conclusion, we also have good reason to rethink the conclusion of the continuum argument about interpreting metaphors and consider what additional (metaphor-specific) details—about the role of constraints due to what is lexically encoded by the words used—might be added to relevance-theoretic comprehension procedures.

Keywords: A*d hoc* concepts, figurative language use, inferential comprehension procedures, linguistic error, literal language use, literal–metaphorical continuum, loose use, malapropism, metaphor, relevance theory, Dan Sperber, Deirdre Wilson.

1. *Introduction*[1]

It has been three decades since Dan Sperber and Deirdre Wilson (1986/1995) first formulated their influential framework for the study of communication: relevance theory. Their paper "A Deflationary Account of Metaphor" (2008) contains the two authors' most recent position on metaphor. They set out to show that

> [t]here is no mechanism specific to metaphors, no interesting generalisation that applies only to them. In other terms, linguistic metaphors are not a natural kind, and 'metaphor' is not a theoretically important notion in the study of verbal communication. (Sperber and Wilson 2008: 97)

Accordingly, Sperber and Wilson hold that relevance theory's inferential procedures for comprehending/interpreting linguistic utterances do not include any metaphor-specific details. In this paper, my primary aim is to reconstruct and assess what I take to be the authors' central argument—which I will call the *continuum argument about interpreting metaphors*—to the conclusion that there are no interesting metaphor-specific generalizations that a study of verbal communication should include. I will show that in the light of considerations about linguistic errors of various sorts, including malaprops, we have reason, first, to reject the continuum argument about interpreting metaphors and, second, to consider what metaphor-specific details are worth making room for within relevance-theoretic comprehension procedures: details that spell out the status of lexically driven constraints on interpretation. This paper is primarily concerned with setting up the first (negative) project, I will briefly discuss the second (positive) proposal about developing relevance-theoretic inferential comprehension procedures, and plan to explore it further in future work.

Mrs. Malaprop, a character in Sheridan's (1775) play *The Rivals* had a tendency to make linguistic errors of a special sort: she would describe people as being "the pineapple of politeness" (when she meant *pinnacle*); or "as headstrong as an allegory on the banks of the Nile" (when she meant *alligator*). Such slips of the tongue have since come to be called malaprops. In a framework like relevance theory, how might we characterize the process of interpreting malaprops as opposed to interpreting literal utterances? We will see that addressing this question exposes a challenge for the relevance-theoretic treatment of *metaphorical* utterances.

Within philosophy of language as well as rhetoric the following claims are widely held, considered platitudinous even: the distinction between literal and figurative discourse carries theoretical importance, and metaphorical utterances clearly fall on the figurative side of the divide, constituting departures from literality. Relevance theory calls into question these time-worn claims.

[1] This paper builds on and expands a shorter predecessor (Zvolenszky 2015).

Relevance theory has, since its inception, become a leading research program in pragmatics. Its founders', Sperber's and Wilson's most recent position on metaphorical utterances is that (i) the interpretation/comprehension procedure for metaphors does not require resources beyond those already needed to account for literal utterances (call this the *procedure claim*), and (ii) metaphorical utterances occupy one end of a continuum that includes literal, loose and hyperbolic utterances (call this the *continuum claim*). Relevance theorists seem to regard the continuum claim as one reason to hold the procedure claim; call this the *continuum argument about interpreting metaphors*.

Sperber and Wilson subscribe to this continuum argument:

> We see this continuity of cases, and the absence of any criterion for distinguishing literal, loose, and metaphorical utterances, as evidence not just that there is some degree of fuzziness or overlap among distinct categories, but that there are no genuinely distinct categories, at least from a descriptive, psycholinguistic or pragmatic point of view. Even more important than the lack of clear boundaries is the fact *that the same inferential procedure is used in interpreting all these different types of utterance.* (2008: 111–112, emphasis added)

In this paper, I aim to show that the continuum argument about metaphors, if it were to work, would face an unacceptable consequence: the argument would license *a continuum argument about interpreting malaprops* (and more generally, a continuum argument about linguistic errors):

> *Continuum premise for malaprops*: The literal–loose–metaphorical continuum extends to malaprops.

> *Procedure conclusion for malaprops*: The relevance-theoretic comprehension procedure for malaprops does not require details beyond those needed to account for literal utterances.

We have good reason to resist the malaprop conclusion: surely, when we manage to interpret Mrs. Malaprop as having meant 'alligator' when she said 'allegory', the fact that the lexically encoded meaning of 'allegory' becomes wholly irrelevant is a detail that is bound to be featured in a full description of our process of interpreting her. And if we want to resist the malaprop conclusion, then we have to find fault with the continuum argument about interpreting malaprops. There are two strategies we could follow: we could fault the premise or fault the argument itself as non-truth-preserving. I will argue that the former strategy is not open to us, so our remaining option is to regard the malaprop argument as non-truth-preserving. But then (I shall argue) we have to say the same about the continuum argument about interpreting metaphors also. Whether the comprehension procedure for interpreting metaphors includes any details specific to metaphor (or nonliteral discourse) therefore remains an open question.

After some background on relevance theory (Section 2) and preliminary considerations about a recurring analogy-based reasoning strategy prevalent in the philosophical literature on metaphor (Section 3), I will formulate the continuum argument as a special instance of the analogy-based reasoning strategy (Section 4). I will then outline my malaprop objection (Section 5) and deflect a counterobjection to it (Section 6), concluding with some remarks on how the inferential comprehension procedure might be supplemented to include metaphor-specific details (Section 7).

2. *Relevance theory, the literal–loose–metaphorical continuum, and* ad hoc *concepts*

Sperber and Wilson's (1986/1995, also, Wilson and Sperber 2012) relevance-theoretic framework outlines an inferential comprehension procedure that hearers follow in arriving at an interpretation of speakers' linguistic utterances. Crucially, the comprehension procedure is delimited and guided by specific assumptions about relevance (i)–(iii), accepted by speakers and hearers alike. (i) Cognition (generally, not just in the case of communication) aims to maximize relevance (this is the *cognitive principle* of relevance). (ii) Linguistic utterances communicate a presumption of their own optimal relevance (this follows from the *communicative principle* of relevance[2]). And (iii) an utterance is *presumed to be optimally relevant* if and only if it is at least relevant enough to be worth the speaker's effort to process it, and it is the most relevant utterance compatible with the speaker's abilities and preferences. The kind of inference involved in the relevance-theoretic comprehension procedure is inference to the best explanation (Allot 2013). The concepts encoded by the words the speaker has used on a given occasion are mere starting points for arriving, via inferential steps, at an interpretation of her utterance: her utterance's explicit content (the speaker's explicit meaning) on the one hand, and its implicit content (which consists of implicit premises and conclusions) on the other.

By explicit and implicit content, we mean content that was *intended* as such by the speaker. The hearer's task is to *reconstruct* the explicit content and implicit premises and conclusions that the speaker has intended to communicate. Of course, rarely, if ever do hearers converge on the very same concepts as those that speakers actually meant. Nor is this required for successful communication. It suffices that the concepts reconstructed by the hearer be ones that allow him to draw (nearly enough) the same inferences as those intended by the speaker; it is enough that the reconstructed concepts "activate contextual implications that make the utterance relevant as expected" (Sperber and Wilson 2008: 110).

[2] Our concern here is with acts of *linguistic* communication, but the communicative principle and the relevance-theoretic framework are intended to apply to a broader range of cases: acts of ostensive communication which include, besides linguistic utterances, certain kinds of non-linguistic acts also.

A recurring example of Wilson and Sperber's (2002, see also Sperber and Wilson 2005, 2008) exemplifies *loose use*:

(1) Holland is flat.

uttered in the context of the following conversation: Peter and Mary are discussing their next cycling trip. Peter has just said that he feels rather unfit. Mary replies: "We could go to Holland. Holland is flat." Wilson and Sperber (2002) illustrate the inferential comprehension procedure via which Peter interprets Mary's second sentence as follows.

(a) Mary has said to Peter, 'Holland is flat'.	*Decoding of Mary's utterance.*
(b) Mary's utterance is optimally relevant to Peter.	*Expectation raised by the recognition of Mary's utterance as a communicative act, and acceptance of the presumption of relevance it automatically conveys.*
(c) Mary's utterance will achieve relevance by giving reasons for her proposal to go cycling in Holland, which take account of Peter's immediately preceding complaint that he feels rather unfit.	*Expectation raised by (b), together with the fact that such reasons would be most relevant to Peter at this point.*
(d) Cycling on relatively flatter terrain which involves little or no climbing is less strenuous, and would be enjoyable in the circumstances.	*First assumption to occur to Peter which, together with other appropriate premises, might satisfy expectation (c). Accepted as an implicit premise of Mary's utterance.*
(e) **Holland is FLAT*** (where FLAT* is the meaning indicated by 'flat', and is such that Holland's being FLAT* is relevant-as-expected in the context).	*(Description of) the first enriched interpretation of Mary's utterance as decoded in (a) to occur to Peter which might combine with (d) to lead to the satisfaction of (c).* ***Interpretation accepted as Mary's explicit meaning.***
(f) Cycling in Holland would involve little or no climbing.	*Inferred from (d) and (e). Accepted as an implicit conclusion of Mary's utterance.*
(g) Cycling in Holland would be less strenuous, and would be enjoyable in the circumstances.	*Inferred from (d) and (f), satisfying (b) and (c) and accepted as an implicit conclusion of Mary's utterance.*

Table 1. Interpretation of Mary's utterance 'Holland is flat'.

As indicated on line (e) (in boldface), the explicit content of Mary's utterance of (1) is 'Holland is FLAT*'. FLAT* is an *ad hoc concept* Peter arrived at that is distinct from, broader[3] than the lexicalized concept encoded by the word 'flat' in the given context of utterance: say, FLAT_1. Unlike FLAT*, the extension of FLAT_1 doesn't include imperfectly flat surfaces like the Dutch landscape.

Loose use, as in (1), is a type of literal discourse[4] that involves some departure from the lexically encoded concept. While the departure is greater than in many other instances of literal discourse, Sperber and Wilson (2008: 107) stress that the comprehension procedure for *some* literal utterances (to wit: cases of loose use and narrowing) already involves the formation of *ad hoc* concepts. They suggest further that even in literal utterances that do not involve a departure from the lexically encoded concept, the process of disambiguating the expressions used involves inferential steps similar to those in Table 1. For example, Mary's and Peter's idiolect may have (at least) two senses associated with the word 'flat', one of which amounts to, say, "having a smooth, even surface" while the other, to "is in a horizontal position"; Sperber and Wilson (2008: 111) suggest that if Mary uttered

(2) My computer screen is flat,

the process of interpreting her utterance and deciding that she has in mind the first and not the second sense of 'flat' would take a similar inferential procedure as the one seen in Table 1.

Sperber and Wilson (2008) gradually build up a continuum of cases with no clear boundaries in between them. The continuum includes cases of disambiguation like (2), various examples of

- loose use (or broadening), covering a broad range:
  - *Approximation*: 'Holland is flat';
  - *Limited category extension*: 'Here is a Kleenex', said of a piece of non-Kleenex-brand tissue;
  - *Creative category extension*: 'For luggage, pink is the new black';
- *Hyperbole*: 'Joan is the kindest person on earth';
- *Nonpoetic metaphor*: 'Joan is an angel';
- *Poetic metaphor*: 'The fog comes on little cat feet' (from Carl Sandburg's poem *The Fog*).

[3] Alternatively, according to another prominent relevance theorist, Robyn Carston (2002), the formation of *ad hoc* concepts involves conceptual narrowing as well as broadening. My malaprop objection can be straightforwardly adapted to work against Carston's proposal also.

[4] In their early work, Sperber and Wilson (1986/1995: 234) already stress the literal status of instances of loose use:

> [i]f someone says, It's 5 p.m., she should not be taken to task if it turns out to be five minutes or two minutes to, unless the relevance of the utterance depends on that kind of exactitude… The examples discussed so far would normally be treated as loose uses of language, but would not be regarded as figurative: there is no temptation to invoke the substitution of a figurative for a literal meaning.

A central claim of relevance theory (besides Sperber and Wilson's work, see also Carston 2002) is that each of the listed cases involves the formation of an *ad hoc* concept, one that—as we go down the list of examples—exhibits a gradually greater degree of departure from the concept lexically encoded by the word used, that is, the concept that serves as one of the starting points for the comprehension procedure. The *ad hoc* concepts are then featured as part of the explicit content attributed to the speaker (as in line (e) in Table 1). The *ad hoc* concepts for the listed examples (except for poetic metaphors, to be discussed in detail in Section 6) are as follows:

- FLAT*, whose extension includes imperfectly flat surfaces like the Dutch landscape;
- KLEENEX*, whose extension includes paper tissues that aren't Kleenex brand;
- BLACK*, whose extension includes (roughly) objects of a fashionable, trendy color, among them pink suitcases;
- KINDEST PERSON ON EARTH*, whose extension includes people who are very kind, but not even close to being among the kindest;
- ANGEL*, whose extension includes nonangelic human beings who are very kind.

3. *A recurring analogy-based reasoning strategy in the metaphor literature*

Imagine us in the middle of a discussion about Woody Allen's 2005 film *Match Point*; I contribute the following metaphorical[5] utterance:

(3) The film's plot is flat.

It is customary to distinguish the *literal import* of such an utterance (which is quite outlandish: about a non-concrete thing like a story line *literally* having an even surface) and its *metaphorical import* (roughly: that the story line of the film lacks imagination / is banal, prosaic / is predictable and simple-minded / is without complexity, layers or depth).

A recurring motif in theories of metaphor is the following analogy-based reasoning strategy.

[5] Arguably, 'flat' as used in (3) has become a dead metaphor, and the sense of 'flat' as dull and prosaic has become lexicalized over time, reflected in major dictionaries. I use this example to keep it close to Sperber–Wilson's 'Holland is flat'. I ask those readers who think (3) is not a metaphorical utterance but contains a dead metaphor to (a) substitute one of the usual examples in (3)'s place like 'Sam is a bulldozer' or 'My chiropractor is a magician', and / or (b) consider that numerous major dictionaries construe these very uses of 'bulldozer' and 'magician' as dead metaphors also, listing as one of their senses, respectively "an overbearing person, a bully" (American Heritage), and "a person who has amazing skills" (Merriam-Webster).

ANALOGY-BASED REASONING STRATEGY
Premise 1: Metaphorical import and *X*-type import resemble one another in a theoretically important respect: both are *P*.
Premise 2: *X*-type import is obviously *Q*.
Conclusion: We have good reason to think that metaphorical import is *Q* also.

We can think of *P* predicates as *premise predicates* (featured in Premise 1) and *Q* predicates as *conclusion predicates* that, based on the analogy at hand, are true of metaphorical import. The analogy-based reasoning strategy is of key importance in Grice's as well as Davidson's arguments about metaphor. Let's take a brief look at these to illustrate the strategy at work before considering the role of this strategy in the context of relevance theory.

In the case of Grice's (1975) theory of metaphor, we can substitute the following for *X*, premise predicates *P* and conclusion predicates *Q*:

ANALOGY-BASED REASONING STRATEGY IN GRICE
X = conversational implicature,
P_1 = unlike what the speaker has (strictly speaking) said (or made as if to have said)
Q_1 = derivable based on conversation-guiding norms (such as the Cooperative Principle and the maxims)
Q_2 = propositional, can be characterized in terms of truth conditions
(P_1, Q_1 and Q_2 are illustrations, we could go on and list further instances of premise predicates and conclusion predicates at play. Also, one might question the sorting of premise predicates and conclusion predicates: in the context of a specific Gricean argument / interpretation thereof. For example, Q_2 could be construed as a premise predicate also.)

This way, in the Gricean reasoning about metaphor, conversational implicatures play a central role: what drives Grice's argument is that metaphorical import is like conversational implicatures in certain (theoretically important) respects, so in other respects they are alike also. We can say that Grice's theory of metaphor crucially relies on the analogy between metaphor and conversational implicatures.

In the case of Davidson's (1978) theory of metaphor, the following can substituted for *X*, premise predicates *P* and conclusion predicates *Q*:[6]

ANALOGY-BASED REASONING STRATEGY IN DAVIDSON:
X = the import/purpose/communicative function of a joke
P_1 =–"can make us appreciate some fact, … but not by standing for or expressing that fact" (Davidson 1978: 46)[7]
P_2 = not propositional, cannot be characterized in terms of truth conditions

[6] This strategy is employed in Lepore and Stone (2010) neo-Davidsonian proposal also.

[7] Notice the similarity between the Gricean and the Davidsonian P_1 predicates.

Q_1 = cannot be characterized as some sort of meaning or content that is distinct from literal meaning/content

Q_2 = is based on the idea that sometimes we use words in such a way that we are relying on their *conventional* meanings in order to make our audience see one thing (for example, a film's plot) as something else (something with an even surface)

(P_1, P_2, Q_1 and Q_2 are illustrations, we could go on and list further instances of premise predicates and conclusion predicates at play. Also, one might, again, question the sorting of premise predicates and conclusion predicates: for example, P_1 could also be construed as a conclusion predicate in specific Davidsonian arguments or interpretations thereof.)

This way, in Davidson's as well as neo-Davidsonian theorists' reasoning (for example Lepore and Stone 2010) jokes play a central role: what drives their arguments is that metaphorical import is like the import or purpose of a joke in certain (theoretically important) respects (particularly P_2), so in other respects they are alike also. We can say that Davidsonian theories of metaphor crucially rely on the analogy between metaphor and jokes.

From its inception, Sperber and Wilson's relevance theory has prominently featured the analogy-based reasoning strategy (1986/1995, 1990, 2004, 2005, 2008, also, Wilson and Sperber 2002). And their X is none other than cases of loose use or broadening, as in (1) ('Holland is flat').

Analogy-based reasoning strategy in Sperber–Wilson, with brief commentary:

X = the import (explicit content) of utterances involving loose use

P_1= propositional, can be captured in terms of truth conditions

P_2 = is on the same continuum with other instances of language use traditionally thought to be literal,[8] loose, hypberbolic and metaphorical: it is only a matter of degree to what extent the lexically encoded concept is being broadened to yield the concept featured in the explicit content of the utterance

Q_1 = is arrived at via the process of constructing an *ad hoc* concept

- Recall that an *ad hoc* concept is broader than the lexically encoded concept associated with the expression uttered (which is no more than a departure point for the comprehension process).

[8] The upshot of Sperber and Wilson (2008) is that the traditional notion of literal doesn't serve a theoretical purpose in a theory of communication: the category of literal utterances turns out to encompass loose use, hyperbole as well as metaphor (poetic as well as nonpoetic). When they talk about a literal–loose–metaphorical continuum, it would be odd to apply their own sense of 'literal': for then the first label would be an overarching one that covers cases of the other two also: loose use and metaphor. Instead, plausibly, when Sperber and Wilson discuss the 'literal–loose–metaphorical' continuum, they are using 'literal' in the traditional sense instead. This way, literal (in the traditional sense) does not include cases of loose use or metaphor.

- For example, according to Sperber and Wilson, the explicit content of (1) is 'Holland is FLAT*' (as in Table 1), where FLAT* is an *ad hoc* concept whose extension includes *approximately* flat things also, ones that are not included in the extension of the lexically encoded concept for the given use of flat, say, FLAT_1. Holland is an example of something included in the extension of FLAT* but not FLAT_1.

Q_2 = is interpreted via an inferential comprehension procedure that contains only such details that are already needed to interpret literal utterances

Q_3 = involves no departure from literal discourse

- Importantly, as we have seen above, Sperber and Wilson consider it evident that (1) (featuring loose use) is an instance of literal language use.

(In what follows, I will set aside Q_3: even though it is a prominent claim in Sperber and Wilson's argument, it doesn't add to the specific argument of theirs that I am about to examine, which already mentions literality in Q_2.) Notice that here, again, we could construe differently which predicates count as premise predicates and which as conclusion predicates; for example, Q_1 could be construed as a premise predicate instead.)

4. *The continuum argument about interpreting metaphors*

We are now in a position to formulate in far more depth and detail Sperber and Wilson's (and other relevance theorists') central argument about interpreting metaphors.

SPERBER–WILSON'S ANALOGY-BASED REASONING ABOUT METAPHOR

Premise 1: Metaphorical import and the content of loose use resemble one another in a theoretically important respect, P_2: they are on the same continuum with other instances of language use traditionally thought to be literal (including loose use) as well as hyperbolic, metaphorical language use.

Premise 2: The content of loose use is obviously

Q_1 (is arrived at via the process of constructing an *ad hoc* concept), and

Q_2 (is interpreted via a relevance-theoretic inferential procedure that contains only such details that are already needed to interpret literal utterances).

Conclusion: We have good reason to think that metaphorical import is

Q_1 (is arrived at via the process of constructing an *ad hoc* concept), and

Q_2 (is interpreted via an inferential comprehension procedure that contains only such details that are already needed to inter pret literal utterances).

For exposition, let us simplify, relabel and reword things a bit:

> THE CONTINUUM ARGUMENT ABOUT INTERPRETING METAPHORS (final version)
> *Premise 1. The continuum premise for metaphors*:
> All metaphorical utterances (poetic and nonpoetic alike) can be located on a continuum of cases that includes loose use (a kind of literal use) as well as hyperbolic and metaphorical uses.
>
> *Premise 2.* Ad-hoc-*concept premise for metaphors:*
> The process of forming *ad hoc* concepts to arrive at the explicit content attributed to the speaker is all that the inferential comprehension procedure requires in order to capture the process of interpreting instances of loose use.
>
> *Procedure conclusion for metaphors:*
> The process of forming *ad hoc* concepts to arrive at the explicit content attributed to the speaker is all that the inferential comprehension procedure requires in order to capture the process of interpreting metaphorical utterances.

The upshot of the procedure conclusion is this: equipped with the relevance-theoretic comprehension procedure and the *ad hoc* concept formation tool, both already required for interpreting literal utterances like loose use, we have all the resources needed to describe the comprehension procedure at play during the interpretation of metaphorical utterances. No further details specific to metaphor (or figurative language use) are needed in a comprehensive account of interpreting metaphors.

In Section 5, I will raise an objection that purports to show that the continuum argument about interpreting metaphors is flawed: even if we accepted both of its premises, that is not reason enough to accept its conclusion also. I will motivate this by giving what I think is an entirely parallel argument about malaprops—the *malaprop objection*—with a clearly false conclusion. Someone might then raise a counterobjection: the argument about malaprops has a false conclusion because it has either a false premise or fails to provide a strictly parallel argument. We have two options, the counterobjection goes. (i) We can maintain (as relevance theorists do) the continuum premise for metaphors while resisting its counterpart about malaprops. Or (ii) we can claim that the two arguments are not entirely parallel after all because of the specifics of the *ad hoc* concept tool. In Section 6, I will elaborate this counterobjection and deflect it by showing that (i) the continuum premise for metaphors is no more plausible than its counterpart for malaprops, and (ii) relevance theorists' *ad hoc* concept tool is highly unconstrained and powerful, so nothing about its specifics prevents its extension to malaprops; therefore, the metaphor and malaprop arguments are exactly parallel after all. The upshot: the malaprop objection has traction and there is room to reject the procedure conclusion for metaphors, despite relevance theorists' arguments to the contrary.

5. *The malaprop objection*

Once we have accepted the continuum argument about interpreting metaphors, along with its premises and conclusion, we have, I claim, no reason to resist making the same moves with respect to a parallel argument about malaprops (and more generally, about linguistic errors):

THE CONTINUUM ARGUMENT ABOUT INTERPRETING MALAPROPS
Premise 1. The continuum premise for malaprops:
All malaprops can be located on a continuum of cases that includes loose use (a kind of literal use) as well as hyperbolic and metaphorical uses.

Premise 2. Ad-hoc-*concept premise for malaprops:*
The process of forming *ad hoc* concepts to arrive at the explicit content attributed to the speaker is all that the inferential comprehension procedure requires in order to capture the process of interpreting instances of loose use. (Same as the previous Premise 2 for metaphors.)

Procedure conclusion for malaprops:
The process of forming *ad hoc* concepts to arrive at the explicit content attributed to the speaker is all that the inferential comprehension procedure requires in order to capture the process of interpreting malaprops.

The upshot of the procedure conclusion for malaprops is this: equipped with the relevance-theoretic comprehension procedure and the *ad hoc* concept formation tool, both already required for interpreting literal utterances like loose use, we have all the resources needed to describe the comprehension procedure at play during the interpretation of malaprops. No further details specific to slips of the tongue (or more broadly: linguistic errors) are needed in a comprehensive account of interpreting malaprops.

But—the *malaprop objection* goes—there is a flaw in this argument: (a) its conclusion is clearly unacceptable and (b) it remains unacceptable even if we accept its premises. And if we accept all this, we have exposed a flaw in the original continuum argument about interpreting *metaphors*. In the rest of this section, I aim to establish (a), in the next section, (b).

The procedure conclusion for malaprops leads to the following bizarre results:

- *Allegory example.* In interpreting Mrs. Malaprop's utterance "She is as headstrong as an allegory on the banks of the Nile", the explicit content that hearers arrive at involves an *ad hoc* concept ALLEGORY*, which is constructed by broadening the concept lexically encoded by the word 'allegory' (about a certain kind of trope or figure of speech) in such a way that its extension includes *alligators*. The comprehension procedure is basically the same as that in Table 1, it's just that the degree of departure to get from FLAT$_1$ to FLAT* is not as great as that from ALLEGORY$_1$ to ALLEGORY*.

- *Spanking example.* In interpreting George W. Bush's utterance in the context of a speech he gave at a school "I want to spank all teachers" (he meant *thank all teachers*), the explicit content that hearers arrive at involves an *ad hoc* concept SPANK*, which is constructed by broadening the concept lexically encoded by the word 'spank' (about slapping) in such a way that its extension includes acts of *thanking*. The comprehension procedure is basically the same as that in Table 1, it's just that the degree of departure to get from FLAT$_1$ to FLAT* is not as great as that from SPANK$_1$ to SPANK*.

As mentioned before, the continuum argument about malaprops is readily extended to linguistic errors of all sorts, including slips of the tongue other than malaprops as well as mistaken translations like the following:

- *Steak example.* In interpreting a German speaker's order in a restaurant "I want to become a steak" ('bekommen' in German means 'get'), the explicit content that hearers arrive at involves the *ad hoc* concept BECOME*, which is constructed by broadening the concept lexically encoded by the word 'become' in English (about 'turning into') in such a way that its extension includes one thing *getting* another. The comprehension procedure is basically the same as that in Table 1, it's just that the degree of departure to get from FLAT$_1$ to FLAT* is not as great as that from BECOME$_1$ to BECOME*.

It is bizarre to think that when we manage to interpret successfully the German speaker's request to "become a steak", we are broadening the concept lexically encoded by the English word 'become'. After all, our grasping that he's talking about getting a steak rather than turning into one happens *despite* his use of the English word 'become'. We can say the same about understanding Mrs. Malaprop's and George W. Bush's utterances: it is *despite* the lexically encoded meaning of the words they have used that we manage to interpret them as having said something about alligators and thanking, respectively.

In the light of this, it seems that relevance-theoretic comprehension procedures, as they stand, are missing key details that distinguish malaprops (and more broadly, linguistic errors) from utterances that are literal or metaphorical. To wit: the procedure has to specify that in utterances like 'Holland is flat', 'Joan is an angel' (loose and metaphorical uses alike), the speaker has *not* committed a linguistic error; further, that the speaker (and hearer) takes the lexically encoded concept associated with her words to be in force, and would not retract her words when confronted with the concept lexically encoded by her words. By contrast, in the case of linguistic errors including malaprops, the hearer is rerouting the inference such that he sets aside the lexically encoded concept entirely, and the speaker, when confronted with the lexically encoded concept, would retract his or her words: “I didn’t mean spanking teachers was desirable, I wanted to talk about thanking them.” “I didn’t mean there were allegories on the banks of the Nile, I wanted to talk about alligators”. But we would have absolutely no grounds for seeking such additional details if we thought the continuum argument about malaprops worked and moreover featured true premises. If, despite the argument about malaprops, we thought the additional details were needed, then we open the door to seeking additional details with which to supplement the comprehension procedure for metaphorical utterances also. And we thereby open the door to rejecting the conclusion of the continuum argument about metaphors.

An analogy helps illuminate what my objection, if successful, shows with respect to Sperber and Wilson’s continuum argument about interpreting metaphors. If you are at Columbus Circle in Manhattan and want to take the subway to the Museum of Natural History (at 81st Street), then don’t get on the A train (the 8th Avenue Express); despite the fact that you would initially approach your desired destination, eventually, your train would whizz right past the Museum of Natural History, taking you all the way to 125th Street in Harlem, far away from your desired destination. Likewise: if you don’t want an inferential comprehension procedure for malaprops (and other linguistic errors) that invokes no more than the formation of *ad hoc* concepts at work in the comprehension procedure you posited for cases of loose use, then don’t apply the continuum argument to metaphorical utterances, for you won’t be able to get off there but will be whisked straight to a place where you don’t want to be: the continuum argument about interpreting malaprops.

6. *A counterobjection deflected*

It seems natural to respond to the foregoing objection as follows: a distinguishing feature of linguistic errors, malaprops included, is that the speaker makes a mistake about which *word form* is associated with the lexically encoded concept that he or she wants to express: G. W. Bush has said ‘spank’ even though his intended concept is expressed by the word form ‘thank’; Mrs. Malaprop has said ‘allegory’ even though her intended concept is expressed by the word form ‘alligator’. Proponents of this counterobjection may then claim: of course the swapping of word forms, and the fact that the hearer recognizes the swap and reroutes the inference accordingly, will be part of the comprehension procedure via which he interprets malaprops and the like. We are in no way forced to regard the alligator, spanking and steak examples as cases involving simply the formation of *ad hoc* concepts with extreme degrees of departure from the lexically encoded concepts that had served as starting points for the construction of the *ad hoc* concept. This is how the counterobjection goes.

Someone could maintain this line while holding on to the continuum argument about *metaphors* and its conclusion, by denying either (i) the first premise of the continuum argument about *malaprops* or (ii) the link between the second premise and conclusion of the malaprop argument (without undermining the link in the metaphor argument). This would amount to showing either (i) that—in the context of relevance theory—extending the literal–metaphorical continuum to malaprops (and other linguistic errors) is unfounded, or (ii) that—again, in the context of relevance theory—the tool of *ad hoc* concept construction is such that it is readily applicable to loose use and metaphors but not to malaprops and other linguistic errors (the nature of the tool would then be such that it would license the transition from premises to conclusion in the metaphor argument but not the malaprop argument). In what follows, I will show that neither (i) nor (ii) will work and hence the counterobjection fails. My response consists of two parts:

> (i) With respect to malaprops (and other linguistic errors also) we can talk about a continuum of cases ranging from limited to extreme degrees of discrepancy between the intended concept and the lexically encoded one. And the limited-discrepancy cases can be readily placed on the literal–metaphorical continuum Sperber and Wilson had posited.
>
> (ii) In the case of poetic metaphors, the *ad hoc* concept departs greatly from the lexically encoded one, yet Sperber and Wilson (and others) do not doubt that here, too, explicit content is arrived at via the construction of an *ad hoc* concept. If the *ad hoc* concept tool is capable of that, then it is plausibly suited for capturing cases like the allegory, spanking and steak examples also. Meanwhile, the limited-discrepancy cases fit squarely the *ad hoc* concept formation paradigm. To resist these moves, substantial constraints would need to be in place about the *ad hoc* concept formation tool.

I discuss (i) and (ii) in reverse order.

(ii) is in part about poetic metaphors. We've already encountered the example from Sandburg's poem "The fog comes on little cat feet". According to Sperber and Wilson, the explicit content arrived at in the comprehension procedure for interpreting this line of the poem involves the *ad hoc* concept: ON-LITTLE-CAT-FEET*. What Sperber and Wilson say about this concept signifies that it involves a great degree of departure from the lexically encoded concept: the *ad hoc* concept is supposed to help convey that the fog is spreading in a smooth, quiet, stealthy and deliberate way. Yet it remains quite vague what this *ad hoc* concept is, in what direction it takes off from the lexicalized concept, what does and does not belong in its extension. The authors offer us limited guidance on these matters: ON-LITTLE-CAT-FEET*…

> is the concept of a property that is difficult or impossible to define, a property possessed in particular by some typical movements of cats (though not

all of them—little cat feet can also move in violent or playful ways) and, according to the poem, by the fog. (Sperber and Wilson 28: 122).

As Sperber and Wilson see it, the great distance between lexicalized and *ad hoc* concepts and the vague description of the latter is no obstacle to applying the *ad hoc* concept formation paradigm to highly creative, poetic metaphors. Then comparably great distances and vagueness characterizing ALLEGORY* (whose extension includes certain reptiles) and SPANK* (whose extension includes acts of thanking) should be no obstacle to applying the *ad hoc* concept formation paradigm to malaprops (and other linguistic errors).

(ii) is also about examples involving limited-discrepancy between the encoded concept and the intended one. These examples fit squarely within the *ad hoc* concept formation paradigm, comparable to the "Here is a Kleenex" and "For luggage, pink is the new black" type examples.

> *Ocean example* (a slip of the tongue involving limited discrepancy). G. W. Bush said once: "I didn't grow up in the ocean—as a matter of fact—near the ocean—I grew up in the desert. Therefore, it was a pleasant contrast to see the ocean. And I particularly like it when I'm fishing." In interpreting the first portion of Bush's utterance, via *ad hoc* concept formation, from the encoded lexical meaning IN-THE-OCEAN$_1$, we arrive, by broadening, to IN-THE-OCEAN*, whose extension includes events and things *near* the ocean.
>
> *Library example* (a mistaken translation involving limited discrepancy). A French speaker says: "There is a library around the corner" to mean that there is *bookshop* around the corner (in French 'libraire' means bookshop). In interpreting the utterance, via *ad hoc* concept formation, from the encoded lexical meaning of LIBRARY$_1$, we arrive, by broadening, to LIBRARY*, whose extension includes bookshops. (Such an utterance could also exemplify a slip of the tongue involving limited discrepancy.)

In the ocean example, the distance between IN-THE-OCEAN$_1$ and IN-THE-OCEAN* is no greater and no less vaguely delineated than that between KLEENEX$_1$ and KLEENEX*. The same can be said about LIBRARY$_1$ and LIBRARY* also.

The underdefined nature of the process of *ad hoc* concept formation (as observed with ON-LITTLE-CAT-FEET*) makes it even clearer that much too little is settled about this tool (what it can and cannot do) to prevent its application to the gradually greater departures we find in the examples of linguistic error spanning from the library and ocean examples to the allegory, spanking and steak ones.

Turning to (i): limited-discrepancy examples of linguistic error (like the library and ocean examples) already suggest that we can plausibly construct a continuum of examples spanning from such examples to the extreme-discrepancy ones (like the allegory, spanking and steak examples). Further, the limited-discrepancy examples plausibly fit right onto the literal–loose–metaphorical continuum. Given all this, the limited-discrepancy examples make clear that considerations about a literal–metaphorical continuum support a literal–metaphorical–linguistic-error continuum also.

This concludes my justification for (i) and (ii), which together show that the counterobjection about swapped word forms does not undermine the malaprop objection I had formulated against the continuum argument about interpreting metaphorical utterances. After all, the limited-discrepancy examples of linguistic error make clear that the continuum premise for malaprops (and other linguistic mistakes) is just as plausible as the continuum premise for metaphors. In addition, the lack of constraints on the *ad hoc* concept formation tool makes clear two things: that nothing prevents its application to gradually greater and less determinate departures from what is lexically encoded and that there is no principled reason for deeming the tool fit to handle highly poetic metaphors but not linguistic errors. We therefore have at hand two entirely parallel arguments, both with true premises, and the one about malaprops boasting a clearly false conclusion. Hence, the other argument, about metaphors, is also undermined: the truth of its premise is no guarantee for the truth of its conclusion.

7. *Concluding remarks*

The continuum argument about interpreting metaphorical utterances is central to Sperber–Wilson's conclusion that "the same interpretive abilities and procedures" are at play in the case of loose use as well as metaphor (poetic and nonpoetic) (Sperber and Wilson 1986/1995: 235, see also their 2008: 97, and Wilson and Carston 2007: 231). My aim has been to show that we need not accept this conclusion given that the continuum argument about interpreting metaphors is flawed, as shown by its application to malaprops (and other linguistic errors).

In the wake of the malaprop objection to the continuum argument, several questions arise.

First, what shall we make of empirical considerations about metaphor processing, according to which, for example, the interpretation procedure for simpler metaphors is similar to that for literal utterances, while interpreting highly creative or novel metaphors involves a markedly different procedure (for example, Gibbs 1994)?[9] The dialectical situation is as follows: such considerations support or undermine, *independently of the continuum argument about interpreting*

[9] More recent experimental results (for example Forgács, Lukács and Pléh 2014) cast doubt on earlier views positing a marked difference in the processing of novel metaphors and literal utterances. Carston (2010), a central figure of relevance theory, parts ways with Sperber and Wilson (2008) and posits two distinct modes of processing metaphorical utterances. Her distinction provides the basis for one way of incorporating metaphor-specific generalizations in relevance-theoretic comprehension procedures.

metaphors, the claim that a similar comprehension procedure applies to literal utterances and certain types of metaphorical utterances. The continuum argument doesn't—cannot—provide an objection to or further support for such claims, because (as I have tried to argue, successfully, I hope) if it were to work, it would show too much, so it doesn't work. Therefore the tenability of the claim about a literal–loose–metaphorical continuum and the application of *ad hoc* concept formation in the interpretation of metaphorical utterances will depend on *other* (experimental-data-driven) arguments.

Second, how might relevance theorists maintain the procedure conclusion about interpreting metaphors? They can, in response to the malaprop objection, explore two options. On the one hand, they may fill in various details about the nature of the literal–loose–metaphorical continuum in a way that makes clear why the continuum cannot extend to malaprops. On the other hand, they may fill in various details about and constraints on the process of constructing *ad hoc* concepts in a way that makes clear why this tool is applicable to poetic metaphors but inapplicable to malaprops. By framing the continuum argument and the malaprop objection as a special case of an analogy-based reasoning strategy, I hope to have provided a useful backdrop for clarifying the challenge confronting relevance theorists who are keen on preserving the procedure conclusion.

Third, it is worth considering an alternative approach: what options lie ahead if we decide to give up the procedure conclusion. This involves formulating a positive proposal about how to supplement the relevance-theoretic comprehension procedure for interpreting metaphors. I address this question in work in progress (Zvolenszky Ms.), drawing in part on some of the considerations that provide missing details with which to supplement the comprehension procedure for interpreting malaprops and other linguistic errors (these were briefly discussed in Section 5). In the case of metaphorical utterances (but not malaprops), the speaker (and hearer) takes the lexically encoded concept associated with her words to be in force, and would not retract her words when confronted with the concept lexically encoded by her words. "The fog doesn't really walk on feline legs," someone might challenge the poet. And he might reply: "I was speaking metaphorically. But I stand by my words: The fog does come on little cat feet". Notice that the poet could not say (instead of: I was speaking metaphorically) "I was speaking *loosely"*; loose use does not license the poem's sort of departure from the lexically encoded concepts at hand. This point gives preliminary motivation for resisting exactly alike treatment (and exactly alike comprehension procedures) for loose use and metaphor. Meanwhile, loose use as well as metaphor are markedly different from malaprops (and linguistic errors): Mrs. Malaprop, when challenged, "There are no such things as pineapples of politeness," would (likely) respond: "I *retract* my previous words; I meant to speak about a *pinnacle* of politeness. Some of the commitments my original words had accrued were *inadvertent* and I now reject them."[10] Such differences in the response to being challenged about the lexically encoded concepts associated with one's words—in instances of metaphor, loose use and malapropism—do, I think, offer a promising starting point for the sorts of details that a relevance-theoretic comprehension procedure can incorporate in an account of metaphor. Such an account would part ways with Sperber–Wilson's stance, claiming instead that there are, after all, interesting details and generalizations specific to metaphors. More generally: within relevance theory (and any theory of communication), the various ways in which lexically encoded concepts systematically constrain speakers' meaning is a worthy area for in-depth exploration, whether an utterance involves deliberate departure from the lexically encoded concept (as in loose, hyperbolic and metaphorical utterances) or inadvertent departure (as in linguistic errors like malaprops).[11]

References

Allott, N. 2013. "Relevance Theory". In A. Capone, F.L. Piparo, M. Carapezza (eds.). *Perspectives on Linguistic Pragmatics* (*Perspectives in Pragmatics, Philosophy & Psychology 2*). Berlin: Springer:57–98.

Camp, E. 2008. "Showing, Telling and Seeing". *The Baltic Yearbook of Cognition, Logic and Communication* 3: 1–24.

Camp, E. 2012. "Sarcasm, Pretense, and the Semantics/Pragmatics Distinction". *Noûs* 46 (4): 587–634.

[10] See Camp (2008, 2012) about how deniability reveals distinctive features of metaphorical utterances.

[11] I have received many incisive comments in connection with this research project. I thank audiences and organizers at three conferences (Philosophy of Linguistics and Language X, in the Special Session on Dan Sperber and Deirdre Wilson's Philosophy of Language, Interuniversity Center Dubrovnik, September, 2014; Meaning and Experience (a conference in Hungarian) held at Kaposvár University, January 2015; the 8th AISB Symposium on Computing and Philosophy: The Significance of Metaphor and Other Figurative Modes of Expression and Thought, AISB Convention 2015 (The Society for the Study of Artificial Intelligence and Behavior), University of Kent, Canterbury, April 2015), one workshop (Metaphor workshop, in Hungarian, organized by the Hungarian Coaches Association and Erasmus Collegium's Language Research Group, Budapest, May, 2013), and two invited presentations (the MASZAT Colloquium Series (Round Table Society of Hungarian Semanticists), MTA Research Institute for Linguistics, Budapest, September 2014; the Institute of Philosophy of the Czech Academy of Sciences, Prague, October 2014). Special thanks are due to Nicholas Allott, Tibor Bárány, John Barnden, Ágnes Bende-Farkas, Robyn Carston, István Danka, Ingrid Lossius Falkum, Bálint Forgács, Hans-Martin Gaertner, Michael Glanzberg, Dunja Jutronić, Tamar Kalandadze-Tabakhmelashvili, Karen Lewis, Nenad Miščević, Kim Phillips Pedersen, Mihaela Popa, Paolo Santorio, Adam Sennet; and, last yet foremost, Craige Roberts, Dan Sperber and Deirdre Wilson for invaluable comments, help and encouragement. The present research was supported by three grants received from the Hungarian Scientific Research Fund (OTKA): Grant No. K–109456 entitled "Integrative Argumentation Studies", Grant No. K–109638 entitled "Connections Between Analytic Philosophy and Phenomenology Within the Philosophy of Mind", and Grant No. K–116191 entitled "Meaning, Communication; Literal, Figurative: Contemporary Issues in Philosophy of Language".

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

Carston, R. 2010. "Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images". *Proceedings of the Aristotelian Society* 110: 295–321.

Davidson, D. 1978. "What Metaphors Mean". *Critical Inquiry* 5 (1): 31–47.

Forgács, B., Lukács Á. and Pléh, C. 2014. "Lateralized Processing of Novel Metaphors: Disentangling Figurativeness and Novelty". *Neuropsychologia* 56: 101–109.

Gibbs, R. W. 1994. *The Poetics of Mind: Figurative Thought, Language and Understanding*. Cambridge: Cambridge University Press.

Grice, H. P. 1975. "Logic and Conversation". In P. Cole and J. L. Morgan (eds.). *Syntax and Semantics 3: Speech Acts*: 41–58.

Lepore, E. and Stone, M. 2010. "Against Metaphorical Meaning". *Topoi* 29: 165–180.

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

Sperber, D. and Wilson, D. 2005. "Pragmatics". In F. Jackson, M. Smith (eds.). *Oxford Handbook of Contemporary Analytic Philosophy*. New York: Oxford University Press: 468–501. (Reprinted in Wilson and Sperber 2012.)

Sperber, D. and Wilson, D. 2008. "A Deflationary Account of Metaphors". In R.W. Gibbs (ed.). *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press: 84–105. (Reprinted in Wilson and Sperber 2012. Page numbers are to this edition.)

Wilson, D. and Carston, R. 2007. "A unitary approach to lexical pragmatics: relevance, inference and *ad hoc* concepts". In Noel-Burton Roberts (ed.). *Pragmatics*. London: Palgrave: 230–259.

Wilson, D. and Sperber, D. 2012. *Meaning and Relevance*. Cambridge: Cambridge University Press.

Wilson, D. and Sperber, D. 2002. "Truthfulness and Relevance". *Mind* 111: 583–632. (Reprinted in Wilson and Sperber 2012.)

Zvolenszky, Z. 2015. "Relevance Theoretic Comprehension Procedures". In *Proceedings for the 8th AISB Symposium on Computing and Philosophy: The Significance of Metaphor and Other Figurative Modes of Expression and Thought, AISB Convention 2015* (The Society for the Study of Artificial Intelligence and Behavior), University of Kent, Canterbury, http://www.cs.kent.ac.uk/events/2015/AISB2015/proceedings/philosophy/zvolenszky.pdf

Zvolenszky, Z. Ms. "Why Inflate Sperber & Wilson's Account of Metaphor?" Manuscript, Department of Logic, Institute of Philosophy, Eötvös Loránd University (ELTE), Budapest, Hungary.

Croatian Journal of Philosophy
Vol. XV, No. 44, 2015

Pragmatics and Epistemic Vigilance: The Deployment of Sophisticated Interpretative Strategies

DIANA MAZZARELLA
Laboratoire sur le Langage, le Cerveau et la Cognition (L2C2), CNRS-Université de Lyon, France
University College London, United Kingdom

Sperber (1994) suggests that competent hearers can deploy sophisticated interpretative strategies in order to cope with deliberate deception or to avoid misunderstandings due to speaker's incompetence. This paper investigates the cognitive underpinnings of sophisticated interpretative strategies and suggests that they emerge from the interaction between a relevance-guided comprehension procedure and epistemic vigilance mechanisms. My proposal sheds a new light on the relationship between comprehension and epistemic assessment. While epistemic vigilance mechanisms are typically assumed to assess the believability of the output of the comprehension system (Sperber et al. 2010), I argue that epistemic assessment plays an additional role in determining this very output.

Keywords: Pragmatics, epistemic vigilance, deception, misunderstanding.

1. *Communication and speaker's intentions*

Current research on linguistic communication is grounded on the well-established assumption that "[h]uman communication is characterised, among other things, by the fact that communicators have two distinct goals: to be understood, and to make their audience think or act according to what is to be understood" (Sperber et al. 2010: 364). This is captured by Sperber and Wilson's (1986/1995) definition of ostensive-inferential communication. According to them, ostensive-inferential communication takes place when communicators produce an utterance (or any other ostensive stimulus) with the following two intentions: (i) an *informative intention* to inform the audience of something, and (ii) a *communicative intention* to inform the audience of one's informative intention (Sperber and Wilson 1986/1995: 29).[1] More technically, the *informative intention* is defined as the intention to make manifest or more manifest to the audience an array of propositions I, whereas the *communicative intention* is defined as the intention to make it mutually manifest to audience and communicator that the communicator has this informative intention.[2]

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

Crucially, an audience can understand an utterance without believing what they have understood. In this case, the communicative intention is fulfilled without the corresponding informative intention being fulfilled. As Wilson and Sperber (2004: 611) suggest, typically "[w]hether the informative intention itself is fulfilled depends on how much the audience trusts the communicator". While understanding is underpinned by the pragmatic ability to recover the speaker's communicated meaning on the basis of linguistic and contextual cues, the epistemic assessment which leads to its acceptance (or rejection) is carried out by what Sperber et al. (2010) call 'epistemic vigilance'.

Before exploring the nature of epistemic vigilance mechanisms, I intend to show that the possibility of comprehension without acceptance/belief is warranted by the definition of communicative and informative intentions provided by Sperber and Wilson. To begin with, it is worth noting that their definition of manifestness is epistemic in nature: a proposition is manifest to an individual at a given time to the extent that he is likely, to some positive degree, to entertain it and *accept it as true*. Given that communicative and informative intentions are conceived of as intentions to change the degree of manifestness of certain

[1] In contrast with Grice (1957), Sperber and Wilson reject the idea that the communicator must have a third-level intention that the addressee's recognition of her informative intention should be at least part of his reason for fulfilling this. This allows Sperber and Wilson's account of ostensive-inferential communication to cover the whole continuum from pure cases of 'showing' to pure cases of 'meaning'. See also Sperber and Wilson (2015).

[2] See Sperber and Wilson (2015) for a detailed discussion on the notion of 'manifestness'.

assumptions, the following question seems to arise. Is the rejection of the communicated content compatible with the idea that manifestness involves acceptance on the interpreter's part?[3] In order to answer this question, it is worth stressing that in those circumstances in which the addressee understands a piece of communicated information without ending up believing it, the communicative intention is fulfilled without the corresponding informative intention being fulfilled. The informative intention is *recognised* (rather than fulfilled). The fulfilment of the communicative intention entails that the fact that the communicator has a certain informative intention, let's call it "i_i", is made mutually manifest. That is, the fact that the communicator has the intention i_i is likely to be accepted as true by the audience (and the communicator) and this is itself manifest. This does not seem to be problematic: while the audience is likely to accept as true the fact that the speaker has the intention i_i (i.e. the intention to make manifest or more manifest to the audience an array of propositions I_i), the audience is not bound to accept as true any of the propositions which are included in I_i. The audience accepting as true the array I_i would correspond to the fulfilment of the informative intention (but it is not a condition for its recognition).

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

2. *Epistemic vigilance*

Sperber et al. (2010) suggest that humans have developed "a suite of cognitive mechanisms", which is targeted at the risk of misinformation in communication. Each of them 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 interactions suggests it is), humans have to deploy an ability to calibrate their epistemic trust. This ability is 'epistemic vigilance'.

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

Sperber et al. (2010) conceive of epistemic vigilance as a cognitive adaptation for social exchange. As Cosmides and Tooby (1992: 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.[4]

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

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

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

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

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

[4] 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.

The reliability of the source of information, however, is not the only factor affecting the believability of a piece of communicated information. The content of information may itself be more or less believable, independently of its source (with tautologies and logical contradictions lying at the two extremes of a continuum of believability). Thus, Sperber et al. (2010) argue for the existence of a second cluster of epistemic vigilance mechanisms, that is, mechanisms which assess the quality of the incoming information (i.e., *what* to believe). They check its factual plausibility by assessing its consistency with existing knowledge and its degree of evidence. According to Sperber et al. (2010), the beliefs against which the communicated information is tested are those that are automatically activated by the comprehension process and used in the pursuit of relevance. These are a subset of the mental encyclopaedia of the addressee, and provide the ground for an "imperfect but cost-effective epistemic assessment (Sperber et al. 2010: 374). When the result of this assessment is a contradiction, there are three possible outcomes: (i) if the source is taken as trustworthy and the background beliefs of the interpreter that conflict with the incoming information are not held with much conviction, these beliefs are corrected; (ii) if the source is not regarded as trustworthy, the new information is rejected; (iii) if the source is regarded as authoritative and the conflicting background beliefs are held confidently, some process of (typically conscious) coherence checking is triggered. The choice among (i), (ii), and (iii) partly depends upon the output of epistemic vigilance mechanisms focused on the source (the speaker).

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

3. *Comprehension and epistemic assessment*

3.1 Sperber et al.'s (2010) 'pragmatic' model

Sperber et al. (2010) have recently proposed a model of the relation between comprehension and epistemic assessment that, for reasons which will soon become apparent, we shall call the 'pragmatic' model. According to Sperber and colleagues, comprehension and epistemic assessment are *parallel* processes which are triggered by the very same act of ostensive communication. While comprehension is underpinned by a relevance-guided comprehension procedure, epistemic assessment is carried out by dedicated mechanisms which contribute to the capacity for 'epistemic vigilance'.

Relevance Theory claims that comprehension is driven by the expectations of relevance which are raised by every ostensive stimulus. Specifically, every ostensive stimulus conveys a presumption of its own optimal relevance, that is, the expectation that the stimulus will be relevant enough to the addressee (to be worth processing) and that it is the most relevant one compatible with the speaker's abilities and preferences. This presumption justifies the adoption of the following comprehension heuristic:

(1) **Relevance-guided comprehension procedure**
Follow a path of least effort in computing cognitive effects:
a. Test interpretative hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.
b. Stop when your expectations of relevance are satisfied.
(Wilson and Sperber 2004: 613)

Interestingly, Sperber et al. (2010) maintain that comprehension and epistemic assessment are interconnected aspects of a single process whose goal is to make the best of communicated information and they suggest that considerations of believability play a crucial role in the comprehension process itself:

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

This is, however, a 'hypothetical' role: comprehension initially proceeds *as if* the interpretative hypotheses under construction were to be accepted as true. Given such an assumption, the first hypothesis that satisfies the addressee's expectations of relevance is attributed to the speaker as her intended meaning. No *actual* assessment of the believability of that hypothesis needs to take place at this stage. Importantly, the presumption of optimal relevance which is communicated by every ostensive stimulus need not to be true or accepted as true: the speaker might fail to achieve relevance either because of incompetence or malevolence (Sperber and Wilson 1986/1995: 158–159). However, the very fact that the presumption is communicated is enough to guide the interpretative process. To illustrate this point, Sperber and colleagues discuss the following example. Andy and Barbara have decided to throw a party and Barbara has asked Joan to bring a bottle of champagne.

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

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

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

It is crucial to notice, thought, that while Sperber et al. (2010) see comprehension and epistemic assessment as parallel processes triggered by the same piece of communicative behaviour ('The speaker has uttered *u*'), they think of their interaction as limited in scope. The only role of the epistemic vigilance system is to assess the believability of the interpretation *resulting from* the comprehension process (in light of considerations about both communicator's reliability and content's believability). In what follows, I suggest that the interaction between comprehension and epistemic assessment has a wider scope than previously assumed and, as a result, it may be more finely articulated.

3.2 *Competence, benevolence and interpretative strategies*

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

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

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

These examples, taken together, suggest that consideration of the speaker's benevolence and competence may affect the way in which we interpret what she says. An adequate account of pragmatic processing should shed some light on how this happens to be the case. With regard to this, Sperber (1994) suggested that competent interpreters have sophisticated interpretative strategies at their disposal, which allow them to cope with deliberate deception or to avoid misunderstandings due to speaker's incompetence. This proposal has been given only a relatively marginal role within the development of Relevance Theory in the following years. In what follows, I present the interpretative strategies proposed by Sperber (1994) and suggest the existence of an interesting link between these strategies and the operations of epistemic vigilance mechanisms.

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

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

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

b. Stop when your expectations of relevance are satisfied.

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

Three versions of the relevance-guided comprehension procedure:

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

Successful interpretative paths may require the adoption of sophisticated interpretative strategies. Let us consider, for instance, the examples discussed above. With regard to the delivery-example, if John were a naively optimistic interpreter, he would attribute to Carol the first interpretative hypothesis that is relevant enough to him. The first interpretation that comes to John's mind is that the delivery is late.

Given its relevance to John, a naïve interpreter would retain it and mistakenly attribute it to the speaker. But what if John adopted the cautiously optimistic version of the relevance-guided comprehension procedure? John would not take for granted Carol's competence and he would be vigilant to the possibility that Carol may not know what he knows (and may consequently fail in her attempt to make the relevant information that she intends to convey more accessible than any other possible interpretation). John would realise that Carol could not have intended the interpretative hypothesis *The delivery is late* to occur to him (precisely because she does not know that he is waiting for a delivery). Carol could not have thought that this interpretation would be relevant enough to him as she, Carol, has no thoughts of any sort involving this delivery. Thus, the comprehension procedure would go further and test the next most accessible interpretative hypothesis. For instance, it would access and assess the interpretation that it is time to go home because of the baby sitter. Since John takes it that Carol might have thought this interpretation to be relevant enough to him (as in fact it is), the interpretation is selected and attributed to Carol. With regard to the deceptive version of this example, it is possible to show how both a naïve and a cautiously optimistic interpreter would fail in attributing to the speaker the intended interpretation. Let us assume that the first interpretation to come to John's mind is the (intended) interpretation that it is time to go home because of the baby sitter. However, John immediately realises that this is not the case, as he knows that Carol and the baby sitter have made a special arrangement for that night and that the baby sitter will leave later than usual. If John were a naively optimistic interpreter, he would discard that interpretation, as it is not relevant to him (he knows it to be false). If he were a cautiously optimistic interpreter, he would also discard it given that it is not an interpretation that Carol might have thought would be relevant to him (having made the arrangement herself, Carol knows that it is not the case that it is time to go home because of the baby sitter). Only the adoption of a more sophisticated interpretative strategy would allow John to correctly attribute this interpretation to Carol. John would realise that Carol might have thought that this interpretation would *seem* relevant to him (as she does not know that John is aware of this special agreement) and, if he had reasons to think that she might want to deceive him, he would end up attributing that interpretation to her.

3.3 *The role of epistemic vigilance in comprehension*

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

If the interpreter is vigilant towards the speaker's competence and finds reasons to doubt it, he will expect 'attempted' optimal relevance. As a consequence, he will stop at the first relevant interpretation that the speaker might have thought would be relevant to him (as described in the cautiously optimistic version of the relevance-guided comprehension procedure). If the interpreter is vigilant towards the speaker's competence as well as her benevolence and he realises that the speaker may want to deceive him, he will expect 'purported' optimal relevance. In this case he will stop at the first interpretation that the speaker might have thought would seem relevant enough to him.

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

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

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

The definition of trustworthiness provided by Sperber et al. (2010) is thus intrinsically context-dependent; and it could not be otherwise. To elaborate on this point, let us focus on speaker's competence. For every speaker, there is always some information that she does not possess and some false assumptions that she takes to be true. However, this is not what 'competence' is about. If this were the case, every speaker would have to be classified as incompetent and would not be entitled to receive our epistemic trust. Competence has a narrower and context-sensitive scope: the same communicator may be competent on one topic but not on others. The investigation of epistemic vigilance mechanisms that can assess competence (as well as benevolence) in a context-sensitive way will prove to be crucial for a general understanding of epistemic vigilance, and my proposed interaction with the comprehension system.

To clarify the dynamics of the hypothesised interaction between the relevance-guided comprehension procedure and epistemic vigilance mechanisms, let us consider again the example "It's late", where the first interpretative hypothesis to come to the hearer's mind is that the delivery is late but this does not correspond to the intended interpretation. My suggestion is that the construction of an interpretative hypothesis provides a hypothesised topic of conversation. This, in turn, serves as input to epistemic vigilance mechanisms which assess the competence of the speaker on a particular topic. In this case, the interpretative hypothesis that the delivery is late provides a hypothesised topic of conversation (i.e. the delivery) with regard to which epistemic vigilance mechanisms assess Carol's competence. These mechanisms access the piece of information that Carol does not know that John is waiting for a delivery. As a consequence, an incompatibility between the speaker's system of beliefs and the interpretative hypothesis under construction is detected. This inhibits the comprehension procedure and prompts it to access (and assess) the next most accessible interpretative hypothesis.

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

As the discussion of these examples illustrates, not only do epistemic vigilance mechanisms affect the *believability* of a piece of communicated information (as proposed by Sperber and colleagues), but they also contribute to the assessment of the interpretative hypotheses under construction. That is, not only do they establish whether an interpretation attributed to the speaker (i.e. the output of the comprehension procedure) is allowed to enter the 'belief box' of the interpreter, but they also assess whether an interpretative hypothesis under construction is to be retained and attributed to the speaker as the intended interpretation (i.e. whether it ends up being the output of the comprehension system or not). As far as this latter role is concerned, epistemic vigilance mechanisms may filter out interpretative hypotheses that, although relevant, are incompatible with the speaker's mental states (i.e. her beliefs and desires). In this case, they may prompt the comprehension process to continue and assess further interpretative hypotheses. In other circumstances, they may prevent the comprehension procedure from abandoning an interpretative hypothesis that is irrelevant (to the hearer, e.g. he knows that it is false) but compatible with the speaker's mental states (e.g. her intention to induce false belief in the hearer).

3.4 *Implications*

It is worth investigating the implications of this proposal for the relationship between comprehension and epistemic assessment. Specifically, the question of whether the epistemic assessment that is involved in comprehension exhausts the validation-process for communicated information shall be addressed. According to my proposal, the effects of epistemic vigilance on comprehension and acceptance/belief can be either *simultaneous* (when no further epistemic assessment beyond that involved in comprehension is needed) or *serial*. When the addressee arrives at the intended interpretation via the recognition that the communicator is trying to deceive him, it is plausible to assume that no further epistemic assessment will be undergone and the output of the comprehension process will be automatically prevented from entering the addressee's belief box. This seems to be the case in the deceptive version of the example discussed above. If John correctly reaches the intended interpretation that it is time to go home because of the baby sitter via the recognition of Carol's deceptive intention, he won't further assess the believability of this interpretation after attributing it to Carol. However, when the role played by epistemic vigilance is that of warranting an interpretation that is compatible with the speaker's system of beliefs, it is still an open issue whether or not the interpretation selected is worth being accepted as true (i.e. believed). For instance, in the delivery-example discussed above, John recognises that Carol could not intend to communicate that the delivery is late, as she is not aware that he is waiting for a delivery. As a consequence, he looks for an interpretation that is compatible with Carol's system of beliefs, e.g. that it is late if they want to catch the last train. Once this interpretation has been attributed to Carol as the intended interpretation, it is still an open question whether John should accept it as true or not. For instance, it might contradict some of his strongly held beliefs concerning the recent introduction of a 24-hour train service and could thus be rejected as false. In this case, rejection (or acceptance/belief) will be the result of a process of epistemic assessment that takes place on the *output* of the comprehension system.

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

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

In line with this, if we assume that more sophisticated interpretative strategies are implemented by the operations of epistemic vigilance mechanisms, we can endorse the idea that optimistic interpretative strategies might represent a 'preferred' option, but reject the claim that optimistic strategies must fail in order for the interpreter to resort to 'sophisticated understanding'. Rather, addressees may start by being vigilant and only grant trust if they have no reasons to doubt the communicator's competence and benevolence. If they do have reasons to doubt her trustworthiness, they will downgrade their initial expectations of 'actual' optimal relevance to expectations of 'attempted' or 'purported' optimal relevance. As emphasised by Sperber et al. (2010), epistemic vigilance is not the opposite of trust, but it is the opposite of *blind* trust. To clarify this point, Sperber et al. (2010: 346) develop the following enlightening analogy: when we walk down the street through a crowd of people, we typically do not have any hesitation about walking among them, despite the risk of being accidentally or intentionally hit by them. However, we do monitor the trajectory of others and if we detect the presence of a careless or aggressive individual, we raise our level of vigilance. This low-level, unconscious, vigilance allows us to enjoy our stroll while preventing us from any risk of collusion. In other words, our mutual trust is buttressed by our mutual vigilance.

4. *Conclusions*

While the interaction between the comprehension system and epistemic vigilance mechanisms has not been much explored, its centrality has already been recognised:

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

This passage suggests three different perspectives that are relevant to the investigation of epistemic vigilance in communication: an evolutionary perspective, a developmental perspective, and a 'pragmatic' perspective. The main focus of this paper was to explore the 'pragmatic' dimension of this interaction.

According to Sperber et al. (2010), comprehension and epistemic assessment are triggered by any act of ostensive communication: while comprehension follows a relevance guided procedure which selects the interpretative hypothesis that would be relevant to the hearer if he accepted it, epistemic vigilance mechanisms assess the speaker's reliability and the factual plausibility of the communicated content and establish whether the selected interpretation is indeed worth being accepted as true. Going beyond this proposal, I argued that, not only do epistemic vigilance mechanism assess the believability of the output of the comprehension system, but they also play a crucial role in determining that output. As discussed above, epistemic vigilance mechanisms towards the source modulate the expectations of relevance that drive the comprehension procedure. If they detect that the speaker is not competent, they downgrade the initial expectation of actual optimal relevance to *attempted* optimal relevance. Similarly, if epistemic vigilance mechanisms towards deception detect that the speaker is not benevolent, they set the expectations of relevance at an even lower grade, that is, *purported* optimal relevance. As a consequence, epistemic vigilance mechanisms may filter out interpretative hypotheses that are incompatible with assumptions about the speaker's epistemic state or retain interpretative hypotheses that are irrelevant to the addressee (e.g. because he knows them to be false) but compatible with the speaker's deceptive intention. This has important consequences for the relationship between comprehension and epistemic assessment. Crucially, epistemic assessment taking place *after* comprehension can be made redundant whenever the selection of the intended interpretative hypothesis is grounded on the recognition that the speaker has a deceptive intention. If it is by realising that the speaker has the intention to mislead him that the addressee recognises what the speaker intended to communicate, there is no need to further assess the believability of the misleading pieces of communicated information. However, when the role of epistemic assessment in comprehension is that of warranting an interpretation that is compatible with the speaker's epistemic state, it remains an open question whether the communicated content is worth being accepted as true. In this latter circumstance, further epistemic assessment would typically follow comprehension.

Interestingly, this proposal has potential implications for the development of pragmatic abilities. Sperber's (1994) hypothesised that the three interpretative strategies corresponding to naïve optimism, cautious optimism and sophisticated understanding might represent different developmental stages. That is, children would start as naïve optimistic interpreters and subsequently deploy increasingly sophisticated interpretative strategies, which are based on the recognition that the speaker may not be competent or benevolent. In light of my proposal on the interaction between the comprehension system and epistemic vigilance mechanisms, such strategies may emerge thanks to the unfolding of epistemic vigilance capacities in the child's cognitive development. This sheds a new light on the relation between communicative abilities and other types of metarepresentational capacity (in particular, epistemic vigilance), which is open to further empirical investigation.

Acknowledgements

Many thanks to Robyn Carston, Deirdre Wilson and Dan Sperber for invaluable discussions on the content of this paper. This work has also benefited from the insightful comments of the participants at the *Philosophy of Language and Linguistics Workshop* (Dubrovnik, September 2014).

References

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

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

Grice, H. P. 1957. "Meaning". *Philosophical Review* 66: 377–388.

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

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

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

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

Origgi, G. 2008. "Trust, authority and epistemic responsibility". *Theoria. Revista de Teoría, Historia y Fundamentos de La Ciencia* 23 (61): 35–44.

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

Robinson, E. J. and Einav, S. 2014. *Trust and Skepticism: Children's Selective Learning from Testimony*. New York: Psychology Press.

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

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. and Wilson, D. 2010. "Epistemic Vigilance". *Mind & Language* 25 (4): 359–393.

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

Sperber, D. and Wilson, D. 2015. "Beyond Speaker's Meaning". *Croatian Journal of Philosophy* 15 (44): 117–149.

Willis, J. and Todorov, A. 2006. "First impressions: making up your mind after a 100-ms exposure to a face". *Psychological Science* 17: 592–598.

Wilson, D. and Sperber, D. 2004. "Relevance Theory". In L. R. Horn & G. L. Ward (eds.). *Handbook of Pragmatics*. Oxford: Blackwell Publishing: 607–632.

Croatian Journal of Philosophy
Vol. XV, No. 44, 2015

Pejoratives and Relevance: Synchronic and Diachronic Issues

NENAD MIŠČEVIĆ
University of Maribor, Maribor, Slovenia
Central European University, Budapest, Hungary

The paper considers a possible relevantist treatment, in the spirit of Wilson and Sperber's work, of pejoratives and argues for three claims concerning them. On the level of synchronic issues it suggests that the negative content of pejoratives, at least in its minimal scope, is the normal part of their lexical meaning, and not a result of extra-semantic enrichment. It thus suggests an evaluative-content approach for the relevantist, in contrast to its neutral-content alternative. On the more general side, it suggests that the relevance theorist owes us a clear story about what kind of material is normally encoded. Concerning the issues of diachronic behavior of pejoratives, the paper suggests primarily the application of relevantist theory of irony, and secondarily some links with theory of metaphor. A relevantist theory of echoic use, and proposed for irony, can be used to understand the appropriation of pejoratives by their original target group, and the reversal of valence that goes with it. There is an interesting parallel between the echoing-cum-reversal processes Wilson and Sperber propose for irony and the repeating-and-reversing process typicall of appropriation of pejoratives. Finally, a brief application of the relevantist understanding of metaphor is proposed as a tool for understanding the genealogy of pejoratives of figurative origin. The dynamics, history and development of pejoratives has not been systematically addressed by philosophical theories of pejoratives: a collaboration with relevance theory might prove a useful strategy.

Key words: Pejoratives, relevance, irony, metaphor, semantics, pragmatics.

1. *Introduction*

Let me start on a more personal note.[1] I have known Dan Sperber from Paris, at least since the early ninetees, from visits to Jean Nicod institute. I had a very superficial knowledge of his (and Deirdre's) work, didn't even read *Relevance* from beginning to end. It has all changed with Dan's arrival in Budapest some five years ago. Since then we used to teach in the same, winter semester, I used to visit regularly his seminars and his encounters with local linguists and cognitive psychologists, and we had lots of time to discuss our work. I have learned enormously from all these encounters, and I am very happy we managed to have Dan in Rijeka and Dubrovnik. All this time I was curious about Dan's co-author, Deirdre Wilson. Finally, we had her in Dubrovnik as well, and this is a good reason for celebration.

In this paper I want to inquire about the relevance of the relevance theory for my favorite small area in philosophy of language, theory of pejoratives. To anticipate briefly, I end up showing that "relevance" is here a homological word: relevance theory has the property it talks about, and it is highly relevant to for the issues at hand. Let me first briefly introduce the topic of pejoratives.

In this paper we shall be considering bad epithets for ethnic, racial and gender groups, presumed social kinds, like "Nigger", "Boche", "faggot". We can also call them "generic epithets" in a contrast to non-generic ones like "bastard"; we shall often mention the two groups in the same breadth. Our examples will be pejorative sentences like "Jack is a Nigger." combining reference to a group (e.g. Afro-Americans and, in general, persons of African descent) with ascription of some negative properties (all this with sincere apologies for mentioning the offensive words). Their use normally expresses negative attitude, so one would expect that pejoratives carry negative content, both descriptive and prescriptive. Pejoratives have been an object of study since the pioneering work of Michael Dummett (1973), continued in the same spirit by Robert Brandom (1994). Nowadays the area is in the full development, with names like Robin Jeshion (e.g. 2013), Tim Williamson (2009), Christopher Hom (2010) and Ernie Lepore (with L. Anderson, 2013) among the prominent recent authors.

Let me note that, following Hom (2010), I shall distinguish derogation from offense. While the two phenomena are closely related, they are not reducible either one to the other. Derogation is a matter of use; by using a pejorative sentence the speaker is downgrading the target. Offense is what happens at the receiving end: the target, or a neutral member of the audience might have the psychological reaction of being offended. If someone says "John is not a Nigger", he is not intentionally offending John, but is still derogating Afro-Americans. And many people, including Whites (Causasians), might be offended by the very use

[1] Thanks go to Dan Sperber, Deirdre Wilson, Michael Devitt, Dunja Jutronić, Martina Blečić and Julija Perhat.

of the term. In this paper I shall not talk about offense. When I mention "negative content" I mean primarily content that is derogatory.

Now the preview. The paper is organized around three topics. The first is synchronic, the other two diachronic. The first, occupying section 2, concerns the semantic status of the derogatory, negative material. Is it part of the meaning of pejorative, i.e. strictly semantic? Transposed into relevance theory, the question becomes one of its linguistic status: is it encoded, part of the litterary lexical meaning? Relevantists are friends of austerity, so we should concentrate upon the minimal negative material, something like "bad", "despicable" (e.g. "Boche" would contain "German and therefore despicable", or "...therefore bad"). Is this minimal negative content encoded? If not, the encoded meaning of "Boche" is just "German", and everything else is inferred. A larger question looms in the background: What are the criteria that decide what is encoded?

The second topic concerns the diachronic issue at which I find relevance theory most useful. It is the reversal of valence of pejoratives. Let me start with a quote from Hughes and a meta-quote from Rushdie:

> Salman Rushdie's provocative novel *The Satanic Verses* (London: Penguin/Viking, 1988) contains the comment:
> "To turn insults into strengths, Whigs, Tories, Blacks all chose to wear with pride the names they were given in scorn." Rushdie then proceeds to give "our mountain-climbing, prophet-motivated solitary ... the Devil's synonym: Mahound" (93). This was a xenophobic medieval name for Mahomet. (Hughes 2006: 382)

"Whig" was once a pejorative; its value then became reversed. Contemporary examples abound. Many theorists notice the phenomenon, but there is no clear account of it as yet. I think relevance theory might help, in particular Wilson and Sperber's ideas about irony, and section III is dedicated to this topic.

The third and final topic are figurative pejoratives, in particular those originating from metaphores. Many pejoratives are such: "bitch" is a metaphor (from dogs to humans), "Hun" as well,[2] "cunt" a synecdo-

[2] "Hun", derives from a belligerent speech by Kaiser Wilhelm II in 1900; Kaiser enjoined his fellow countrymen to be like ancient Huns, merciless with the prisoners, and the British press made a scandal our of it. On http://germanhistorydocs.ghi-dc.org/...ocument_id=755 one finds the following quote from the Kaiser, listed as the source of the pejorative:

> "Should you encounter the enemy, he will be defeated! No quarter will be given! Prisoners will not be taken! Whoever falls into your hands is forfeited. Just as a thousand years ago the Huns under their King Attila made a name for themselves, one that even today makes them seem mighty in history and legend, may the name German be affirmed by you in such a way in China that no Chinese will ever again dare to look cross-eyed at a German."

The German source quoted is listed as "Johannes Prenzler, ed., Die Reden Kaiser Wilhelms II. [The Speeches of Kaiser Wilhelm II]. 4 volumes. Leipzig, n.d., 2. pp. 209-12.," and the following futher info is listed: "Unofficial version of speech reprinted in Manfred Görtemaker, Deutschland im 19. Jahrhundert. Entwicklungslinien [Germany in the 19th Century. Paths in Development]. Opladen 1996. Schriftenreihe der Bundeszentrale für politische Bildung, vol. 274, p. 357.

che (from part to the whole person, and then presumably from women to men), "Lime" and "Kraut" as well; in my native Croatian, "Ustaša" derives from the name of Croatian Nazis, and when used for all Croats it is a clear synecdoche; similarly "Četnik" for Serbs. The figurative origin is quite telling, for instance, about the unity of the transferred material: with "Hun" it is the negative characteristics (savagery, cruelty, aggressiveness), negative evaluation, expression of negative feelings, and the prescripton (they should be fought!). If the transferred material is so unitary, then the whole must be unitary as well; it reminds one of thick concepts in ethics, combining factual, evaluative, prescriptive and probably expressive dimension. However, before looking at consequence, we should understand the transfer; so, we need a theory of figurative broadening. In section 3.2, I try briefly to deploy the Wilson and Sperber's theory of metaphorical inference, and apply it, also very briefly to the issue at hand.

The conclusion is optimistic: relevance theory can help our understanding of the diachronic behavior of pejoratives. Its proposals can be worked out within the theory, but they can also be exported into other approaches, and do some work outside of their native territory.

2. *The negative content: encoded or inferred?*

A crucial contrast concerning the semantic content within relevance theory is the contrast between being encoded, part of a lexical meaning and being inferred. Some inferences develop the logical form encoded by an utterance into a fully propositional form (explicatures), whereas others enrich it by non-explicit material (implicitures) (2012: 12—references with years only refer to the two books by Sperber and Wilson (1995 and 2012)). As we mentioned, the first question to be asked when it comes to a possible relevantist treatement of pejoratives concerns the status of the negative content. If a term is officially considered to be derogatory, it is plausible that it carries some fixed negative content with it. Not all agree; LePore and his collaborators are recent counterexamples (e.g. Anderson and LePore 2013). How would a relevantist treat the negative content? Why think it is part of encoded meaning?

Let me start with a pertinent quote, of two passages of *Relevance* in which Wilson and Sperber offer a sketch of a framework, and an interesting example. Here is the framework. It begins with a linguistic description determined by the grammar, which is considered to be quite universally valid. It then passes to semantics:

> Second, this linguistic description yields a range of semantic representations, one for every sense of the sentence uttered. Each semantic representation is a schema, which must be completed and integrated into an assumption about the speaker's informative intention, and can be as complex as the speaker cares to make it. Moreover, each schematic sense is generally quite different from all the others, and can be completed in quite different ways. (1995: 175)

It looks as if these semantic representations, associated with the lexic, are relatively stable. We now need an example of a word with bad content, and our authors offer one:

> Consider utterance (1), for example:
>
> (1) He's a bastard.
>
> Let us assume that on the basis of a linguistic analysis of (1) and an assignment of contextually accessible referents, the speaker might be taken to be asserting any of (2a–d):
>
> (2) (a)Peter is a nasty man.
> (b)Bob is a nasty man.
> (c)Peter is illegitimate.
> (d)Bob is illegitimate. (1995:175)

It is (2)(a) and (2)(b) that are relevant here. They seem to pick up the negative meaning of "bastard", namely "a nasty man". To my ears the negative content of "bastard" is not more stable or better known than the negative content of "Nigger" and "bitch". If the first is there, part of the lexical meaning, the other two are there as well.

Prima facie, the bad assumptions are part of the meaning (in a wide sense of the term) of the pejorative. Are there obstacles to identifying the core bad assumptions about the targets of a given use of a pejorative P with the meaning of P? A popular but not super reliable test is the following question: suppose a person, A, does not know about these assumptions. Does she know what "P" means? Does a visitor who knows that "nigger" refers to Afro-Americans, but has no idea about the stereotype of Afro-Americans associated with "nigger" know the meaning of the term? Is it appropriate to tell her that you will teach her the shocking *meaning* of the term? My hunch is that it is.

Further, how does the non-native speaker find out the meaning of "bastard"? It is what you find in the ordinary dictionary. For pejoratives, most often you find some of the bad material in the dictionary, as a part of the lexical meaning.

Here is the entry for "bitch" in John Ayto: *The Oxford Dictionary of Slang*. It begins by listing the most general meaning, common to several bad words:

> "An unpleasant or despicable female person."

And goes on to note:

> "The majority of opprobrious epithets applied to women contain, or can contain, some suggestion of immorality, particularly sexual promiscuity. ...
> **bitch** (1400) In early use often applied specifically to a prostitute, and latterly often applied specifically to a malicious or spiteful woman; from earlier sense, female dog ..." (2000: 228) (the year 1400 refers to the first appearance of the term, NM)

Similarly, *Random House Dictionary* (1994) specifies the meaning of "cunt" when applied to a male: "a despicable, contemptible or foolish man." None of the example in itself clinches the point, but taken together they re-enforce the prima facie plausible assumption that claims about "uncivilized manners", "slave-like status or behavior", and the like just enter the meaning of the slurs in question. In other words, if stereotypes associated with pejoratives are listed as their meanings, and the ascription of properties etc. indicates that they belong to the meaning of the slur, then pejoratives are not purely performative and expressive, but semantic.[3]

The same goes for correspondences between languages. The *Collins English-German Dictionary* entry for "bitch": *(=woman,* as they put it*),* offers two direct equivalents*,* "Miststück" and "Hexe", noting that they are "spiteful". The expression "don't be a bitch" is rendered as "sei nicht so gemein or gehässig", and "she's a mean bitch" as "sie ist ein gemeines Stück."

Of course, this is all merely *prima facie* evidence. But where should we look further? Let me borrow from the traditional view of lexical-semantic meaning and offer you five more criteria, compiled from literature and listed by Zoltan Gendler Szabo in his introduction to *Semantics vs. Pragmatics* (2005: 6).

> a) Competence. Typically, some but not all of what the speaker conveys could be grasped by any competent speaker without special knowledge.

Seems that if holds for our pejoratives, say "bitch" and "nigger": any competent English speaker can grasp the negative material the speaker conveys.

> b) Encoding. This is just the presence in the dictionaries, we started with.
>
> c) Compositionality. Typically, some but not all of what the speaker conveys is compositionally determined (by the syntax and the lexicon).

We have seen that our relevantists say the same about the two meanings of "bastard". I see no reason not to apply it to pejoratives.

> d) Rules. Typically, some but not all of what the speaker conveys can be ascertained by following rules, as opposed to elaborate cognitive strategies.

[3] Standard dictionaries indeed talk about senses of slurs, exactly in terms of stereotypes and figurative origins that we sketched. Geoffrey Hughes notes for instance the following:

> In his *Classical Dictionary of the Vulgar Tongue* (1785), Francis Grose noted that the basic term negro carried the sense of "slave" in uses like "I'm no man's Negro." (2006: 327).

And he comments:

> The history of the term in the southern United States is obviously colored by the slave relationship.

Similarly, with the South African equivalent, the term "kaffir" for black people:

> From the earliest accounts, the stereotype of the savage predominated, especially in the categorization of the "red" or "raw" kaffir, so called because of the red ochre that they smeared on their bodies.
>
> In their Zulu Dictionary (1948), C.M. Doke and B.W. Vilikazi included the following usage note: "Term of contempt for a person (black or white) of uncivilized manners (a swearword if used direct to a person)." (2006: 281).

What elaborate cognitive strategy do you need to understand what the speaker conveyed by calling his female boss a bitch? None.

> e) Truth-conditionality. Typically, some but not all of what the speaker conveys is truth-conditionally relevant.

This is of course, debatable, and denied by relevance theorecians, but we don't need it here.

> f) Intention-independence. Typically, some but not all of what the speaker conveys is independent of the speaker's specific intentions to talk about this or that.

Holds indeed! Most theoreticians agree that the derogatory content of pejoratives is independent from speaker's intentions.

So, nearly all criteria offered classify pejoratives in the same way; this strengthens the case for semantic nature of the bad content.

In the discussion in Rijeka professor Sperber has distanced himself from the idea that negative meaning is always semantically lexically encoded. Tradeoff between lexical meaning and inference can be achieved in various ways. "We dont have to take a stand on that (i.e. lexical meaning) in order to develop our understanding of the use of pejoratves. We don't need the same lexical meaning in the head of each speaker/hearer", he said, pointing to varieties of idiolects and small-range sociolects. Specific groups use the term each in its own way. But, it is not the case that anything goes, he added.

On the methodological side he has proposed the following: in order to be sure, look at language learning, the acquisition; only then will you have a principled criterion. Finally, he allowed that some of the content might be "linguistic but not involving strictly speaking semantics".

I thank him. The methodological remark seems to me constructive, and I hope someone will do the work. In the meantime, let me note that the negative character is often the first thing a foreigner learns about the meaning of a pejorative. To mention German examples, I have learned that two words, "Kanake" and "Tschusch", have negative meanings, before realizing that the first is applied to Turks and the second to south Slaves; and way before realizing the exact extension of the first, namely Turks, Arabs and Mediterraneans in general. I am still not sure about Tscusch, for instance whether a Kosovo Albanan counts as a Tschusch. But I am absolutely certain of the negative character of the term. How children learn these words is an open question, but the differences should not be dramatic.

The pluralism remark is relevant, but there are limits to its scope. If there are widely based sociolects (especially ones involving a vast majority of the language speakers) that use the term basically in a uniform way, then the term is like "bastard", endowed with a fixed negative meaning. If we can never be sure, than anything goes, the option Sperber wants to avoid.

The contrast points to a wider dilemma: the relevantist account presupposes some fixed lexical meaning, and sufficient commonalities between speakers to insure the inferential work. If for some word, these conditions are satisfied, then we have fixed lexical meaning. Whether we call it "semantic" or not, makes little difference. If for all sorts of terms, then for pejoratives as well. If these conditions are practically never satisfied, than anything goes. An important morals of all this is that *the relevance theorist owes us a clear story about what kind of material is normally encoded*; relevance theory starts from such material, and for all its richness, its plausibility may ultimately to a large extent depend upon the kind of basic units it is willing to deploy. If we are offered no clear criteria, then ultimately anything goes.

For the negative content of pejoratives, I would opt for the first horn of the dilemma: it satisfies most of the usual criteria for lexcal meaning: it is well known and stable accross various social groups, except the targeted group; but the members of that group understand very well its derogatory meaning. So, it should be part of the lexical semantics of the pejorative. How much material goes into negative content is a further issue. I would place in it quite a lot of material (see Miščević 2012, 2014), but for the relevantist purposes a minimum should be OK, like the meaning "a nasty man" for "bastard". So much for the synchronic issue, with apologies for brevity.

3. *The life of pejoratives: A diachronic perspective*

Here is a different aspect of the account that calls for the dialogue with relevance theory, namely the dynamics, the history, or development of pejoratives. Here I would like to thank again Dan for all I have learned from him.

The issue of the diachronic behavior of pejoratives has practically not been addressed systematically within the philosophy of language. Linguists, in particular historians of language, have done extensive work and there are some impressive and readable books on the topic, strictly on generic pejoratives Kennedy (2002), or on bad language more generally, for instance Keessen (2009) and Saunders (2011, see the critical discussion by Julija Perhat (2012). I would like to use the opportunity to start addressing it in a more systematic way, with the help of some ideas from relevance theory. I shall briefly discuss two areas. The first I shall approach using strictly the material from relevance theory, namely the issue of appropriation of pejoratives by the members of the targeted group. My second area is the metaphorical or more widely, figurative origin of many pejoratives. Here I shall appeal to relevance theory as one source of ideas and methods.

3.1 *Echoic use, appropriation and the reversal of valence*

In this section I want to borrow an idea, or rather a small but efficient tool-box of ideas, from Wilson and Sperber, in order to explain the famous phenomenon of the change of valence of pejoratives. I believe that the relevantist story about echoic use and about irony can be exported to this new context and do the main explanatory work. Let me start by presenting some background.

Pejoratives often change their valence, from negative to positive. The members of the target group chose "to wear with pride the names they were given in scorn", to quote Rushdie again. The famous cases are, in American culture, the word "Nigger", and in German culture "Kanake", these days normally pejorative for Turkish or Kurdish or Arabic immigrants. Afro-Americans have appropriated the first one, the younger generation of the descendent of Turkish immigrants the second: "we Niggers" and "wir Kanaken" has become a respectable way of internal characterizing of the relevant community.[4] With appropriation goes reversal of valence: "Niggers are stupid and brutal" vs. "we Niggers are cool". Similarly, for other groups or "social kinds". Take the word "bugger" and its French ancestor, word "bougre"; nowadays in French "Mon bougre" is very friendly. "Bitch" as used by many women is positive: "We, bitches are super". "Bitch" as used by a male chauvinist is negative. The same with "witch", "dyke" , and other formerly pejorative epithets turned up in the names of small feminist groups, as Hughes points out: "…radical feminist groups similarly chose provocative acronyms such as (…) witch (women's international-terrorist conspiracy from hell), which, according to its manifesto, was born on halloween 1968." (Hughes 2006: 352). Finally, a group of friends, women dissident activists in Zagreb, were called witches by their right wing colleagues. Then, they started using the word to refer to themselves (more on this in a moment). As a rule, only members of the targeted group can appropriate the pejorative, reverse its valence and use it possitively.

How does all this happen? Little is said in the literature at the general level. However, writing about irony, Wilson and Sperber have pointed out two processes: the first is the ironical echoing by the speaker of someone else's utterance, and the second is the re-evaluation of the utterance echoed. I want to argue that there is an interesting parallel with the quoting-reappropriation-and-reversal processes that seems to take place in the valence change of pejorative. Let me remind the reader of the way in which Wilson and Sperber bring in echoing in their account of irony. The example is Marry's comment on the party she found very boring: "The party was a fun." Wilson and Sperber proposed that what whe is doing is "expressing an attitude of scorn towards (say)

[4] For "Kanake", the reversal is linked with the publication of the novel by Feridun Zaimoğlu: *Kanak Sprak: 24 Mißtöne vom Rande der Gesellschaft.* 1995, 6. Auflage. Rotbuch, Hamburg 2004.

the general expectation among the guests that the party would be fun" (2012: 125). She is echoing the sentence: "The party was a fun", dissociating herself from it, and re-using it to poke fun on the naive guests who expected somethnig better. In general:

> ...irony consists in echoing a thought (e.g. a belief, an intention, a norm-based expectation) attributed to an individual, a group, or to people in general, and expressing a mocking, sceptical or critical attitude to this thought. On this approach, an ironical utterance typically implies that the speaker believes the opposite of what was said, but this is neither the meaning nor the point of the utterance (2012: 125).[5]
>
> The central claim of the echoic account is that what distinguishes verbal irony from other varieties of echoic use is that the attitudes conveyed are drawn from the dissociative range: the speaker rejects a tacitly attributed thought as ludicrously false (or blatantly inadequate in other ways). Dissociative attitudes themselves vary quite widely, falling anywhere on a spectrum from amused tolerance through various shades of resignation or disappointment to contempt, disgust, outrage or scorn (2012: 130).

Doesn't this remind one (if one is working on pejoratives) of quoting the pejorative and adding a strong dissociative attitude? With the stress on contempt, disgust, outrage and scorn. Let me re-use the "witch" example from Zagreb which i witnessed personally and where I know the participants ell. It was a conflict between nationalist intellectuals with their female anti-nationalist colleagues (some of which are also feminist activists). Let me call the main nationalist professor Victor, and the feminist opponent Rada. Imagine the situation, Zagreb 1989, the tensions between Croatia and Serbia are at their peak and the war is just about to start; right time for a witchunt. Imagine Victor saying (and writing) about his anti-nationalist female activist colleagues: "These women are real witches."

Rada pretends to agree: "Now, we are real witches." Rada is echoing her enemy, and like the Afro-American speaker in the Nigger case, she does not need not to pretend she is a nationalist (viz. white racist); the echoing is already clear enough. So, what is Rada echoing? First, both

[5] And here are the higher genera, the echoic and the attributive:

> We define echoic use as a subtype of attributive use in which the speaker's primary intention is not to provide information about the content of an attributed thought, but to convey her own attitude or reaction to that thought. Thus, to claim that verbal irony is a subtype of echoic use is to claim, on the one hand, that it is necessarily attributive, and, on the other, that it necessarily involves the expression of a certain type of attitude to the attributed thought (2012: 129).

Echoing is metarepresentational:

> An echoic utterance indicates to the hearer that the speaker is paying attention to a representation (rather than to a state of affairs); it indicates that one of the speaker's reasons for paying attention to this representation is that it has been entertained (and perhaps expressed) by someone; it also indicates the speaker's attitude to the representation echoed. An echoic utterance achieves relevance by allowing the hearer to recognise, and perhaps to emulate, the speaker's interest in, and attitude to, someone else's thought. The speaker may express any one of an indefinite variety of attitudes to the representation echoed (2012: 93).

the content and the attitude of the opponent—Wilson and Sperber are obviously right. But, of course, her uttering goes further: she dissociates herself strongly from Victor, the original speaker and to his buddies, patriarchal nationalists. So, the content is: we the women in the group are witches, but the attitude is anger outrage and scorn. Here, the echoic theory works perfectly. Let us generalize.

The meta-representational echoing of pejorative sentence can express the spectrum from "resignation or disappointment to contempt, disgust, outrage or scorn". Consider Wilson and Sperber again:

> ... two features of attributive utterances in general which are also found in echoic utterances. First, attributive utterances (including tacit indirect reports) can be used to inform the hearer about the content not only of thoughts or utterances attributed to a particular individual on a particular occasion, but of those attributed to certain types of people, or to people in general. These may have their roots in culturally-defined social, moral or aesthetic norms, or general human hopes or aspirations (2012: 130).

They stress the normative bias of irony:

> There is a widely noted normative bias in the uses of irony. The most common use of irony is to point out that situations, events or performances do not live up to some norm-based expectation. Its main use is to criticise or to complain. Only in special circumstances is irony used to praise, or to point out that some proposition lacking in normative content is false. This bias is unexplained on the classical or Gricean accounts. To illustrate: when someone is being clumsy, it is always possible to say ironically, "How graceful", but when someone is being graceful, it takes special circumstances to be able to say ironically "How clumsy" (2012: 127).

The ironical reversal goes from positive to negative. The reversal of pejoratives will take the opposie direction. Let us go back to the "bitch" example. Start with the source group, SG, of male users of the term. Imagine then a dissenting woman, call her Jane. At the start Jane's cognitive system represents the (inimical) concept-representation $\mathrm{BITCH}_{\mathrm{SG}}$ originated by source group SG, and probably creats a meta-representation $\mathrm{mBITCH}_{\mathrm{SG}}$. She then works on the concept $\mathrm{BITCH}_{\mathrm{SG}}$ and refigures it, re-evaluating some qualities, deleting them, replacing negative qualities like dangerous, aggressive, promiscuous with positive or neutral ones dangerous-to-one's enemies, sexy**.** The negative -positive- contrast yields a very stable symmetry, it works even better here than for irony; in the case of irony, many examples do not presuppose the deployment of the relevant item by interlocutor, see "What a beautiful wheather!" example. In the case of pejoratives, the target group is surrounded by the deployers of the pejorative, so, the assumption of echoing is much more natural

Consider now the stages:

stage zero:

The word "bitch" is used by a big male group M, for the negative concept BITCH of the target group, women, call it $\mathrm{BITCH}_{\mathrm{SG}}$.

At the beginning the pejorative is sometimes merely quoted: "we, 'bitches' as called by males" (or "we 'witches' as called by the right-wing colleagues"). Here, the idea of echoing is very helpful.

stage one, echoing: mention, not use

Jane, a member of the target group quotes "bitch" in a positive context, "We 'bitches' are dangerous to our enemies", meaning, "We, called derogatively 'bitches' by male group M are are dangerous to our enemies". Similarly, some Afro-American speaker in the hypothetical initial historical situation is echoing the deeply rooted rasist use of the word with strong negative valence: contempt, disgust, outrage or scorn. Here, the utterance of the typical rasist is being echoed and no need for pretense. The same with the "witch" example from Zagreb.

So, there is an interesting parallel between the echoing-cum-reversal processes Wilson and Sperber propose for irony and the repeating-and-reversing process typicall of appropriation of pejoratives. (In the discussion in Dubrovnik Sperber expressed agreement with my understanding of echoing, and interest in the application to pejoratives. I thank him.)

So much for the direct borrowing from Wilson and Sperber, but their idea of echoing with dissociation is also useful in the next step—step two: after echoing, reverse the evaluation. Echoing and pejorative reversal go naturally together. "Nigger" is negative when white racist's use it, but we Blacks think highly of ourselves, so, "Nigger" will be good when we use it. The idea is to reverse the evaluation of some quality from the conception-stereotype (or encyclopedic entry in one's head):—"bitch" involves a sexual component; well it is sexy, "ustaša" involves being similar to Croatian Nazis; well, they were not that bad. The idea is acceptable for relevantsts, but also for both semanticist and expressivist theory.

The reversal is worthy of a closer look. Obviously, some initial qualities (I shall use Q for quality) are replaced by some others, or some new qualities are simply added (I shall use the arrow "→" to indicate the change). Also, the positive (+) and negative (–) evaluations are in play. Let me propose a few possibilities. First, the basic ones:

1. Simple reversal of presentation and valuing of given quality Q. The presumably negative quality Q-, is now evaluated positively, and becomes Q+:

(1) Q–→ Q+

We can distinguish two cases. The first is purely indexical. Take the Zagreb nationalists disparaging the anti-nationalist feminist group: "The witches hate *us*, nationalists, and this is one feature that makes them bad." Now, the group starts appropriating the pejoratives: yes, we hate *them*, the nationalists, and this makes us good ("We, 'witches' hate you, and we are great!"). The reference in contexts of the two underlined pronouns is the same, the modes of presentation are opposed (us/them), and the valuation follows the modes of presentation.

The second is the non-indexical reversal of evaluation–still echoing content.

Take "Nigger" and consider the reversed property-valuation: it is good to be uncivilized. Being civilized is bad. The same for "bitch": it is good to be sex-hungry (and analogously, it is good for a male to be gay, for a women to be aggressive against males). A trace of egocentric attitude might be still there in specifying the rationale: the quality is good because it is us who have it.

2. Denial of having of some bad quality from the conception-stereotype:

(2) Q– → not–Q–

We "bitches" are not sex-hungry, we "Niggers" are not uncivilized, we "buggers" love our partners just as heteros do. Of course, change (2) is incompatible with (1) for the same quality. And indeed, the members of the target group sometimes divide over which one to follow, the typical 'moderates' following (2) and typical 'radicals' following (1). Outside the appropriation context, the move (2) is, of course, part and parcel of normal, enlightened education against racial, gender and other prejudices: the members of the target group just do not have the traits ascribed to them by the stereotype.

3. replacement of the relevant quality from the stereotype with its positive relative, call it Q*+.

(3) Q–→Q*+

"Bitch" involves a sexual component; well it is sexy (if you need evidence, just look up under "bitch" at Google images: you find a series of attractive young women, with the word "bitch" written on their shirts). The same for Zagreb "witches": "Witches are powerful and dangerous, we, witches, have our powers and we are dangerous to people like Victor, but we do this in the name of justice and of tolerance between peoples of Yugoslavia".

In practice, (3) is the middle road between the extremes (1) and (2). You don't want to be depicted as frustrated, sex-hungry person, but also not as an a-sexual and cold one. Replacing "sex-hungry" with "sexy" solves the problem: the link with sexuality is preserved, and in addition, now it is the others who are 'hungry', who yearn for you, and not vice versa.

4. Adding new positive qualities, possibly from the stereotypical self-image of the targeted group:

(4) Ø→Q'+

For "Nigger" one can add sexy, talented for music and dance, for "bitch" cool, and so on. Here the new elements go way beyond echoing, and the original analogy with ironical echoing-cum-reversal is at its weakest.[6]

[6] Dictionaries sometimes mention some of the resulting meanings. Here is a selection from entry for "bitch" at dictionary.com:

Next comes possible combination of moves, where several options are possible; I shall mention only two.

First, (2) the denial of having some bad quality Q- and (3) introduction of some new attractive relatives Q*+ of the initial negative one, i.e. the (2) & (3) combination.

(2) & (3) Q– → not-Q–&Q*+

The move is very natural, almost trivial, and also has the advantage of offering the target-group friendly folk explanation of why the negative quality is there in the inimical stereotype. Take the trait of being sexually violent, ascribed to members of some minority group. Well, the minority member reasons, if we are irresistibly sexy, and attractive to majority group females, this explains why majority males are afraid of us and represent us as violent (perhaps the argument is not very persuasive to academic audience, but is surely good enough for street bragging).

Second, the (2) & (3) & (4) combination: negate some, replace other with positive relative and add some new, unrelated ones.

Q– → not-Q-&Q*+ & Q'+...

This is a promissing maximal combination: some negative qualities are just too bad to be retained (e.g. being stupid), some can be re-valued (e.g. being dangerous, if the dangerousness is directed to persons that deserve punishment), some have attractive relatives or apparent relatives and some are just convenient positive qualities from the self-stereotype of the group.

The more semantically and less pragmatically oriented theorist can add a few more stages.

STAGE 3—new concept

Jane has created a positive concept, linked with concepts DANGEROUS TO OUR ENEMIES, SISTERS TO EACH OTHERS, AND SEXY

STAGE 4—from mention to use

Jane completely disquotes “BITCH”, and now uses it for the positive, BITCH $^{+}$ concept.

STAGE 5 –conventionalization. Other members of the target group converge on the usage, on which “bitch” has BITCH^{+} concept as its sense.[7]

- –noun 1. a female dog.
- 2. a female of canines generally.
- 3. Slang . a. a malicious, unpleasant, selfish person, especially a woman.
- b. a lewd woman.
- Slang . a. a complaint.
- b. anything difficult or unpleasant: The test was a bitch.
- c. anything memorable, especially something exceptionally good: That last big party he threw was a real bitch.

[7] To summarize, the more semanticsts account can postulate the following stages:

STAGE 0: The word P is used by members of a source group (SG) for the negative

To conlude, the relevantist ideas are extremely relevant and useful. But some of them also go well with the conceptual, more semanticist account. But, there are more diachronic matters to discuss, and we turn to the next one.

3.2 *Pejoratives from metaphors*

A lot of pejoratives are of figurative origin. The most hostile English-language term for Germans, "Hun", is a metaphor, ascribing to Germans the savageness, aggressiveness and cruelty of Huns. ("Nigger" and "Boche" are not figurative, and this has obscured things in the discussion.) As we pointed out, "bitch" is a metaphor (from dogs to humans), "cunt" a synecdoche (from part to the whole person, and then presumably from women to men). To give examples from other languages, let me mention the main fashionable political chauvinistic pejoratives in my language which are synecdoches as well: the Serbian use of "Ustaša" for all Croats generalizes from Croatian Nazis to the whole nation) and the Croatian use of "Četnik" for all Serbs.

Now, a figurative use involves a lot of work. The Germans-hater takes the vehicle "Hun" and projects some descriptive features (real or imagined) of ancient Huns upon contemporary Germans. Then, he transmits the negative evaluation: Germans are as bad as Huns were. The sexist takes the vehicle "bitch-dog" and projects some descriptive features (real or imagined) upon the person. Then, he transmits the negative evaluation: same supervenience basis in dogs and persons, same value properties in both. The Croatian chauvinist takes the negative features of Serbian "četniks" from the Second World War or from paramilitary criminals from the nineties, and projects them upon all Serbs. With negative features goes negative evaluation. The mere expressive view has no means to account for the conceptual complexity that goes with the genesis of figurative pejoratives. But there is more. We read the following:

> Bitch has the longest history in the field as a term of abuse, extending from the fourteenth century up to the present. During this long period it has been applied variously to a promiscuous, sensual, mean, or difficult woman, as well as to a man or thing (Hughes 2006: 23).

concept T of the target group, call it $T_{SG.}$.

STAGE 1: A member G of the target group quotes "P" in a positive context, e.g. "We 'Ps' are brave", meaning, "We, called derogatively 'Ps' by SG are brave".

STAGE 2: G's cognitive system represents the (inimical) concept-representation T_{SG}, probably creating a meta-representation mT_{SG}; he then works on the concept T_{SG} (the negative representation), and refigures it
replacing negative qualities Q^-_1 Q^-_2 Q^-_3 Q^-_4
with positive or neutral ones Q^+_1 Q^+_2 Q^+_3 Q^+_4

STAGE 3: Jane has created a positive T^+ concept, with Q^+_1 Q^+_2 Q^+_3 Q^+_4 and some other items as components.

STAGE 4: Jane disquotes "P", and now uses it for the positive concept.

STAGE 5: Other members of the target group converge on the usage, on which "P" has the positive concept as its sense.

Similarly, the word "cur" meaning just dog, has several uses.

> Dr. Johnson comprehensively defined the adjective curish (also first used by Shakespeare) as "having the qualities of a degenerate dog; brutal; sour; quarrelsome; malignant; churlish; uncivil; untractable. (Hughes 2006:137)

Hughes writes:

> Shakespeare uses "cur" in *Midsummer Night's Dream* (1590). In his works it is a term of withering scorn, most memorably in *Coriolanus* (1608), when the hero castigates the Roman plebs as "You common cry of curs!" meaning "You pack of mongrels!" (III iii 118) (Ibid.).

In all these examples, the speaker is encoding relevant properties in his use of the pejorative. Hans is aggressive; "of course, he is a Hun". But Helga is cruel to her subordinates; "well, she is Hun". The listener has to figure out the properties ascribed, answering the why-question. Coriolanus's enemies are curs, since they are of low birth, but Cesar's enemies might be curs since they are so aggressive. Person A is a bitch since she is overly sensual (horny bitch), person B since she is aggressive (and she is, let us suppose non-sensual). Why do you call B a bitch, she never showed any interest in man? Oh, I didn't mean this, I mean that she is dangerously aggressive.

Wilson and Sperber suggest that "the encoded concept helps to activate contextual implications that make the utterance relevant as expected" (2012: 110). On the side of non-specific pejoratives, take the cur example. Ironically, Frege's translators use "cur" for his example of the mere "tone", as opposed to content. But, even if with "cur" as used for dogs the negative attitude is (or were) just a matter of tone, it is not when used for people. Coriolanus transfers the qualities of a degenerate dog to Roman plebeians: they are mongrel, brutal, sour, quarrelsome, malignant, churlish, uncivil and intractable.

vehicle concept:		**target concept:**
MONGREL DOG	$\longmapsto$	CUR-LIKE HUMAN (or ROMAN)
components:		**components:**
MONGREL, BRUTAL, SOUR, QUARRELSOME, MALIGNANT, CHURLISH, UNCIVIL AND INTRACTABLE		NON-PURE ROMAN, BRUTAL, SOUR, QUARRELSOME, MALIGNANT, CHURLISH, UNCIVIL AND INTRACTABLE
qualities pointed to:		**qualities pointed to:**
mongrel, brutal, sour, etc.		non-pure Roman, brutal, sour, etc.

What can relevance theory of metaphor teach us about the process of understanding?

> …interpretation is carried out "on line", and starts while the utterance is still in progress. We assume, then, that interpretive hypotheses about explicit content and implicatures are developed partly in parallel rather than in sequence, and stabilisewhen they are mutually adjusted so as to jointly confirm the hearer's expectations of relevance. And we are not suggesting that the hearer consciously goes through just the steps shown in the tables, with exactly those premises and conclusions (2012: 113).

To apply it to the example at hand, the encoded concept CUR helps to activate contextual implications that make the utterance relevant as expected, and the concept conveyed by the "cur" metaphor is one of degenerateness, brutality; quarrelsome character and impurity of descent, characterised by these implications. The "cur" metaphor is based on fairly central properties of the lexicalised category, i.e. degenerateness, brutality; quarrelsome character and impurity of descent. In discussion with Sperber I have arrived at the following picture of the inference, performed by the Roman plebeian, call him Pauperus.

Coriolanus has said to Pauperus: "You are a cur."

Now, Coriolanus's utterance is optimally relevant to Pauperus, since the situation is one of the protest, the two sides facing each other. According to Sperber's account, Coriolanus's utterance will achieve relevance by addressing Pauperus's behavior at the protest.

Curs (in the lexicalised sense) are degenerate, brutal, quarrelsome and of impure descent.

How does the decoding of Coriolanus's utterance by Pauperus proceed? First, we look at expectations raised by the Pauperus's recognition of the utterance as a communicative act. Next, at expectation raised by the assumption of relevance, given that Coriolanus is responding to Pauperus's demands. The assumption is activated both by use of the word "cur" and by Pauper's wish for a conflict with Coriolanus. So what is tentatively accepted by Pauperus as the point of Coriolanus's utterance is the following:

Pauperus is degenerate, brutal; quarrelsome and of impure descent.

In short, Pauperus is a cur. Or, if you prefer a longer version, Pauperus is a cur who is revolting against the great Corriolanus. This is the implicit conclusion derivable from the knowledge of lexicalized meaninng, together with an appropriate interpretation of Coriolanus's utterance, which would make her utterance relevant-as-expected. Tentatively accepted as an implicit conclusion of the utterance. So we have interpretation of the explicit content of Coriolanus's utterance as decoded, which, together with lexical meaning, would imply that Pauperus is degenerate, brutal, quarrelsome and of impure descent. This interpretation is then accepted as Coriolanus's explicit meaning.

In their work on metaphor Wilson and Sperber offer a fine example of how people interpret a metaphor they meet for the first time. In the example, Peter and Marry discuss whom to invite to Billy's birthday party. Mary says:

(7) "Archie is a magician. Let us invite him."

What if Peter has only one encoded meaning for magician, person with supernatural powers.

> For that matter, some people may have only a single encoded sense for "magician": someone with supernatural powers who performs magic. They would still have no difficulty arriving at an appropriate interpretation of (7) by extending the category of real magicians to include make-believe ones.

> For other people, the metaphorcal sense may have become lexicalised, so that "magician" has the additional encoded sense someone who achieves extraordinary things.

Suppose, such a person hears the utterance of

(29) "My chiropractor is a magician."

How is she going to react?

> They would obviously have no trouble arriving at an appropriate interpretation of (29). Mary did not intend her utterance to be understood literally in (7) and metaphorically in (29); her communicative intentions—like those of all speakers—are about content and propositional attitude... (2012: 114).

And this is the kind of thing Pauperus is doing. The interesting point is that this account is compatible with a range of theories of how figurative pejoratives get lexicalized; my favorite is the account by Deirdre Gentner (see references), but this is a story for another occasion. I have addressed it briefly in Miščević (2014), and hope to do more.

Let me conclude with a piece of morals concerning semantics. Imagine what things would be like if the affect in the use of slurs were in the typical case like phobia, with no awareness of reasons. Consider the ethnic figurative pejorative and our example of conversation: "Hans is aggressive"; "of course, he is a Hun". "But Helga is cruel to her subordinates"; "Well, she is Hun". There is no way to make sense of the conversation(s) if the use of pejorative is blind. The Germans-hater would have no reaction, since he does not "see" that Hans might be aggressive because he is a German, and that the use of pejorative is selling this assumption. The person who uses "bitch" for a woman might accept that she is very sweet and pleasant; after all, *the pejorative has no added cognitive content* on non-semanticist's view; "bitch" just refers to women. (I don't deny that in the situations of rage the use of a slur might be utterly non-cognitive, I only deny that this is the typical, let alone theoretically central case). The figurative slurs thus drive home the lesson already adumbrated by stereotypes. *The figurative basis demands cognitive work both from the speaker who is encoding relevant properties, and from the listener figuring out the properties ascribed, and thus answering the why-question by a because-suggestion.* Since stereotype-linked and/or figurative slurs are so ubiquitous, and since they either assume or demand cognitive achievements (knowledge or encoding, or decoding) we may suggest, contrary to the non-semanticist, that the central and standard uses of slurs are cognitive.

In fact, one can argue from figurative slurs and from the need for cognitive effort (or simple of knowledge of relevant presumed properties of the target), that the central and standard uses of slurs are cognitive. Since cognition has to do with truth and falsity, and since the cognitive task is a good indicator of semantic structure, it seems that the ascription of negative properties, etc. indicates that they belong to the meaning of the slur, and that this meaning might confer truth-aptness. I presume that the (nasty) richness of meaning might vary with pejoratives: all of them involve "contemptible because G" at the very least. The most typical once carry more information. Some of it is given in the form of conceptual links roughly delineating the core stereotype associated with the pejorative, some in the form of figurative transfer of properties from some vehicle to the target member of G. So, slurs are not purely performative and expressive, but semantic in the traditional, truth-directed sense. But this is again, a story for some other occasion.

4. *Conclusion*

Pejoratives are a hot topic in the philosophy of language, and I have tried to point to some interesting connections between relevance theory and the issues connected with them. Although my general sympathies are with a much richer semantic approach, I have limited myself here to the relevantist perspective, and attempted to show how the ideas proposed by Wilson and Sperber might contribute to the understanding of pejoratives.

The first topic discussed has been a synchronic one, namely the meaning of pejoratives. Two options have been contrasted: the neutral-content relevantist vs. evaluative content relevantist one. On the first option, a typical pejorative, like "Kike", just has its referential meaning as the only lexical semantic feature; it means "Jewish", and this is all. All the negative material associated with it is added at various stages of enrichment. On the second, a minimal negative characterization is part of the lexical meaning of "Kike". I have argued that the relevantist should welcome the second option. The negative material is prominent in the dictionaries, the ignorance of it counts as ignorance of language, and it is often the first thing a foreigner learns about the meaning of a pejorative, and often the last he or she forgets.

We have then turned to diachronic matters the "biographies" of pejoratives and the contribution relevance theory could make to the study of these. The life (history, development) of pejoratives has been studied in detail with particular pejoratives in the focus, but no general theory has emerged; we hope it can emerge in a fruitful dialogue with relevance theory.

We have first addressed the topic of echoic use, appropriation and the reversal of valence. We hope that we have detected a possibility of a central contribution the ideas of relevance theory could make here to the understanding of diachronic processes characterizing the carrier of pejoratives. According to Wilson and Sperber, irony involves two crucial moves: first, echoing some (real or presumed) opinion concerning the topic under consideration (say, the expectations that the weather will be fine), and then using the echoed utterance to suggest the contrary ("What a fine weather!" uttered ironically amidst heavy rain).

Pejoratives can be treated, and are often treated, in a symmetric fashion by a member of the target group. First, the pejorative is just re-used, echoed, and then its import is reversed, like in irony, but in the opposite direction, from negative to positive. The qualities culled up from the conception (or, encyclopedic entry, to stay with terminology of Wilson and Sperber) associated with the pejorative have to be reworked: some are just deleted, others are re-valued, still others are replaced by more positive counterparts, and all the consistent combinations of changes are often carried through. I hope that the general characterizations of possible changes, offered in section 3.1, can help understand and organize the rich empirical material culled from the history of particular pejoratives. Although the model comes from relevance theory, it is applicable within various approches; a thoroughgoing semanticist can use it as well. Of course, the model needs testing on particular cases, but the job of the theoretician is to offer plausible ideas for testing.

The final topic, also from the diachronic group, is the topic of pejoratives of figurative origin, which we have narrowed down to those of metaphorical origin. Here, the relevance theory can help with proposals about ordinary understanding of metaphorical utterances. Wilson and Sperber (and also Carston 2010) stress the ordinary character of what is classically seen as metaphorical inference; for them it is just broadening of the same kind as the one that happens in ordinary loose talk, when people use "Kleenex" not for the brand, but simply for paper tissue, and the like. Section 3.2 presents stages of such reasoning, illustrated by the metaphor from Shakespear's "Coriolanus", where the hero addresses the rebellious plebeians as "curs". Again, the model is applicable within a range of approaches: the semanticists can claim, as I would do, that it successfully accounts for the crucial first phase of a biography of a figurative pejorative. The phase culminates in the creation of a new, figurative concept "cur", and the word then enters the second phase, with the dominant litteral meaning and a fledgling figurative meaning. The later, if successful in actual practice of communication, becomes more and more fixed.

In short, the relevantist proposal are extremely relevant and useful, both on their native pragmatist ground, and as sources of tools for different, say conceptual semanticist accounts. So, let me mention a couple of topics for further research, concerning the career of pejoratives.

The first seems to be the easiest one, namely change of meaning of terms, like with "Führer", "macho" and "Übermensch", that started their carreer as positive ones. This negative reversal seems to follow the same pattern of echoing and dissociation we saw with positive reversals.

The second is much more difficult, and to my knowledge, it has not been much studied. It is pejorativization, the turn from neutral to bad. A fine example is "bastard", according to the dictionary "acknowledged child of a nobleman by a woman other than his wife."[8]

[8] Here is the entry from *Online etymological dictionary*:
Illegitimate child, early 13c., from Old French bastard (11c., Modern French bâtard), "acknowledged child of a nobleman by a woman other than his wife," probably from fils de bast "packsaddle son," meaning a child conceived on an

An additional twist is added by the group of expressions that have started their life as politically correct replacements for some very nasty pejorative, and then become pejoratives themselves. In many European languages, from French to Croatian and Bulgarian, there is a word deriving from "paiderastos", the Classical Greek term for homosexual (with preference for young partners). It has been introduced at least into some of them as a high-sounding, politically correct expression for nasty ordinary words for homosexuals (like "bougre" in French). The word and/or its derivatives (say "pedè", or "peder" or "pedal") has soon become the canonical nasty term for homosexuals. These days one finds the use of politically correct terms in politically incorrect contexts, like commercials for "gay away" pills. In Croatia and Serbia one finds walls with graphiti like "Kill Roma"; the politically correct term "Roma" is re-used in the context of intense hatred. Some journalists have argued from this fact that political correctness in all its forms is just pointless; seems to me an over-reaction. But what about a theoretical account?

Here is a simple proposal for a starting point: there is a pejorative concept for homosexuals, or Gipsies, that is associated with the original pejorative (first, "bougre", second "Cigan", rougly the same as "Gypsy") but can be dissociated from it, *preserving all the ascriptions of negative properties.* The old concept then just gets transferred and pasted onto the new word. To put it more picturesquely, there is a nasty pejorative concept stalking the newly introduced politically correct counterpart-terms, and sometimes succeeding in capturing them. Relevance theory might offer a model for understanding how an ordinary reader interprets new graphiti like "Kill Roma": she starts from the clash of politically correct conception (or entry) associated with "Roma" and the "Kill..." suggestion, keeps the meaning of "Kill..." fixed and then reworks the "Roma" part, or something along these lines. Again, a lot of work needs to be done, and we need a general theory, of the sort creative philosophers of language can produce in collaboration with creative linguists, psychologists, and in the case of pejoratives, sociologists. Professor Sperber belongs to the almost incredible intersection of all four categories, so no wonder that his suggestions turn out to be so fertile.

References

Anderson, L. and LePore E. 2013. "Slurring Words", *Noûs* 47 (1): 25–48.
Ayto, J. 2000. *The Oxford Dictionary of Slang*. Oxford University Press.

improvised bed (saddles often doubled as beds while traveling), with pejorative ending -art (see -ard). (...)

Not always regarded as a stigma; the Conqueror is referred to in state documents as "William the Bastard." http://www.etymonline.com/index.php?term=bastard

Carston, R. 2010. “Metaphor, ad hoc concepts and word meaning – more questions than answers”. *Proceedings of the Aristotelian Society* 110 (3): 295–321.

Brandom, R. 1994. *Making It Explicit*. Cambridge: Harvard University Press.

Dummett, M. 1973. *Frege, philosophy of langua*ge. London: Duckworth.

Gentner, D. and Bowdle, B. 2008. “Metaphor as structure-mapping”. In Gibbs, R.W. Jr. (ed). *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press.

Hom, Ch. 2010. “Pejoratives”. *Philosophy Compass* 5 (2): 164–185.

Hughes, G. 2006. *An encyclopedia of swearing: The social history of oaths, profanity, foul language, and ethnic slurs in the English-speaking world*, Armonk: M.E. Sharpe.

Jeshion, R. 2013. “Slurs and Stereotypes”. *Analytic Philosophy* 54: 314–325.

Keessen, J. 2009. *Cardinal Men and Scarlet Women, A Colorful Etymology of Words That Discriminate*. Milwaukee: Marquette University Press.

Kennedy, R. 2002. *Nigger: The Strange Career of a Troublesome Word*. New York: Vintage Books.

Miščević, N. 2012. “Offensive Communication. The case of pejoratives”. *Balkan Journal of Philosophy* 4 (1): 45–59.

Miščević, N. 2014. “The Geometry of Offense. Pejoratives and Conceptual Spaces. *Croatian Journal of Philosophy* 14 (41): 195–214.

Perhat, J. 2012. “On Saunders, W. K. (2011), Degradation”. *Croatian Journal of Philosophy* 12 (34): 111–116.

Saunders, W. K. 2011. *Degradation: What the History of Obscenity Tells Us about Hate Speech*. New York: New York University Press.

Szabo, Z. G. 2005. “Introduction”. In Szabo, Z. G. (ed). *Semantics vs. Pragmatics*. Clarendon Press, Oxford.

Williamson, T. 2009. “Reference, Inference, and the Semantics of Pejoratives”. In Almog, J. and Leonardi, P. (eds.). *The Philosophy of David Kaplan*. Oxford: Oxford University Press: 137–158.

Wilson, D. and Sperber, D. 1986/1995. *Relevance*. Oxford: Blackwell. (page numbers are from 1995 edition).

Wilson, D. and Sperber, D. 2012. *Meaning and Relevance*. Cambridge: Cambridge University Press.

Croatian Journal of Philosophy
Vol. XV, No. 44, 2015

Argumentation as a Means for Extending Knowledge

NENAD SMOKROVIĆ
University of Rijeka, Rijeka, Croatia

In this paper I am developing the theses that argumentation is a means for extending knowledge. The theses are founded on two focal points:1. Reasoning is designed for argumentation, and 2. Argumentation process is an exceptionally successful media that provokes usage of methods reliable for the extension of knowledge. The first point relies on Sperber's and Mercier's evolutionary psychological approach to argumentation which I consider the most convincing theory in the field. Taking this ground as a departing point, the goal of the paper is to broaden this approach with epistemological insights that I base on Williamson's safety theory of knowledge.

Keywords: Argumentation, evolutionary psychology, knowledge, safety.

In this paper I am going to combine two apparently distinct approaches to argumentation. The first one is an evolutionary psychological approach to argumentation while the second is a rather highly theoretic epistemic understanding of knowledge. I am going to argue for the claim that the argumentation process is a particularly good means for the extension of knowledge. Also, there are two focal points around which the paper is organised. The first point is

a) The function of reasoning and the nature and the structure of the argumentative process, while the second is
b) Argumentation process as an exceptionally successful media that provokes usage of methods reliable for the extension of knowledge.

Concerning the first point, Dan Sperber and Hugo Mercier do excellent work in a series of articles. In their valuable contribution to the cognitive science, their naturalistic, evolutionary-oriented theory develops and explains the topic of my first point. Their great result contributing to the cognitive science is their explanation of the relationship between inference and argumentation.[1] They have shown convincingly that reasoning is evolutionally designed for argumentation, more precisely, "that reasoning is best adapted for its role in argumentation, which should therefore be seen as its main function" (Mercier and Sperber 2011: 59). When a reasoning mechanism, they argue, is employed to do what it is designed to do—finding and evaluation of reasons through argumentation—it works well and produces good performance. In this paper I will heavily rely on their results but also extend their evolutionary approach offering an epistemic contribution to the naturalistic theory of argumentation.

To connect evolutionary cognitive theory of argumentation and argumentative process' aptitude for the extension of knowledge, I need a suitable epistemological theory. The epistemological theory I am endorsing here and the notion of knowledge it develops perfectly suits my purposes. The theory I am going to employ is Timothy Williamson's highly innovative epistemological theory (2000, 2009), which bases the notion of knowledge on the safety principle. I will explore Williamson's *safety* theory in its general form as the theory of *safe knowledge*, but also in its dynamic aspect accounting for the process of *safe derivation*.

My basic idea, combining these two approaches, is to show that the argumentation process naturally guides participants to extend their knowledge given that a) reasoning has a biological function to work optimally in argumentation, and b) that the very argumentative process is structured so that it optimally supports participants' epistemic curiosity to acknowledge whether a proposition P (the object of discussion) is true.

Let me take my starting point at two statements emphasised by Sperber and Marcier. Supporting their theses with a great number of research, they claim:

1. Advantage of group reasoning theses (AGRT): Groups do better at reasoning tasks than individuals, and, in some cases, do even better than *any* of their individual members (occasionally even better than their best member).

An effective account for this asymmetry between individual and group performance is:

2. Evolutionary thesis (ET): reflective reasoning has been designed by evolution as a communicative competence (rather than aiming at enhancing individual inference).

They claim, more precisely "that reasoning is best adapted for its role in argumentation, which should therefore be seen as its main function" (Mercier and Sperber 2011: 59). When a reasoning mechanism is employed to do what it is designed to do—finding and evaluation of reasons through argumentation—it works well and produces good per-

[1] See Mercier and Sperber (2011) and Sperber and Mercier (2012).

formance. My view concerning argumentation as a means for extending knowledge is quite in accord with these two claims.

Let me put forward some preliminaries. In a very general sense, I consider argumentation to be the most advanced form of communication in the sense that it requires participants' reflection and engagement of their inferential abilities in a much higher degree than it is the case in other forms of communication. I also understand argumentation as a social, two-sided (usually informationally asymmetric) temporal process in which two sides enter, each with some initial stock of beliefs, and through the process consisting of producing and evaluating arguments, each of them eventually reach the point where they know more than they did before.

It is important to emphasise here that as a central and important situation of argumentation I consider the case where addresser sincerely believes that P but does not know whether P. She has some initial beliefs about proposition P, but neither she nor the addressee still know whether P is true or false. I am arguing that *curiosity* whether P is true is, at lest implicitly, an important motive for entering the argumentation process. It might seem that the urge to convince the other side is the strongest motive in argumentation. But, it is hard to convince someone that P is true if it is not the case and the addresser is uncertain about that.

The basic form of curiosity in the argumentation process is whether a proposition P, relevant to the cognizer, is true or false. I am calling this type of curiosity *propositional curiosity*, in contrast to other types, such as the curiosity whether the bus will depart on time, or whether she is pretty as it is said. In the argumentation process, the propositional form of curiosity underlines all other curiosity forms. One might object that participants in the argumentation process can be, and often are, curious in a non-propositional way. For instance, one might enter the argumentation process just being curious whether he is good enough to win in a discussion. But, the discussion is always organized around a claimed proposition P. After all, the addresser, claiming that P, counts as a winner only if she has convinced the other side that P is true. Although the question of P being true or false at the beginning of the process might be non-existent for her, at some point in the argumentation process the question whether P is true will arise and become important. Both sides are propositionally curious whether P. But, I am arguing that the propositional form of curiosity is implicitly present as a motivating element. The propositional curiosity, which up to a certain point in argumentation was implicit, at that point turns out to be explicit and important.

Here is a rough idea about the structure of the argumentation process. Since the argumentation process is the most advanced form of communication, let us see how to get from communication to argumentation. Communication is a central form of social practice, which, besides other purposes, has a goal of giving and receiving information. In a communication process one can be informed truly or be deceived. The advantage of true information is obvious as well as the danger of being misled. To judge the putative information correctly one should possess a number of cognitive abilities among which the inferential ability is one of the most important. To use Sperber's and Mercier's formulation, the inferential ability and its result, reasoning, is traditionally meant as "mental action of working out a convicting argument, the public action of verbally producing this argument so that other will be convinced by it and the mental action of evaluating and accepting the conclusion of an argument" (Sperber and Mercier 2011: 59).

Communication as a social practice is a process in which the informer claims a statement (conveys a piece of information) and an addressee evaluates the acceptability of the conveyed information. The information we are interested in might have a form of a sentential claim that something is the case (that P). The informer is claiming that P, while an addressee, on the other hand, evaluates (in most cases implicitly, intuitively) the acceptability of the claim. In such a simple communication process, the addressee evaluates the trustworthiness of the source and checks the consistency of the content of a claim with her previously held beliefs. It is hard to believe that one will blindly trust the source. Rather, the degree of trust depends on the context, the addressee's interest and the relevance of information.

However, human communication is usually not as simple as it is stated. In human communication, participants usually not only give and receive information but the addresser also offers supporting reasons or good ground for her claim. The addressee not only evaluates the trustworthiness of the source and the consistency of the content of a claim with her other beliefs, but also the connection between the reason and the claim. Here the reasoning, as the ability for producing arguments and evaluating them, comes into play. *It is the form of communication consisting of the claim, reason(s) (or evidence) supporting it and the relation between the claim and the reasons, where the addresser produces a claim and reasons while an addressee evaluates it, that I will consider as argumentation.*

Now, we have a familiar structure. In such a simple argumentation situation the addresser claims that P and P is a proposition of the form "an object O has a property F". In support of her claim she provides reasons $r_1, \ldots, r_n$. Claim-content P stands in a kind of consequential relation to $r_1, \ldots, r_n$. The addressee evaluates the argument finding reasons–claim relation acceptable or refutable. In both cases one makes a decision whether to accept or refute the relation.

Let us take the central case of argumentation situation to be, as it has been mentioned, the one in which the addresser sincerely believes what she is claiming to be true (or highly probable) and has a good ground (according to her own lights) for it, while the addressee wants to make an effort to find out whether this is the case. Each of the participants is primarily interested (curious) in the truth of the claim. The addressee wants to avoid false beliefs by hypothesis, while the addresser, sincerely believing that P is committed to the belief that belief in P is true (because P implies that P is true).[2] This is the argumentation format I am interested in, letting aside other possible argumentation situations. The stipulation that each participant is primarily interested in the truth of the claim is to be examined in more detail.

In their attempt to convince other, people actually tend to neglect the objective search for the truth and "they are typically looking for arguments and evidence to confirm their own claim, and ignoring negative arguments and evidence unless they anticipate having to rebut them" (Mercier and Sperber 2011: 63). *The confirmation bias,* Mercier and Sperber claim, "is a feature of reasoning when used for the production of arguments" (Mercier and Sperber 2011: 63). Namely, given that addresser, believing (and claiming) that P is committed to the belief that a belief in non-P is false (for P implies that P is true), has no motivation to find out whether P is false. The confirmation bias has been detected as a common fallacy in individual reasoning or in reasoning with like-minded peers. In the argumentation process, the confirmation bias leads participants to reinforce their initial beliefs producing individual polarization. Nevertheless, I will try to show that the argumentation process has resources to remedy the confirmation bias.

Tim Williamson depicted the bias in a stronger form: "One is bound to think any given belief of his own superior in truth-value to the contrary beliefs of others" (Williamson 2007: 247). Fortunately, Williamson offers a remedy that nicely fits my approach. He says:

> But sometimes we step back of our beliefs and regard them as psychological phenomena on a par with the belief of others, in equal need of both psychological explanation and epistemological criticism. (Williamson 2007: 247)

I will try to show that the argumentation situation is exactly the situation that prompts one to "step back" in the described way. Furthermore, I will argue that the very argument structure has the effect of prompting participants to "step back" and that this very structure enhances participant's inferential abilities.

This *stepping back* provides additional motivation for the addresser to evaluate her own beliefs as well as her inferential steps. If it were the case it would be plausible to suppose that firstly, participants are interested in the truth and, secondly, that applying the appropriate notion of knowledge, it would be plausible to show that the argumentation process is a good means for the extension of knowledge. Let me first offer an account of how argumentation enhances individual reasoning abilities and therefore increases the chances for reaching the truth.

[2] It is of course possible that the addresser does not believe that P is true (sincerely believes that P is false) and still wants to deceive the addressee that P is the case, but I am not counting this situation as an argumentation process.

Just as the tendency to commit confirmation bias is a natural tendency in reasoning, the argumentation process has powerful resources for *de-biasing* the bias. The integral part and an important feature of the argumentation process, detected by Sperber, is the 'epistemic vigilance' of participants (Sperber et al. 2010). Participants in an argumentation are likely to display a higher degree of epistemic caution, to be epistemic vigilant, more that it would be the case in individually acknowledging the true state of the affair. This notion of 'epistemic vigilance' is based on the importance of getting the true information and avoiding the false one.

The possible explanation for such articulated epistemic vigilance lies, I suggest, in the *difference in roles* participants may play in the argumentation process. The epistemic status of the addresser and the epistemic status of the addressee points to quite different roles of each of the two in the argumentation process. The role of the addresser is to claim that P and to provide (or make explicit) the reasons that support the claim. As we said before, the addresser is inclined to provide more and more reasons for her claim. Concerning this tendency, empirical investigations report about a strong *confirmation bias* in the argumentation. It should be noted, however, that participants might switch their roles in the process. At the same time, in the role of the addresser, one is under obligation to display reasons for one's claim and to get them in the relation to the claim in a way maximally understandable and clear to the addressee. In other words, one will make it accessible/knowledgeable to the addressee as much as possible. Here is how Dan Sperber describes the process:

> One way to persuade one's addressees is to help them check the consistency of what one is claiming with what *they* believe, or even better if possible, to help them realize that it would be inconsistent with what *they already believe* not to accept one's claim. The communicator is better off making an honest display of the very consistency addressees are anyhow checking. This amounts to, instead of just making a claim, *giving reasons* why it should be accepted, *arguing* for it. (Sperber & Mercier 2012)

On the other hand, the role of the addressee is to evaluate the acceptability of P assessing the reasons and their relation to the claim. The addressee will do her best to find counterexamples to the addresser's argument and in this way falsify her claim, if possible. The role of the addressee thus creates a significant counterbalance to the confirmation bias the addressor is prone to. The fact that participants are committed to playing different roles and their different goals gives the argumentation process the power to add something to the extension of knowledge. Let me elaborate a little. The participants in the argumentation process are individual reasoners with their reasoning competences and abilities as well as their practical and epistemic goals. Let us take, to reiterate, a simple argumentation situation in which one party is claiming that P and the other party is suspicious whether this is the case.

Taking that the addresser sincerely believes that P is true while the addressee does not want to be misinformed and led astray, they, obviously, enter the process with different, even opposing goals. There is, on the one hand, a practical goal of convincing one, and also a practical goal to avoid misinformation, on the other. But, it seems that behind these practical goals, there is a more fundamental one. I suggest that it is the epistemic goal of acquiring knowledge. Each party is doing whatever is in their intellectual power to find out whether that which is claimed is true.

Here is a rough and overall description of the process. In argumentation, both the addresser (S) and the addressee (S') are involved in two distinct courses of action, producing the arguments and evaluating them, because of changing their roles in discussion. Production of the argument as an inferential action is naturally subject to the confirmation bias. The evaluation of the produced argument, on the other side, is not. It has an effect of de-biasing the process. These two inferential actions mutually support each other creating a method for broadening initial knowledge possessed by both participants.

This structural description, stated so far, obviously needs its dynamic part, the epistemic account for the advancement of knowledge in the process. I take the initial epistemic situation to be something like the following. Both parties have some initial knowledge concerning P that they arguing about. S sincerely believes P, but does not know whether P is true. Still, she does know some facts concerning P. S' is also acquainted with some facts concerning P, in less certain way then S. S' also doesn't know whether P. But, both of them, being involved in producing and evaluating arguments, are in a way forced to be curious, whether P.

Let me use an example. Poirot and his assistant, Colonel Hastings, investigate the case of a homicide. In this example, Hastings is the one who is proposing the solution. In the argumentation process, Colonel Hastings claims that it is the gardener who is the murderer and offers reasons. Reasons are based on the following evidence, i.e., there is the knife, which, beyond reasonable doubt, has been proven to belong to the gardener. There is victim's blood on the knife. The gardener has no alibi. Poirot might accept the evidence as sufficient for the conclusion or deny it evaluating the argument, re-examining evidences. He may point out to some equivalence unnoticed to the good colonel. Also, he can help him to recognize some obviousness the colonel did not recognize before. Poirot and Hastings have some knowledge about P, whether the gardener is a murderer, but do not have the knowledge whether this is true or not. Through the argumentation process they have to safely extend their initial knowledge in the way that, in the optimal case, they reach the point of *knowing* who the murderer is, or in the less optimal case, to know more than they did at the beginning of the argumentation.

Let me introduce the epistemic theory I find most appropriate. The epistemic theory that I take as suitable for the idea of extension of knowledge is Tim Williamson's knowledge theory based on the *safety principle*. Williamson's theory is one of the few notable attempts to improve Gettier's *justified true belief* theory of knowledge, and at the same time, to give an alternative solution to famous Nozick's *sensitivity* based theory. Due to space limitation, I am going to give a very sketchy presentation of Williamson's theory.

Avoiding formulating the theory in terms of necessary and jointly sufficient condition for knowledge, Williamson founded the model of knowledge in the concept of safe avoidance of error. Accordingly, S knows P only if S is safe from error. As Williamson stated: "If one knows, one could not easily have been wrong in a similar case" (Williamson 2000: 147).[3] Skipping the subtlety, I am going to maintain only those moments of the theory relevant for my purpose.

The cases or conditions important for determining the similarity can be determined by the set {S,M,P,T}, where S denotes a cognizer, M a method, P proposition and T a time in which S believes P. A further formulation is the following:

"In a case α one is safe from error in believing that (a condition) C obtains if and only if there is no case close to α in which one falsely believes that C obtains" (Williamson 2000: 126–7). A condition C is specified by "that" close relative to an agent and the time. In our Poirot example, C is the belief "gardener is a murderer".

Williamson tackles the concept of safety differently at different places, introducing the concept (together with concepts of "reliability and unreliability, stability and instability, safety and danger, robustness and fragility") he claims that they refer to modal states. They concern "what would easily have happened" (Williamson 2000: 123). In the epistemic situation safety is attributed to knowledge and belief. In such a situation what easily would have happened is the cognizer being wrong or right in respect to some proposition P. S's being right in a particular situation counts as knowledge if S could not easily have been wrong in similar situations.

Since safety principle does not put a necessary and sufficient condition upon knowledge, it is consistent with counterfactual: if P had been false, one would (or might) still have believed P. Accordingly, safety does not imply omniscience. What is important is that one can start with a non-safe knowledge and reach the safe knowledge eventually. Furthermore, *safe* is a gradable adjective. One state of believing can be more or less safer than the other. I will take that safe knowledge is a final state of the process in which one starts with a less safe knowledge reaching a more safe state. As Williamson says, "One's total evidence

[3] This formulation of safety condition significantly differs from Sosa's counterfactual formulation: "If S were to believe P, P would be true" (Sosa 1999: 146). Formally: $(B(P) \square\!\!\rightarrow P)$.

is one's total knowledge" (Williamson 2000: 9). Applying elements of Williamson's theory to the argumentation process, let us start from the following formulation:

> S safely believes P in situation α, if, using method M in time T, S truly believes P and could not easily have been wrong in similar situations β.

Applying Poirot's example at the time T, Hastings believes that the gardener is a murderer, founding his belief on evidence (the knife, blood and alibi) about which he has safe knowledge. Let us take that this knowledge is based on perception and direct experience. Also, Hastings forms a further belief that the gardener is the murderer, basing his belief on the *affirming of consequence* (AC) inference.

Of course, AC is the unreliable *method* and cannot bestow safe knowledge. But, the argumentation process is the reliable media that can provide means for the revision of false beliefs. Poirot's evaluation of Hastings' argument will make it clear to Hastings that he needs to abandon AC and use the *modus ponens* instead. Furthermore, Poirot might re-examine the evidences, refute some of them, find a new piece of evidence and offer the conclusion that it is not the gardener but the driver who is the murderer. To generalise a little, the argumentation process is the reliable media that is able to select, among those methods at hand, the ones that are more reliable. Using reliable methods, participants are in the position to reach safe knowledge. In this way, the argumentation is a media that increases their deductive competences. Learning to use reliable methods and increasing their deductive competences, it is likely that one will acquire safe knowledge (about P), in the way that one could not easily have been wrong in similar cases.

References

Mercier, H. and D. Sperber. 2011. "Why do humans reason? Arguments for an argumentative theory". *Behavioral and Brain Sciences* 34: 57–111.

Sperber D. and H. Mercier. 2012. "Reasoning as a Social Competence". In Landemore and Elster (eds.). *Collective Wisdom: Principles and Mechanisms*. Cambridge: Cambridge University Press: 368–392.

Sperber, D., Clement, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., Wilson, D. 2010. "Epistemic Vigilance". *Mind and Language* 25 (4): 359–394.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, T. 2007. *Philosophy of Philosophy*. Oxford: Blackwell.

Williamson, T. 2009. "Probability and Danger". The Amherst Lecture in Philosophy. (http:/www.amherstlecture.org/)

Croatian Journal of Philosophy
Vol. XV, No. 44, 2015

Cognitive Pragmatics and Variational Pragmatics: Possible Interaction?

DUNJA JUTRONIĆ
University of Split, Split, Croatia

In this paper I attempt to look into a possible way in which cognitive pragmatics can help out variational studies in explaining the processes of language change. After broadly setting the scene this article proceeds by giving basic information about variational pragmatics. Then it concentrates on Sperber and Wilson's relevance theory and its possible interaction with social sciences, namely its possible application in sociolinguistics. I next present my own research of Split (urban) dialect/vernacular change where I concentrate on explanatory side, asking which explanation would be the best one for the changes of some variables in the dialect. The interpretation and discussion of the findings preceed the discussion of salience as the explanatory tool for language change as seen from cognitivists and variationists with the hope that such discussions can bring closer cognitivists, i.e. relevantists, to sociolinguists, i.e. variationists.

Keywords: Cognitive pragmatics, relevance theory, variational pragmatics, sociolinguistics, urban dialectology, salience.

1. *Introduction*

Quite early in the history of pragmatics two different ways of doing pragmatics, or, one might say, two schools of thought established themselves. One of these ways/schools can conveniently be described as the Anglo-American tradition of pragmatics, the other as the (Continental) European tradition (Jucker 2012: 501). Levinson had already pointed out the distinction between these two traditions (Levinson 1983: 2, 5–6). Anglo-American pragmatics is concerned with the study of meaning that arises through the use of language. The (Continental) European school of thought takes pragmatics to have a much wider range of tasks, it is a specific perspective for studying language in general. Representing the European pragmatic turn Verschueren provides a typical definition: "Pragmatics can be defined as *the study of language use,* or, to employ a somewhat more complicated phrasing*, the study of linguistic phenomena from the point of view of their usage properties and processes"* (1999: 1, italics in the original). The difference between these two schools of thought is clearly reflected in the relevant textbooks and handbooks of pragmatics. Horn and Ward (2004), for instance, explicitly exclude the broad, sociologically based European view of pragmatics from their *Handbook of Pragmatics* and focus on the "more narrowly circumscribed, mainly Anglo-American conception of linguistic and philosophical pragmatics and its applications" (2004: xi). The general stand seems to be that the Continental European tradition is too all-inclusive and therefore lacks a clear delimitation and defies an attempt to establish a coherent research agenda (Jucker 2012: 502). On the other hand, the textbooks by Mey (2001) and Verschueren and Ostman (2009a) clearly adopt the wider Continental European approach which means they include questions about the social and cultural contexts in which language is used. Thus Verschueren and Ostman say: "Pragmatics is defined as the cognitive, social, and cultural science of language and communication" (2009b: 1). In this wider approach, pragmatics, very importantly, intersects predominantly with sociolinguistics and, more generally, social sciences, with a focus on interpersonal and social meaning rather than sentential and textual meaning. In Verschueren's words there is "no strict boundary between pragmatics and some other areas in the field of linguistics, such as discourse analysis, sociolinguistics, or conversational analysis" (1991: 1).

After broadly setting the scene this article proceeds as follows: Part 2 gives more information about variational pragmatics. Part 3 concentrates on Sperber and Wilson's relevance theory and its possible interaction with social science, namely its possible application in sociolinguistics. Part 4 presents my own research of Split (urban) dialect/vernacular change where I concentrate on explanatory side, asking which explanation would be the best one for the changes of some variables in the dialect. Part 5. is the interpretation and discussion of the findings. Part 6 concludes on the discussion of salience as the explanatory tool for language change as seen from cognitivists and variationists.

2. *Variational pragmatics*

Recently there has been a growing interest in different dimensions of language-internal variations of pragmatics. Schneider and Barron who have given this field of investigation its name, namely Variational pragmatics, define it as follows: "Variational pragmatics can be considered a twin discipline of historical pragmatics, which was established in the mid-1990s (cf. Juncker 1995). Briefly speaking, historical pragmatics investigates pragmatic variation over time, whereas variational pragmaticis investigates pragmatic variation (geographical and social) in space. Also, while historical pragmatics is conceptualized as the intersection of pragmatics with historical and diachronic linguistics, variational linguistics is conceptualized as the interface of pragmatics with variational linguistics, i. e. with modern dialectology, as a branch of contemporary sociolinguistics" (2008b: 1). Schneider and Barron (2008) develop a framework for variational pragmatics in which they envisage five types of language variation as possible dimensions of investigation: regional, socio-economic, ethnic, gender, and age variation. For my discussion the most important aspect is their stress on the intersection of variational pragmatics and dialectology. Thus looking at a broader scheme (and in order to help the reader) that the field of pragmatics encompasses I offer the following divisions as presented in Figure 1. Pragmatics studies 1. cognitive, 2. intercultural, 3. historical and 4. variational aspects of language. Furthermore variational pragmatics interfaces with variational linguistics, i.e. with its most prominent subfileds, (urban) dialectology and sociolinguistics.

PRAGMATICS			
cognitive	**intercultural**	**historical**	**variational**
			variational linguistics
			i.e. socioliguistics & dialectology

3. *Relevance theory*

Relevance theory is known primarily as "a cognitive psychological theory" as Sperber and Wilson stress once again in their most recent book *Meaning and relevance* (2013: 281) so one can rightly ask how it can have any relevance to variational pragmatics as defined above? However, Sperber and Wilson have written an article under the title "Relevance theory and the social sciences" where the authors say: "Some commentators have described the relevance-theoretic approach to communication as psychological rather than sociological. Often, this is intended as a criticism. We would like to respond by reflecting in very general terms about possible interaction between relevance theory and research programmes in the social sciences….among other things we would like to help bring about a redefinition of disciplinary boundaries, including those between the cognitive and social sciences, and we see our work as contributing to both domains" (1997: 145). They think that putting stress on inferential communication (in contrast to code model of communication) is in fact pointing to the intrinsically social aspect of communication, not just because it is a form of interaction, but also, less trivially, because it exploits and enlarges the scope of basic forms of social cognition. Furthermore ostensive or non-ostensive uses of the act of communication itself convey claims and attitudes about the social relationship between the interlocutors. They conclude: "*Right or wrong, this is a strong sociological claim*" (1997: 146, italics mine). They also stress that sociolinguists have been particularly concerned with these aspects of verbal behavior, and have studied them with sophistication and insight but that they themselves (in their book Relevance*)*, *largely ignored them* (italics mine). They did not mean by this to deny their importance, or to express a lack of interest in the issues or the work done; they merely felt that, at that stage, they could best contribute to the study of human communication by taking it at its most elementary level, and abstracting away from these more complex (social) aspects. They conclude: "So far, the contribution of relevance theory to the study of human communication has been at a fairly abstract level. However, it seems to us to have potential implications at a more concrete sociolinguistic level" (1997: 146).

It is evident from the above that Sperber and Wilson are much aware of possible and fruitful interaction of cognitive and social since they also conclude their Postface to the second edition of *Relevance* with the words: "Two important and related domains have hardly been explored at all from a relevance-theoretic perspective: the theory has been developed from the point of view of the audience of communicative acts, and *without taking into account the complex sociological factors richly studied by sociolinguistics*." (1995: 259, italics mine).

The only attempt (that I know of) of bringing cognitive and social together within the relevance framework is the study by Gisle Andersen presented in her book *Pragmatic markers and sociolinguistic variation - a relevance theoretic approach to the language of adolescents*.[1] Andersen says that her study is an attempt to combine sociolinguistics and relevance theory. That such a combinatory approach can be fruitful is implied by the abovementioned comment in which Sperber and Wilson claim that social character and context of communication are *essential* to the wider picture.

4. *Sociolinguistic research*

My own work falls under variational sociolinguistics. I have done much research into urban dialectology and what I want to do in the rest of the paper is present just a small part of my findings and try to show that there is a possible meeting place of cognitive and variational studies that can help us to at least clear up the grounds of some explanatory problems in, maybe, both fields.

[1] For more encompassing cognitive pragmatic aspects of communication see Bara (2010) and Schmid (2012).

The investigation into an urban vernacular aims to answer the question: Why did some dialectal variables disappear and why are some in the state of variation and others are still firmly used in, this case Split, urban vernacular? In order to provide a plausible explanation, the chosen variables were analyzed within the sociolinguistic framework using the **principle of salience** as a theoretical explanatory tool. If we say for a linguistic feature that it is salient, then we consider that feature to be perceptually and/or cognitively, or socially marked. In my investigation this principle was formulated in the following way: *Those dialectal characteristics/features that the speaker of standard language feels as socially salient and thus unacceptable, or as some kind of 'mistake' disappear from the dialect first. Stigmatized or salient characteristics go out faster from the dialect than less stigmatized or non-salient characteristics.*

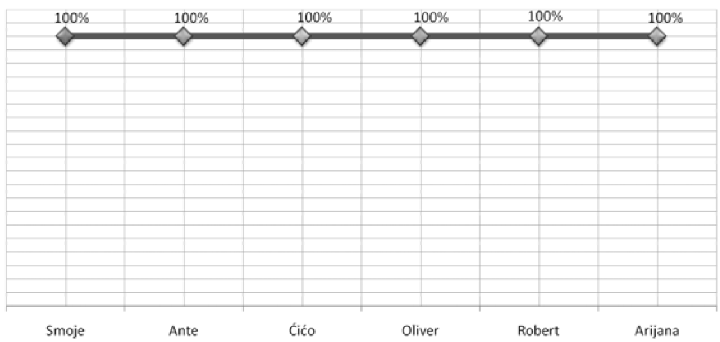
On this occasion I present only the research into syntactic variables and see how the principle of salience can be applied there, i.e., how can we explain dialect change with this theoretical tool? But first some very basic information on Split dialect and methodology used.[2] Split is a city on the Adriatic coast in Croatia. Once a small town (18,500 inhabitants in 1900), it has grown rapidly since World War II so that today it numbers about 200,000 inhabitants. The presented sociolinguistic analysis of Split urban dialect/vernacular is the analysis of the dialect as we find it today. In order to follow the linguistic changes of the Split dialect under the influence of the standard language, it is necessary to know that there are three main dialect groups in Croatia: Štokavian, Čakavian and Kajkavian, named after the interrogative-relative words for "what" in each dialect which is *što*, *ča* and *kaj* respectively. According to their reflexes of proto-Slavic /e/ (called *jat*), these dialects are traditionally also subdivided into *ijekavian*, *ekavian*, and *ikavian* varieties. For example, the word for "*milk*" is *mlijeko/mleko/mliko*, the first word being part of the standard language and the last two of the nonstandard varieties, ekavian and ikavian. But this is, of course, an idealized division, since there are many areas where the mixed varieties occur. Štokavian in its ijekavian form is the official standard language in Croatia. Apart from these most basic and widespread differences that are others as indicated in the presentation of the syntactic variables for this occasion. As far as methodology is concerned, in order to trace the changes speakers of the dialect are divided into 3 age groups which is one of the customary ways in sociolinguistic analyses. The older generation (in the graphs: Smoje and Ante), the middle generation (Ćićo and Oliver) and the third, young/ younger generation (Robert&Arijana; Petra&Marijana). The statistical analysis is obligatory in sociolinguistic studies of this kind if we are to establish with still relative certainty the percentage of features used. For this occasion I am presenting only

[2] For detailed study of dialect change in Split see Jutronić (2010). Summary in English pp. 441–465.

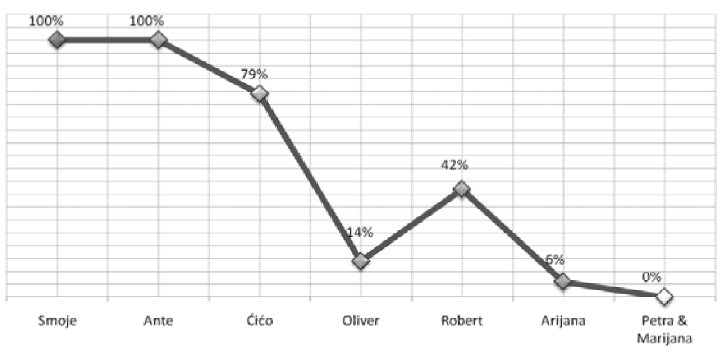
four syntactic variable under change in Split vernacular: 1. The use of the construction from/of+ genitive; 2. The use of the accusative/locative distinction; 3. The use of contracted form *mi je* > *me*; 4. The use of the interrogative/relative pronoun ča.

The results are presented in the graphs.

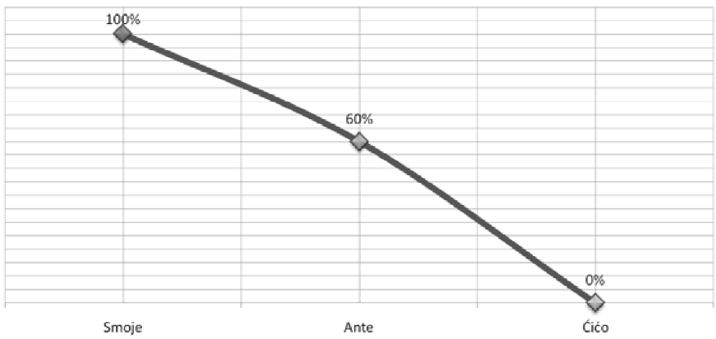
Graph 1. The Use of the Construction from/of + Genitive
Example: Prsten o' znata *insted of* znatni prsten (=ring made of gold *instead of* golden ring)



Graph 2. The Use of Accusative/Locative Distinction
Example: Bija san u Split *instead of* Bio sam u Splitu (I was in Split)



Graph 3. Th Use of Contracted Form: mi je > me
Example: Draga mi je Ravena > Draga me Ravena (I like Ravena)



Graph 4. The Use of the Interrogative-Relative Pronoun ča
Example: Ča si radija? (What did you do?); Reci ča očeš (Say what do you want)

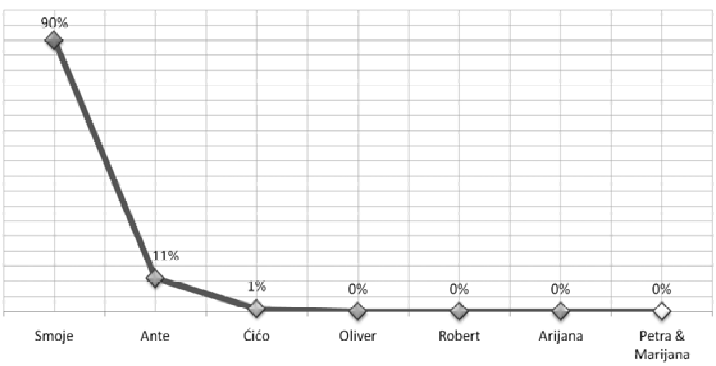


Table 1. Scale of salience -- syntactic variables

salient	variable		nonsalient
(changed)	(varying)	disappearing	(unchanged)
ča	*A/L*	---	*of* + G
mi je > me			

5. *Interpretation and discussion*

A **non-salient feature** is the construction of + genitive (*prsten od zlata* 'a ring of gold') instead of an adjectival attribute as in *zlatni prsten* ('golden ring') is, according to the Croatian linguist Finka, a "quite widespread" feature in Čakavian (1971: 62). This construction appears with all generations in 100% of the cases as evident from the graph. The use of this construction is not specific to Čakavian. It is heard often in other dialects and in the standard language too, in its conversational style. This must be one of the reasons why it is used so much with the young generation. It is nonsalient and thus it is not sanctioned and it firmly remains in the Split vernacular.

A **salient** syntactic characteristic in the Split vernacular are the contraction of *mi je* > *me* (*mi je* literally meaning 'to me is') and the interrogative-relative pronoun *ča* ('what'). The former construction is found only with the older generation. Here are a couple of extra examples: *ruku me deboto izija* ('he almost ate my hand'), *kad me skočija na posteju* ('when he jumped on my bed'), *puno me drago* ('I like it a lot'). This contraction is obviously stigmatized and we do not find it in use with the middle generation. The young generation does not even know about this feature. When you use the phrase they are rather surprised and often they do not understand what you mean.

Another **salient** feature is the interrogative-relative pronoun *ča* ('what'). The Čakavian dialect and its various local manifestations got its name from the interrogative-relative pronoun *ča.* Finka says: "Wherever we find a trace of the pronoun *ča* there we find other very vital and essential Čakavian characteristics" (1979: 15).[3] The pronoun *ča* is not a matter of prestige in the Split vernacular any more. The Štakavian pronoun *šta* or Štokavian form *što* (what) has replaced *ča* in all contexts. It is not consistently used even with the older generation. However, *ča* is used in 100% of the cases in the songs sung by the popular Dalmatian singer Oliver Dragojević and other singers from Dalmatia, and it is used in a kind of nostalgic way in order to strengthen the Dalmatian timbre and spirit of those songs. Speaking about the use of some archaic dialectal forms Croatian sociolinguist Kalogjera remarks: "… using from time to time this (archaic) variety the speaker is aiming at the 'authentic' old, local speech. As if he had some covert feeling of 'historicity' of his local vernacular which 'today it is not as it used to be'" (1992: 129). Within the framework of the principle of salience this pronoun is felt as being overly Čakavian and thus it is being avoided or not used at all.

The Čakavian characteristic of mixing locative and accusative in the example sentence: *Bija san u Split* instead of *Bija san u Splitu* ('I was in Split') shows that the latter sentence has the correct locative ending of *–u (Splitu).* This is a feature that **varies** but still persists

[3] Note that this was stated in 1979.

today. The question arises as how to explain this variation? It is not that easy or straightforward to apply the principle of salience in this case as it was in the previous cases. Why? This is a syntactic feature that should be salient since the speakers use the wrong case endings which one definitely hears as a 'mistake' and consequently it should be avoided and/or stigmatized. It is interesting that Finka speaks about it as "*the most serious disorder* in the Čakavian forms which was probably the result of the influence of a language called Dalmata" (1971: 46, italics mine). Thus this feature should be dying out of the Split vernacular today – but it is not. It is obviously not felt as a 'disorder' but it rather seems that this feature has a covert (social) prestige for Split vernacular speakers. Here we have to go back to the question of salience and its explanatory potential.

6. *More on salience*

British sociolinguist Peter Trudgill (1988) offers a very careful elaboration of salience and its most explicit application to language change. In my research I primarily relied on his discussion and further criticism of P. Kerswill and A. Williams presented in their 2002 paper entitled "Salience as an explanatory factor in language change."[4]

In language change factors that play explanatory role are: 1. Language internal factors (linguistic factors) and 2. Extralinguistic factors. Extralinguistic factors are usually subdivided into: a. perceptual-attitudinal factors and b. social (socio-linguistic) factors.

If we suspect that a feature is salient for speakers because of its particular patterning, we start by checking first for language-internal factors. The received view among variationalists is that language internal factors are, in most cases, not sufficient for the explanation of language change. We must immediately look for extra-linguistic factors that might be, in this case, linked to salience. These factors might be extremely varied and sometimes complex. But discussing salience in a way that divorces it from extra-linguistic factors leads to a failure to gain insights into the social patterning of linguistic features.

What are then some questions and the problems? If we say for a linguistic feature that it is salient, then we consider that feature to be perceptually and/or cognitively or socially marked. Consequently salience is defined both perceptually/cognitively and socially. The question is: Which of the extra linguistic factors are primary? Cognitive/perceptual or sociolinguistic? Kerwill and Williams seem to hold the view the sociolinguistic factors are primary. They talk about sociolinguistic sensitivity, social relations, time scheme and intensity of contact. I myself also put the stress on the social factors in my discussion about the

[4] There are a number of newer books on salience but not that useful for the linguistic explanation. See Giora (2003) and Chiarcos, Claus, Grabski (2011). The most recent book that is more relevant and that summarizes the sociolinguistic discussion on salience is Racs (2013).

persistence of A/L mixture in Split dialect. Thus I concluded: "Although it (A/L mixing) should be socially stigmatized in the wider context of the standard language, it seems to be taken as an acceptable sign of localism, as something that every speaker of standard Croatian knows it is a "mistake" but takes it as a characteristic feature (a little quirk so to speak) of speakers from Dalmatia" (2010: 458). And everything Dalmatian is in most cases taken as positive since it is connected to the sea, the characteristic lazy Dalmatian attitude and happy-go-lucky behavior. Kalogjera says: "Thus in Zagreb the Dalmatian dialect is connected to a vigorous temperament, fickleness in love, garrulousness, pleasant laziness and at times with an unscrupulous brazenness in social life" (1965: 29).

On the other hand we have voices that favor the perceptual/cognitive factors as more explanatorily important. For example, Willem Hollmann and Anna Siewierska (2006: 27) argue that cognitive perceptual factors are primary, not the social ones. Their reasoning is that linguistic items will normally be more or less free from social values when *they first arise*. They claim that it is only after they have emerged that social forces can start working on them. In other words, they suggest that ultimately it is the cognitive-perceptual constraints that make a form more or less liable to becoming subject to social evaluation and patterning. Here we turn back to Sperber and Wilson's relevance theory for some possible insight.

7. *How can relevance theory help?*

Sperber and Wilson ask: "How does verbal communication convey social claims and attitudes, and play a role in the 'negotiation' of mutual relationships? The cognitive processes at work in the communicator, and the social character and context of communication are, of course, essential to the wider picture, to the study of which we hope relevance theory can contribute, and from which it stands greatly to benefit" (1995: 279). They have two proposals: Proposal A: The most basic claim of relevance theory (the First, or *Cognitive Principle of Relevance*) is that the pursuit of relevance is a constant factor in human mental life. Proposal B: *Communicative Principle of Relevance* is that ostensive communication creates uniquely precise expectations of relevance in others. Ostensive communication is the most important means by which the psychological tendency to maximise relevance is socially exploited.

With these two proposals or principles in mind we can pose our question again: Why are some non-standard realizations, although salient, so resistant to change? Taking Sperber and Wilson Proposal A the suggestion is that it is possible that the subjective relevance that speakers and listeners attach to salient phenomena is a result of processes of appraisal and possibly speech accommodation. Thus salience might primarily be seen as *unreflective intuitions of relevance* together with procedural inferential process. In other words we can say that

Cognitive principle of relevance suggests *that ultimately it is the cognitive/perceptual constraints that make a feature more or less liable to becoming subject to social evaluation and patterning.* This would give support to Hollmann and Siewierska (2006) who argue that cognitive perceptual factors are primary, not the social ones. On the other hand Kerswill and Williams (and myself) might retort that (unreflective) perceptual/cogntive factors have to be completed with reflective social evaluation. Otherwise they would not surface at all! There surely are different (unreflective) cognitive/perceptual factors that contribute to a variant's sociolinguistic salience but again if they are not completed with reflective social evaluations we would not know about them at all.

8. *Conclusion*

There has recently been a plea for sociolinguistics to integrate both theoretical and methodological developments from cognitive linguistics. Gries says: "I believe it is necessary to recognize that something can by definition only be sociolinguistically relevant if, as some point of time, it has passed through the filter of human mind (2013: 7). On the other hand cognitivists apologize that their approach to communication is primarily psychological rather than sociological, too. At the same time cognitive linguists claim that cognition is embedded in interaction. Social aspects of language must be taken into account. Some like Sperber and Wilson (2012) try to stress the possible interaction between their cognitivist relevance theory with research programs in the social sciences *and they see their work as contributing to both domains*. In our discussion of language/dialect change where salience was used as a possible explanatory tool it hopefully became evident that a worked out explanatory role of salience in explaining language change can profit from cognitivists as well as sociolinguists.

In sum to restate: Pragmatic research in the first decade of the 21st century has been characterized by an unprecedented diversification of subfields of pragmatics**.** My discussion here can be taken as an attempt to look into a more coherent way in which cognitive pragmatics can help out variational studies in explaining the processes of language change and the hope is that such discussions can bring closer cognitivists, in this particular discussion relevantists, to variationalists. In other words there is room for relevantists and variationalists to find a fruitful meeting ground.[5]

[5] A short version of this paper was presented at Sperber symposium held March 18th 2013 in Rijeka and was organized by the Philosophy Department in Rijeka together with the Croatian Society for Analytic Philosophy. I thank the participants, particularly Dan Sperber, for many comments.

References

Allan, K. and Jaszczolt, K. (eds.). 2012. *The Cambridge Handbook of Pragmatics*. Cambridge: Cambridge University Press.

Andersen G. 2001. *Pragmatic markers and sociolinguistic variation: A relevance theoretic approach to the language of adolescents*. Amsterdam: John Benjamins.

Chiarcos, Ch., C. B. and M. Grabski (eds.). 2011. *Salience: Multidisciplinary perspectives on its function in discourse*. Berlin: Mouton De Gruyter.

Bara, B. 2010. *Cognitive Pragmatics, the Mental Processes of Communication*. Cambridge: The MIT Press.

Gries, S. 2013. "Sources of variability relevant to the cognitive sociolinguist, and corpus – as well as psycholinguistic methods and notions to handle them". *Journal of pragmatics* 52: 5–16.

Finka, B. 1971. "Čakavsko narječje" [Chakavian Dialect]. *Čakavska rič* 1 (1): 11–71.

Finka, B. 1979. "O novim tendencijama i pojavama u čakavskom narječju" [About new trends and tendencies in Chakavian dialect]. *Filologija* 9: 145–148.

Giora, R. 2003. *On our mind: salience, context and figurative language*. Oxford: Oxford University Press.

Hollmann, W. and A. Siewierska. 2006. "Corpora and other methods in the study of Lancashire dialect". *Zeitschrift für Anglistik und Amerikanistik* 54 (1): 21–34.

Horn, L. and G. Ward. 2004. *The Handbook of Pragmatics*. Oxford: Blackwell Publishers.

Jones, M. and E. Esch (eds.). 2002. *Langugage change: The interplay of external, internal and extra-linguistic factors*. Berlin: Mouton de Gruyter.

Jucker, A. H. (ed.). 1995. *Historical pragmatics: Pragmatic developments in the History of English*. Amsterdam: John Benjamins.

Jucker, A. H. 2012. "Pragmatics in the History of Linguistic Thought". In Allan and Jaszczolt: 495–512.

Jutronić, D. 2010. "Spliski govor, od vapora do trajekta" [Split vernacular, from steamboat to ferryboat]. Split: Naklada Bošković.

Kalogjera, D. 1992. "'Arhaizacija' dijalektalnog teksta" ['Archaization' of dialectal text]. *Suvremena lingvistika* 4: 127–132.

Kalogjera, D. 1965. "O odnosu regionalni dijalekt-standardni jezik" [About the relationship of regional dialect and standard language]. *Jezik* 1: 27–30.

Kerswill, P. and A. Williams. 2002. "Salience as an explanatory factor in language change. Evidence from dialect levelling in urban England". In Jones and Esch: 81–110.

Levinson, S. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Mey, J. 2001. *Pragmatics: An Introduction*. Oxford: Wiley Blackwell.

Racz, P. 2013. *Salience in Sociolinguistics*. Berlin: Mouton de Gruyter.

Schmid, H. 2012. *Cognitive pragmatics*. Amsterdam: De Gruyter Mouton.

Schneider, K. and A. Barron (eds.). 2008a. *Variational Pragmatics*. Amsterdam: John Benjamins.

Schneider, K. and A. Barron. 2008b. “Where pragmatics and dialectology meet: Introducing variational pragmatics”. In Schneider and Barron 2008a: 1–35.

Sperber, D. and D. Wilson. 1995. *Relevance, Communication and Cognition* (sec. ed.) Oxford: Blackwell.

Sperber, D. & D. Wilson. 1997. “Relevance theory and the social science”. *Multilinguist* 16: 145–51.

Trudgill, P. 1986. *Dialects in Contact*. Oxford: Blackwell.

Verschueren, J. 1999. *Understanding pragmatics*. London: Arnold.

Verschueren, J. and J. Ostman (eds.). 2009a. *Key notions for pragmatics*. Amsterdam: John Benjamins.

Verschueren, J & J. Ostman 2009b. “Introduction: The pragmatic perspective”. In Verschueren and Ostman 2009a:1–28.

Wilson, D. and D. Sperber. 2012. *Meaning and Relevance*. Cambridge: Cambridge University Press.

Croatian Journal of Philosophy
Vol. XV, No. 44, 2015

Book Review

Philip Kitcher and Gillian Barker, *Philosophy of Science: A New Introduction,* Oxford: Oxford University Press, 2014, pp. 192

Philip Kitcher's and Gillian Barker's *Philosophy of Science: A New Introduction* is the recent contribution to textbooks in Philosophy of Science. This accessible introduction is intended not only for philosophy students but also for students and interested professionals from related fields, such as science and technology studies, humanities or social sciences. Aside from professionals the book is useful and informative for every reader interested in the subject assuming she has at least minimal knowledge on the subject. Since both authors work extensively on topics that broaden the traditional discussions in philosophy of science with accounts from social epistemology, sociobiology, sociology of science, political philosophy and ethical theory, *Philosophy of Science: A New Introduction* seems as a natural summary (extension) of their previous work. Although discussions in philosophy of science address these issues for more than a decade and have moved further from the debates that dominated the twentieth century, the book represents a novel attempt to incorporate contemporary philosophical accounts of science in a fruitful introduction. Authors reconsider the core questions in philosophy of science by taking into account debates about climate changes, the role of values in scientific practice, science policies, feminist and ecological critique, the interdisciplinarity and diversity of science considering the changes that occurred in the scientific practice and sciences themselves.

The book is structured around six chapters and can be divided into two parts. The first three chapters focus on content of the sciences while the last three consider the contexts in which scientific work is done. Chapter 1 gives a good overview of the connection between science and philosophy introducing the relevant questions through climate change debates and disputes concerning racial differences. The examples force us to ask philosophical questions. What is the evidence and what does it entitle people to believe? Who has the authority to make scientific judgments? How should we decide about future science policies? Questions like these lead to more general concerns about whether *natural sciences are the uniquely best sources of human knowledge, setting standards that ought to be achieved in all fields of inquiry* (p. 3). A brief history of science gives us further insight into how philosophers and scientist addressed these questions through notions of causality or development of new quantitative experimental methods that established professional science. At the end of the chapter authors introduce three *images of science* that dominated historical, sociological and philosophical conceptions of science. The most enduring is the traditional *image* that views *science as a reliable means for accumulating useful knowledge* with observation as the main scientific method.

Therefore, the following chapter describes the analytic project in philosophy of science. Authors introduce the problem of demarcating science which resulted in attempts to reduce science to analytical truths with respect to scientific explanation, confirmation and structure of scientific theories. In a critical discussion about historical attempts to discover the scientific method through contexts of discovery and justification, Kitcher and Barker provide a few interesting examples from scientific practice such as *The Discovery of Insulin*, *The Biology of Race* or *Philosophical Sources for the Analytic Project*. Apart from displaying a brief historical overview it seems that the intent of the chapter is framing the problem of viewing science as a value free-zone and demonstrating the limitations of such an approach to the scientific practice.

This line of criticism extends to the third chapter where authors explore questions that where neglected in the analytical debates about science, namely the diversity of the sciences. Through various theories integrated under the *ideal of unified science* such as naturalism and reductionism Kitcher and Barker present the most prominent attempts to define scientific methodology. As in previous chapter authors insist on the modest form of methodological naturalism emphasizing the importance of looking into the scientific practice and argue for pluralistic approach to the scientific methodology.

The remaining three chapters form the core of the textbook and its primary intention which is to introduce philosophical picture of science as a social enterprise. In order to do that, authors start with notions of success, truth and progress within historical reference to Kuhn's revolutionary account of science. The main epistemological and methodological problems – theory leadenness of observation and incommensurability thesis, illustrated with two famous examples *Tycho and Kepler Observe the Dawn* and *The Devonian Compromise* – display relativistic concerns about scientific progress. In order to preserve pluralistic view of scientific knowledge against the *authoritative perspective* authors accept the *natural reasonableness thesis: in different areas and in different places, people are equipped with the same cognitive faculties, and those faculties are put to work in similar ways* (p. 91). Since this does not solve the problem of relativism about scientific progress authors defend a special kind of pluralism which is compatible with realism. They suggest two ways of dealing with this problem, first, *divorcing the idea of progress from claims about truth* and second, abandoning the thesis that there is *a view from outside in which we can compare the world and our representations of it*. According to the first idea we should recognize the fact that people from different societies have different problems, depending on significance they pose for a particular culture, which they try to resolve by developing theories or models that differ from the ones accepted in our society, since we are focused on different problems. Therefore, scientific activity is determined by its pragmatic purposes leaving relativistic worries about scientific truth outside the debate. The second line of reasoning is also concerned with purposes of scientific activity and the main suggestion is that instead of looking from the outside *we can observe one another*. In this way Barker and Kitcher preserved both realism with the attractive but problematic thesis of *natural reasonableness* suggesting that, when it comes to success, our ordinary considerations provide enough evidence that our representations of the independent world are accurate, and that scientific progress consists in getting closer and closer to a true account of the parts of the world we consider important.

This approach to scientific *practice gives basis for developing an image of science as a social endeavor*. The argument is developed from the critical challenges from feminist, cultural and ecological critique which all have in common the problem of insufficient diversity within scientific practice. Whether we are concerned with predominance of particular social or gender group or by cultural and ecological uniqueness of specific parts of the world, there are damaging parts of scientific practice as well as unfair outcomes of a particular research. We can learn about some interesting cases from history of science, such as substantive developments in Primatology or Genome research which occurred when women entered research. Authors examine issues that arise from oversimplification of analytical research which is illustrated with an example from the history of modern Agriculture. The main suggestion is that *recognizing the social environment of researchers should not inspire the conclusion that attention to the evidence is inevitably overwhelmed by baser urges* (p. 125). In somewhat condense passage authors elaborate familiar themes and problems from social philosophy of science like the problem of scientific consensus, cognitive variation, distribution of scientific labor or the reward and recognition system in science. Those familiar with Kitcher's previous work can recognize his contribution in claims that we should abandon the myth of pure science (scientist) and assert that *extra-scientific motivations can promote good community strategies* (p. 128) or that the knowledge we have is dependent on past social decisions. All these claims support the main agenda of the textbook which is to provide a philosophical perspective of science that is not and should not be value-free.

The last chapter focuses on framing the problem of values within philosophy of science debate with respect to contemporary scientific controversies. In order to elaborate their view authors start by stating that the aim of science is to provide us with true answers to significant questions. The view is developed by integrating two traditional answers, first is that scientific research aims at explanation, prediction and control and the second, which states that science aims at truth. The definition according to which *significant truths are those that enhance our understanding or that enable us to predict events or to intervene in nature* (p. 137) opens debate for normative issues about whose goals ought to be considered in respect to significant questions. Kitcher and Barker want to develop an account of scientific inquiry that goes against reductivist strategies which according to them *neglected the complex causal interconnections*. What they have in mind specifically is the fact that science often excludes certain truths or values and the reason for that is because it excludes *certain kind of people*. From this, authors develop an argument for an account of science that should not try to be value-free and construct basis for their own picture of scientific research which debunks the autonomy of sciences suggesting an ideal of well-ordered science. The first problem addresses the question of determining what kind of autonomy is possible. Specifically the direction of scientific research. When we observe relationship between scientists, larger society and contemporary scientific practice it seems that (whether we are inclined toward governmental or market control view) *today's science is far from autonomous*. Therefore, authors suggest an ideal as an attempt to approach the notion of significance in order to make way for determining the aims of sciences. They suggest Kitcher's concept of well-ordered science. Scientific research must involve a concept of *ideal deliberation* which is defined as *a discussion among representatives of the different predicaments and perspectives found in the inclusive human population (our entire species, past, present and future). Those representatives are required to readjust the wishes with which they come to the discussion, by taking account of the best available information about nature and about the prospects for research of different kinds, and by recognizing the equal worth of their fellow discussants and of their perspectives and preferences. (...) They endeavor to reach consensus on how research should be directed (what it should aim at), how it should be conducted, what standards should be used in adopting potential new items of knowledge, and what uses might be made of the knowledge that research delivers* (p. 151). There are further conditions that have to be met in order to achieve well-ordered science which introduces democratic values and procedures in scientific practice. The concept is of course an ideal to which we should strive since alternatives are, according to Kitcher and Barker, far from acceptable and sketched by examples like *Climate Change Controversies* or *Disease Research and Global Health*. The examples are simple illustrations of not only what happens in contemporary science when we depart from the proposed ideal but also of how complicated and interdisciplinary are today's research and science policies. Therefore, in a field that is rapidly evolving Kitcher and Barker suggest we keep up. Image of philosophy of science developed in the textbook frames familiar problems in more expansive way taking into account current scientific practice and controversies. From that point we can move beyond traditional debates in an effort to develop epistemological and metaphysical accounts of science that is not value-free. Furthermore, authors suggest that scientific research should aim toward an ideal of a well-ordered practice which is defined as a social endeavor.

It is important to mention that every chapter in the book contains an extensive further reading lists and numerous illustrative examples from science which portrait very well some of the problems the book addresses. However, it is not always clear who are the intended readers since some parts of the book clearly demand background knowledge rendering the book a bit ambiguous for uninformed reader but there are also parts of the text where claims are put forward in somewhat condense and theoretically oversimplified manner. For example, one entire chapter is concentrated on Kuhn's revolutionary account of science. Parts of his theory as its critics are elaborated in detail as is the case with the analytic project in philosophy of science. Both are of course important and necessary for introducing the picture of science as a social endeavor. However, authors completely neglected philosophers like R. Merton, R. Hull, H. Longino, L. Laudan or I. Lakatos. If the book was intended for students then sometimes it lacks clarity and completeness in sketching historical and theoretical background of a particular problem. However, for the ones teaching philosophy of science it can provide a useful overview of the current state of the field and of contemporary scientific practice with some new perspectives on philosophy of science debates. Therefore, the textbook reaches its agenda in providing a more expansive and novel approach to teaching philosophy of science. Regardless of a few shortcomings, *Philosophy of Science: A New Introduction* is a valuable contribution to textbooks in philosophy of science. It frames the most important problems in philosophy of science in contemporary context taking into account real scientific practice. Aside from that, it introduces many issues that are open for debate and contemplation which makes it both instructive and useful for students and professionals in the field.

SMILJA CUKROV
University of Rijeka, Rijeka, Croatia