*Relevance and Prosody**

DEIRDRE WILSON AND TIM WHARTON

Abstract

Prosody provides both 'natural' and properly linguistic input to utterance comprehension. It contributes not only to overt communication but to covert or accidental forms of information transmission. Its function is typically to convey emotions or attitudes or to alter the salience of available interpretations. How should these aspects of communication be described and explained? This paper takes a relevance-theoretic approach, focusing on four main issues: (a) how should the communication of emotions or attitudes be analysed? (b) how do 'natural' prosodic elements contribute to communication? (c) what does prosody encode? (d) what light can prosody shed on the place of pragmatics in the architecture of the mind?

1 Introduction

Commentators on the effects of prosody on comprehension are broadly agreed on three main points. First, prosodic inputs to the comprehension process range from the 'natural' (e.g. an angry, friendly or agitated tone of voice) to the properly linguistic (e.g. lexical stress or lexical tone). Second, the effects of prosody are highly context-dependent: prosodic information interacts with information from many other sources during the comprehension process, and the same prosodic input may have different effects on different occasions. Third, prosody typically creates impressions, conveys information about emotions or attitudes, or alters the salience of linguistically-possible interpretations rather than expressing full propositions or concepts in its own right; it is sometimes described as forming the 'packaging' rather than the 'content' of the message (cf. House 1990, forthcoming).

A point less often noted in the literature is that the effects of prosody may be either accidental or intentional, and if intentional, either covertly or overtly so. For instance, a speaker's tone of voice may create an impression of boredom or impatience without her being aware of it. Knowing this, a sophisticated speaker may covertly manipulate her tone of voice to create an apparently accidental

* A version of this paper will appear in *Journal of Pragmatics*. We would like to thank Diane Blakemore, Coralie Chevallier, Marjorie Lorch, Neil Smith, Dan Sperber, John Wells, Anne Wichmann, Yi Xu, the participants in the Prosody and Pragmatics Conference and two anonymous referees for valuable discussion and comments.

impression that she would not want to acknowledge as part of her meaning in the full Gricean sense (Grice 1957, 1969, 1982). Grice distinguishes covert and accidental forms of information transmission from *overt communication,* or *speaker's meaning*, where the speaker not only intends to convey a certain message but wants the audience to recognise this intention, and would acknowledge it if asked. Whether some information is accidentally, covertly or overtly conveyed makes a difference both to pragmatics and to social interaction more generally: the effects of accidentally betraying or covertly revealing one's anger, nervousness or triumph, for instance, are different from those of overtly showing it (Sperber and Wilson 1986/95: chapter 1, section 12).

These diverse effects of prosody present a challenge to existing theories of communication: how should they be described and explained? Here, we will consider how the wealth of insights provided by the literature on the interpretation of prosody might be analysed from a relevance-theoretic perspective by addressing four main issues:

(a) How can both 'natural' and properly linguistic prosodic inputs contribute to utterance interpretation?
(b) What light might relevance theory shed on the non-propositional effects of prosody, whether accidentally, covertly or overtly conveyed?
(c) What (if anything) do prosodic inputs encode?
(d) Where do prosodic effects fit into the architecture of the mind?

In Section 2, we argue that 'natural' prosodic inputs fall into two importantly different classes, *natural signs* and *natural signals*, neither of which is intrinsically linked to overt communication (Wharton 2003b). In Section 3, we show how relevance theory might help to account for the non-propositional effects of prosody, and analyse the contribution of prosody to accidental, covert and overt information transmission. In Section 4, we argue that both 'natural' and properly linguistic prosodic signals may achieve their effects by encoding procedural rather than conceptual information (Blakemore 1987, 2002; Fretheim 2002; Wharton 2003a). In Section 5, we consider the relation between prosody, pragmatics and mindreading in the light of recent debates on pragmatics and modularity (Sperber 2000a,b; Bloom 2002; Wilson and Sperber 2002).[1]

[1] For relevance-theoretic approaches to prosody, see e.g. Sperber and Wilson 1986/95: chapter 4, section 5; Vandepitte 1989; Clark and Lindsey 1990; House 1990, forthcoming; Escandell-Vidal 1998, 2002; Imai 1998; Fretheim 2002.

2 Prosody and the 'showing-meaning$_{NN}$' continuum

Prosodic inputs are sometimes described as ranging along a continuum from 'more to less linguistic', or from 'natural' to language-specific (Gussenhoven 2002; Pell 2002a). The intuition behind these descriptions is that some prosodic inputs are interpreted by specifically linguistic codes, while others are interpreted by non-linguistic systems which are in some sense 'natural' and universal (Ladd 1996; Gussenhoven 2004). Sounding surprised or angry, for example, is seen as a 'natural' prosodic phenomenon which is not interpreted by the grammar, while certain aspects of postlexical intonation, along with lexical stress assignments (e.g. the noun *pérmit* versus the verb *permít*), are seen as properly linguistic. Similar continua between 'natural' display and language-like encoding have been proposed for gesture (e.g. Kendon 1988) and vocalisation generally (Goffman 1981).

How do prosodic inputs achieve their effects? In much of the literature, all prosodic inputs, whether 'natural' or linguistic, are more or less explicitly treated as *signals*, interpreted by linguistic or non-linguistic *codes*.[2] In this section, we will argue that 'natural' prosodic inputs fall into two importantly different categories – natural signs and natural signals – which are worth distinguishing both from each other and from properly linguistic inputs. Natural signals, like linguistic signals, are genuinely coded and inherently communicative; natural *signs*, by contrast, are interpreted by inference rather than decoding, and are not inherently communicative at all. Prosodic inputs of all three types may be exploited in overt communication. We will argue that when exploited in this way, they range along a continuum from *showing* to *meaning*$_{NN}$ which differs from the prosodic continua referred to above in two main respects: first, it includes only inputs exploited in overt communication; and second, it includes not only natural and linguistic signals but also overtly used natural signs (cf. Wharton 2003a,b).

2.1 Signs and signals

Hauser (1996), studying animal communication, applies a traditional distinction between signs and signals to cases of information transmission among animals. *Signs* carry information by providing evidence for it; *signals* carry information by encoding it. The main distinction between signs and signals is one of *function* (in the sense of Millikan 1984; Sperber 2000a, forthcoming a; Origgi and Sperber 2000). Although a sign may happen to carry information for an observer, this is not

[2] A code is a special-purpose system which pairs signals with messages, enabling communication to be achieved by encoding a message into a signal, which is transmitted via a channel and decoded by a receiver with an identical copy of the code (Sperber and Wilson 1986/95: chapter 1, section 1).

its function: it would go on being produced whether or not it carried this information. Hauser (1996: 9-10) provides the following example. While chimpanzee nests provide evidence for forest monkeys of the presence of chimpanzees, this is not their function: chimpanzees would go on building nests whether or not there were any forest monkeys around. One way of describing this situation is to say that natural signs (e.g. chimpanzee nests) are not *inherently* communicative.

Signals, by contrast, are inherently communicative. Their function is to carry information for others; if they did not carry this information, it would be hard to explain why they go on being produced. For instance, the function of the honeybee's dance is to inform other honeybees about the location of nectar; the function of the vervet monkey's alarm call is to inform other vervet monkeys about the presence of predators (von Frisch 1967; Cheney and Seyfarth 1990, 1999; Hauser 1996; Hauser and Konishi 1999; Michelson 1999). If they did not carry this information, it would be hard to see why these behaviours survive. Most animal communication seems to be based on signalling systems of this type. In many cases, the system is so complex that it is hard to see it as governed by anything but an innately determined code.

Coded communication is found in animals with no capacity for inferential intention recognition. Honeybees lack the ability to infer the intentions of others, but they can still inform each other about the location of nectar by means of their dance-based code. Overt communication, by contrast, requires a capacity for inferential intention recognition, and may be achieved in the absence of a code – as when I hold up my half-full glass to show you that I'm not yet ready for another drink. According to Sperber and Wilson (1986/95: chapter 4, section 1), what is unique to human linguistic communication is that it involves *both* coding *and* inferential intention recognition: the speaker produces linguistically coded evidence of her intention to convey a certain meaning. In the domain of human information transmission, it should therefore be possible to find (at least) three distinct types of input: natural signs (which are interpreted purely inferentially); natural signals (which in some cases are interpreted purely by decoding); and linguistic signals (which are interpreted by a combination of decoding and inference). We will suggest that all three types of prosodic input exist.

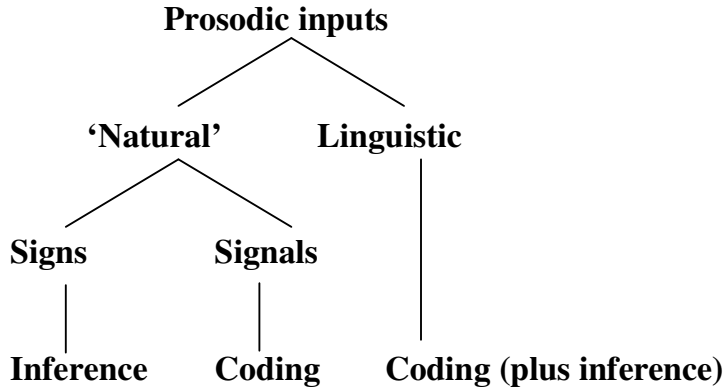
Wharton (2003b) illustrates the distinction between human natural signs and signals by comparing shivering with smiling. Shivering is a natural behaviour whose function is to generate heat by rapid muscle movement. It may provide evidence (for an observer with the appropriate experience or background knowledge) that the individual is feeling cold. However, its function is not to carry this information: it is not a signal but a sign. Smiling, by contrast, appears to have evolved as a signalling activity whose function *is* to convey information to others (van Hooff 1972; Ekman 1989, 1992, 1999; Fridlund 1994). As Ekman (1999: 51)

puts it, smiling and other spontaneous facial expressions "have been selected and refined over the course of evolution for their role in social communication". Like the bee dance, they are signals rather than signs.[3]

It is easy to think of prosodic counterparts to shivering and smiling. For instance, a speaker's mental or physical state may affect the prosodic properties of her utterance, enabling a hearer with the appropriate experience or background knowledge to infer whether she is drunk or sober, sick or healthy, tired or alert, hesitant or assured. As with shivering, these prosodic properties carry information about the speaker's mental or physical state, but it is not their function to do so: they are natural signs, interpreted by inference rather than decoding. On the other hand, affective tones of voice, like affective facial expressions, may well be natural signals, interpreted by innately determined codes. We will explore this idea further in Sections 4 and 5.

If the arguments of this sub-section are right, properly linguistic prosody falls into a still further category. It is part of a *linguistic* signalling system, governed by a *linguistic* code with its own special-purpose principles or mechanisms, whose function is to provide evidence for use in overt intentional communication. This position is summed up in *fig. 1*, where natural prosodic signs are interpreted purely inferentially, natural prosodic signals by decoding, and linguistic signals by a combination of decoding and inference.[4]

Figure 1



[3] The fact that neonates appear able to distinguish basic facial expressions of emotion provides some support for the view that their interpretation is governed by innately-determined codes (cf. Field *et al*. 1982; Phillips *et al*. 1990; Nelson and de Haan 1996).

[4] We will argue in Section 4 that what is encoded by human natural prosodic signals (like what is encoded by human linguistic signals) may be underdeterminate and require some inferential enrichment.

In the next sub-section, we will consider how prosodic inputs of all three types may be used in overt communication.

2.2 Signs, signals and the showing-meaning$_{NN}$ continuum

As noted above, natural signs and natural signals may convey information without being overtly intended to do so. A speaker's tiredness, boredom, frustration or anger may be revealed by her tone of voice or facial expression, even though she is trying to conceal them and even though it is clear to the audience that they are being accidentally revealed rather than intentionally conveyed. In more sophisticated cases, a speaker may covertly manipulate her tone of voice to suggest to an audience that she is accidentally betraying her feelings rather than wanting them to be recognised as part of her meaning in the full Gricean sense.

However, a communicator may also openly *show* her feelings to an audience. She may do this by deliberately producing, and perhaps exaggerating, a natural sign or signal (e.g. a shiver, a frown, an angry tone of voice); or she may do it by making no attempt to conceal a spontaneously-produced natural sign or signal in circumstances where it is obvious to both communicator and audience that she could have taken steps to conceal them. Grice saw an important difference between these two types of case. He was prepared to treat the deliberate simulation of a piece of natural behaviour (e.g. a frown or shiver) as a case of *non-natural meaning*, or *meaning$_{NN}$*, whose interpretation he saw as crucially involving a process of inferential intention recognition. However, he argued that a spontaneous piece of natural behaviour, even if openly *shown* to an audience, did not amount to a case of *non-natural meaning* (Grice 1989: 219). This distinction, if correct, would have important consequences for pragmatics. In Grice's framework, a full-fledged speaker's meaning is a type of non-natural meaning (or meaning$_{NN}$), and to deny that the open showing of spontaneously-produced natural behaviours is a case of non-natural meaning would be to exclude it from the domain of pragmatics.[5] Yet there seem to be clear cases where the open showing of spontaneously produced natural signs and signals makes a difference to the speaker's meaning.

Suppose that Lily utters (1), making no attempt to conceal the spontaneous anger in her facial expression and tone of voice:

---

[5] For a variety of reasons unrelated to the analysis of deliberately shown natural behaviours being proposed here, some philosophers have questioned Grice's distinction between showing and meaning$_{NN}$. Schiffer (1972: 56), for example, argues that cases of overt showing do amount to cases of meaning$_{NN}$; Recanati (1987: 189) argues that while the distinction is well-motivated, we should not necessarily limit 'Gricean communication' to cases of meaning$_{NN}$.

(1) Jack is late.

She would naturally be understood as communicating not only that Jack was late but that she was angry that he was late. Grice's framework appears to exclude such spontaneous expressions of anger from contributing to a speaker's meaning in the sense he was trying to define. Relevance theorists, by contrast, have consistently argued that there is a continuum of cases between showing and meaning$_{NN}$, all of which may fall within the domain of pragmatics and contribute to a speaker's meaning (Sperber and Wilson 1986/95: chapter 1, section 10). In the rest of this paper, we will use the relevance-theoretic term *ostensive-inferential communication* to describe this dual process of showing and inferring the communicator's meaning. A detailed defence of this approach and comparison with Grice's account is developed in Wharton (2003a,b), who argues that the open showing of spontaneously-produced natural signs and signals may be located along the showing-meaning$_{NN}$ continuum in *fig. 2*:

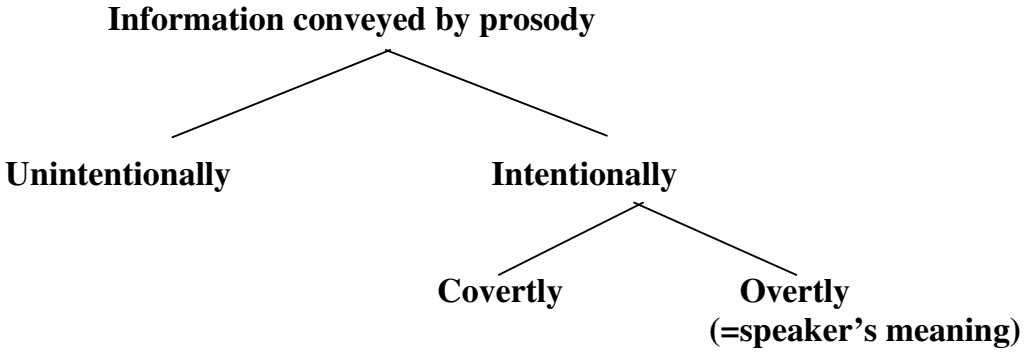
Figure 2

Overt communication and the showing-meaning$_{NN}$ continuum

Showing Natural signs Natural signals Linguistic signals **Meaning$_{NN}$**

In this section, we have argued that prosodic inputs fall into three broad types: natural signs, natural signals and linguistic signals. Natural signs are not inherently communicative, and convey information by providing evidence for it. Natural signals are inherently communicative, and convey information by coding. Both natural signs and natural signals may be used in ostensive-inferential communication, but may also exist without it. Linguistic signals, like natural signals, are inherently communicative and convey information by coding. Unlike natural signs and natural signals, they are primarily geared to use in ostensive-inferential communication, so that the output of linguistic decoding (which is typically quite fragmentary and incomplete) provides input to a further process of inferential intention recognition. The position is summed up in *fig. 3*:

Figure 3



In Section 5, we will suggest a range of possible test cases based on these distinctions which might shed further light on the place of prosody in the architecture of the mind. In the next section, we will consider what light relevance theory might shed on the non-propositional effects of prosody, whether accidentally, covertly or overtly conveyed.

3 Relevance theory and the analysis of prosody

Relevance theory is based on a definition of relevance and two general principles: a Cognitive and a Communicative Principle of Relevance (for recent accounts, see Blakemore 2002; Carston 2002; Wilson and Sperber 2004). *Relevance* is characterised in cost-benefit terms, as a property of inputs to cognitive processes, the benefits being *positive cognitive effects*, and the cost the *processing effort* needed to achieve these effects. Other things being equal, the greater the positive cognitive effects achieved by processing an input in a context of available assumptions, and the smaller the processing effort required, the greater the relevance of the input to the individual who processes it.

Although mainly concerned with ostensive-inferential communication, relevance theory has implications for the role of natural signs and signals in other forms of information transmission. In this section, we will consider how the contribution of prosody to accidental, covert and overt information transmission might be described within this framework, and propose an analysis of non-propositional effects. We will end by distinguishing *strong* from *weak* communication, and show how the ostensive use of natural signs and signals may create a wide array of weakly communicated effects.

3.1 Relevance, cognition and non-propositional effects

According to the First, or Cognitive, Principle of Relevance, human cognition tends to be geared to the maximisation of relevance. As a result of constant selection pressure towards increasing cognitive efficiency, humans have developed automatic heuristics or procedures for picking out potentially relevant inputs and processing them in the most productive way (cf. Sperber 1994; Sperber and Wilson 2002; Sperber forthcoming a). Prominent among these automatic procedures are *mind-reading* mechanisms for attributing mental states to others in order to explain and predict their behaviour.

An example of an automatic mind-reading mechanism is the Eye Direction Detector described by Baron-Cohen (1995), which infers what someone is seeing or watching from the direction of their gaze. In the terms of Section 2, gaze direction is a natural sign which merely happens to carry information for an observer. However, the potential relevance of this information is such that a relevance-oriented cognitive system might well be improved by the development of a special-purpose inferential mechanism of this type. The natural codes discussed in Section 2, which interpret affective facial expressions or tones of voice in terms of underlying mental states, might be seen as examples of automatic mind-reading mechanisms of a coded rather than an inferential nature, dedicated to the interpretation of natural signals rather than natural signs. What distinguishes a special-purpose inferential mechanism from a coding mechanism is, first, that the inferential mechanism applies to signs rather than signals, second, that it is genuinely inferential (i.e. it draws warranted conclusions on the basis of evidence), and third, that it is not part of a signalling system with corresponding encoding mechanisms at the production end.

What is the output of these automatic mind-reading mechanisms? What type of information do they convey? According to Ekman,

> There is no evidence about precisely what type of information is conveyed when, during an on-going social interaction, one person sees a facial expression of emotion on another person's face (Ekman 1989: 159).

The same might be said of the information conveyed by affective tones of voice. Lieberman (2000) notes that prosodic inputs often create a *diffuse impression* of friendliness, condescension, confidence, nervousness, etc. which may never surface

to consciousness. How are these non-propositional effects of prosody to be described and explained?[6]

According to relevance theory, an input to cognitive processes is relevant when its processing in a context of available assumptions yields *positive cognitive effects* (e.g. true contextual implications, warranted strengthenings or revisions of existing assumptions). Within this framework, different inputs (or the same input at different times) may achieve relevance in different ways. Thus, the sound of a gunshot provides clear, strong evidence that someone has fired a gun, and this conclusion would normally be salient enough, and relevant enough, to pre-empt the attention of anyone within hearing range. By contrast, a faint sigh from a friend working across the room would normally provide weak support for a wide array of possible conclusions, no one of which is likely to be salient or relevant enough to attract the observer's undivided attention, but which, taken together, give a certain insight into the friend's state of mind, and may achieve relevance thereby. We would like to suggest that the diffuse impressions created by affective facial expressions and tones of voice may be analysed along these lines, as involving marginal alterations in the strength or salience of a wide array of conclusions rather than providing strong support for a single, determinate conclusion (Sperber and Wilson 1986/95: chapter 1, section 11; Wharton 2003a,b).

The fact that human cognition is relevance-oriented makes it possible, at least to some extent, for one individual to predict what inputs in the environment another individual is likely to attend to, what contextual information he is likely to use in processing it, and what conclusions he is likely to draw. This predictability may be exploited in covert information transmission: thus, a criminal may fire a gun in the air in order to distract the security guards from a break-in elsewhere; Lily may sigh faintly in a library, expecting her friend to notice and suggest a coffee break. This predictability is also exploited in ostensive-inferential communication, where, as we will argue in the next sub-section, it may have been supplemented by a further automatic procedure specifically geared to inferring a communicator's meaning.

3.2 Relevance and ostensive communication

According to the Second, or Communicative, Principle of Relevance, an utterance or other ostensive act creates a presumption of relevance not created by ordinary inputs.[7] While ordinary inputs (sights, sounds, memories, conclusions) carry no

---

[6] For careful discussion of the wide variety of affective information that may be conveyed by prosodic inputs, see Wichmann 2002.

[7] An ostensive act involves the production of an *ostensive stimulus* designed to attract the addressee's attention and focus it on the communicator's meaning. Clapping one's hands, clearing one's throat, catching the addressee's eye or touching him on the arm are examples of ostensive

particular guarantee of relevance, the addressee of an ostensive act is entitled to presume that it will be relevant enough to be worth his attention, and to look for an interpretation on which it satisfies this presumption. In a relevance-oriented cognitive system, we might therefore expect the task of identifying a communicator's meaning to be facilitated by the development of an automatic comprehension procedure, which would yield the desired results for less effort. In recent work, relevance theorists have been exploring the idea that the following procedure is automatically applied in the on-line processing of attended verbal inputs to construct a hypothesis about the speaker's meaning (Sperber, Cara and Girotto 1995; Sperber 2000; Sperber and Wilson 2002; Wilson and Sperber 2002):

Relevance-theoretic comprehension procedure

(a) Follow a path of least effort in computing cognitive effects. Consider interpretations (e.g. disambiguations, reference resolutions, contextual assumptions, implicatures) in order of accessibility.

(b) Stop when your expectation of relevance is satisfied.

A hearer using this procedure in interpreting an utterance should (a) pay attention to *perceptually salient* aspects of the input; (b) consider the most *accessible* disambiguations, reference resolutions, contextual assumptions, implicatures, speech-act descriptions, etc.; (c) assume that any extra *processing effort* demanded will be offset by *extra or different cognitive effects*, and (d) stop when he has an interpretation that yields enough cognitive effects to satisfy the particular expectation of relevance raised by the utterance. Here are a few brief illustrations of how different types of prosodic input might contribute to the comprehension process along these lines.

It is often noted that one of the functions of prosody is to guide the utterance interpretation process by altering the salience of possible disambiguations, reference resolutions, contextual assumptions, implicatures, speech-act descriptions, etc. What relevance theory adds to this intuitive description is the idea that the salience of interpretations can be affected not only by altering processing effort but also by manipulating the relationship between processing effort and expected cognitive effects. Thus, suppose Lily utters (2) in a neutral tone:

(2) I'm disappointed.

There are many degrees and shades of disappointment that she might have intended to convey, each of which would yield different implications and be relevant in a

---

stimuli; other, subtler indicators of an intention to perform an ostensive act have been insightfully discussed in the literature on turn-taking.

different way. While neutral (or 'expected') prosody would cause the hearer least phonological processing effort, it would give him little guidance on the type of cognitive effects he was expected to derive. By contrast, any departure from neutral (or 'expected') prosody would increase the hearer's phonological processing effort, but would thereby encourage him to look for extra (or different) effects. Which effects should he derive? According to the relevance-theoretic comprehension procedure, he should follow a path of least effort, deriving whatever effects are made most accessible in the circumstances by the type of prosodic input used, and stopping when he has enough effects to justify the extra effort caused by the departure from neutral (or 'expected') prosody. Thus, the utterance of (2) in an angry tone of voice, with a wide pitch range and increased stress on 'disappointed', should indicate a degree and type of disappointment that would warrant the derivation of a particular range of positive cognitive effects via the automatic working of the relevance-theoretic comprehension procedure.

Another idea often found in the literature is that contrastive stress, like pointing, is a natural highlighting device, used to draw attention to a particular constituent in an utterance. This idea is explored from a relevance-theoretic perspective in Sperber and Wilson (1986/95: chapter 4, section 5). Here is a brief illustration of how this approach might work. It follows from the Communicative Principle of Relevance that if two stress patterns differ in the amounts of processing effort required, the costlier pattern should be used more sparingly, and only in order to create extra, or different, effects. Thus, compare the effects on reference assignment of the neutral stress pattern in (3) and the costlier contrastive pattern in (4):

(3) Federer played Henman and he *béat* him.
(4) Federer played Henman and *hé* beat *hím*.

A hearer using the relevance-theoretic comprehension procedure in interpreting the second conjunct in (3) should follow a path of least effort in assigning reference, and interpret *he* as referring to Federer and *him* to Henman (an assignment made easily accessible by syntactic parallelism, on the one hand, and encyclopaedic knowledge, on the other). Use of the costlier contrastive pattern in (4) should divert him from this otherwise preferred interpretation towards the alternative, less accessible interpretation on which *he* refers to Henman and *him* to Federer. On this account, contrastive stress is a 'natural' highlighting device which achieves its effects via the automatic working of the relevance-theoretic comprehension procedure.[8]

[8] In example (3), both syntactic parallelism and encyclopaedic knowledge facilitate the interpretation of 'he' as referring to Federer. As Neil Smith has reminded us, syntactic parallelism

A possible objection to this 'natural highlighting' account is that the acceptability of contrastive stress patterns seems to vary across languages (for an excellent survey of cross-linguistic variations and objections to 'highlighting' accounts of both contrastive and focal stress, see Ladd 1996: chapter 5). However, as noted in Sperber and Wilson (1986/95: 213-4), this is not a particularly compelling objection unless it can be shown that variations in contrastive stress are not explainable in terms of processing effort. For instance, French has a relatively flat intonation contour and a strongly preferred final placement of focal stress, whereas English has a relatively variable intonation contour and freer placement of focal stress. We might therefore expect the use of non-final contrastive stress in French to be more disruptive, hence costlier in terms of processing effort, and the use of alternative syntactic means (e.g. clefting) to be preferred. In the next section, we will consider another possible response to the claim that cross-linguistic variation in contrastive stress automatically shows the need for language-specific rules.[9]

Natural prosodic signs and signals of the type discussed in Section 2 may also contribute to ostensive-inferential communication via the automatic working of the comprehension procedure. Thus, consider what might be communicated by Lily's ostensive sigh in (5b) or (6):

(5) a. *Jack*: How are you enjoying your geography course?
b. *Lily*: [Looks at Jack and sighs]

(6) Lily comes home after a day at work, slams the door, catches Jack's eye and sighs.

In (5b), Lily's sigh provides strong evidence for a definite conclusion (she is not enjoying her geography course) and weaker support for a range of further conclusions (her geography course is difficult, she is worried or anxious about it, is in need of help or sympathy, etc.). In (6), her sigh does not provide evidence for a

---

and encyclopaedic knowledge do not always point in the same direction. Thus, in 'John telephoned Bill and he refused to *spéak* to him', it may be manifest to both speaker and hearer on the basis of encyclopaedic knowledge that, despite the syntactic parallelism, the most obvious candidate referent for 'he' is the direct object 'Bill' rather than the subject 'John'. We claim that in this case, use of a contrastive stress pattern should divert the hearer towards the otherwise less obvious interpretation on which John refused to speak to Bill. More generally, what counts as the 'path of least effort' is determined by a variety of linguistic and non-linguistic factors. This point will be discussed further below.

[9] Notice that we are making a relatively limited point. We are not denying that there are language-specific prosodic rules, but simply pointing out that a well-developed pragmatic theory can provide more resources (on both the effect and the effort side) for analysing cross-linguistic variation than standard appeals to 'relative newsworthiness' or 'relative semantic weight', and that this may have implications for the debate about 'universal' versus 'language-specific' prosody. (See also footnotes 12 and 13.)

single, definite conclusion, but creates a diffuse impression of the type discussed in Section 3.1, by marginally altering the strength or salience of a wide array of conclusions. As these examples show, ostensive use of a natural prosodic sign or signal may convey a more or less definite meaning when addressed to a hearer with more or less definite expectations of relevance.

Relevance theory captures these differences by distinguishing *strong* from *weak* communication, and *strong* from *weak implicatures*. A conclusion is *strongly implicated* (or is a *strong implicature*) to the extent that it (or some closely similar proposition) must be derived in the course of constructing a satisfactory interpretation (i.e. one that satisfies the hearer's expectation of relevance). It is *weakly implicated* if its recovery helps with the construction of a satisfactory interpretation, but is not essential because the utterance provides evidence for a wide array of roughly similar conclusions, any of which would do (Sperber and Wilson 1986/95: chapter 1, sections 10-12, chapter 4, section 6; Wilson and Sperber 2002). Thus, Lily's sigh in (5b) quite strongly implicates that she is not enjoying her geography lectures, while her sigh in (6) conveys a wide array of weak implicatures but no strong implicatures: that is, it creates an impression rather than conveying a definite message. Typically, a spoken utterance involves a mixture of strong and weak communication, with affective prosody generally contributing to the weaker side. Relevance theory provides a framework in which this fact can be accommodated and explained.

In a series of interesting recent papers, Carlos Gussenhoven and his colleagues (e.g. Gussenhoven 2002; Chen and Gussenhoven 2003) have argued that the interpretation of prosody is governed by both biological and properly linguistic codes. An example of a biological code is the Effort Code (Gussenhoven 2003), which links the amount of energy expended in speech production to a range of interpretive effects. Thus, an increase in effort may lead to increased articulatory precision, creating an impression of 'helpfulness', or 'obligingness'; or it may result in a wider pitch range, creating an impression of 'forcefulness' or 'certainty' or conveying affective meanings such as 'agitation' or 'surprise'. This account covers some of the same ground as our suggestion above that variations in the pitch range of (2) ('I'm disappointed') might achieve their effects by 'natural' pragmatic means. The relation between Gussenhoven's biological codes and the notions of natural sign and natural signal discussed in Section 2 above deserves more careful consideration than we are able to give it here, and we hope to explore it in future work. As a preliminary step in this direction, we would like to draw attention to some possible points of comparison between the notion of effort used in relevance theory and the one appealed to in Gussenhoven's Effort code.

The most obvious difference between the two approaches is that the type of effort appealed to in Gussenhoven's Effort code is *speaker's effort*, whereas the type of effort appealed to in the relevance-theoretic account is *hearer's effort*. Although

both speaker's effort and hearer's effort affect the comprehension process, and both will ultimately need to be taken into account, they do not always vary in the same direction.[10] For instance, articulating clearly may cost the speaker some extra effort in production but is likely to diminish the hearer's overall effort in understanding (as the extra effort a writer puts into redrafting a text may save the reader some effort in comprehension). Is clear articulation a natural signal, interpreted (as Gussenhoven suggests) by an innately determined code? We would suggest that it might be better treated as a natural *sign* of the speaker's desire to help the speaker understand, which is interpreted via inference rather than decoding. Like other natural prosodic signs, it may be exploited in ostensive-inferential communication, as long as the fact that the speaker is making a special effort is salient enough, and relevant enough, to attract attention and be picked up by the relevance-theoretic comprehension procedure.[11]

Notice, now, that the hearer's overall processing effort is analysable into several components which may vary in opposite directions. For instance, the addition of an extra word or phrase may increase the hearer's linguistic processing effort but diminish the effort of memory and inference required to access a context and derive the intended cognitive effects, so that the utterance is less costly overall. By the same token, a departure from normal (or 'expected') pitch range may increase the hearer's phonological processing effort but reduce the effort of memory and inference required to arrive at the intended interpretation. Thus, the effort factor appealed to in our account of how pitch range affects the interpretation of (2) above is a special case of a more general factor that affects the pragmatic interpretation of every utterance. If this general account in terms of hearer's processing effort turned out to be descriptively adequate, it might be preferable on theoretical simplicity grounds to Gussenhoven's special-purpose account.[12]

---

[10] Speaker's effort is factored into the presumption of optimal relevance through a reference to the speaker's *abilities and preferences* (for discussion, see Sperber and Wilson 1995: section 3.3)

[11] As Dan Sperber has pointed out to us, there is also a third possibility: some forms of clear articulation (e.g. in certain traditions of theatrical speech, or stylised emphasis of final consonants in utterances such as the comedy catchphrase "I don'T believe iT") may be neither purely natural nor properly linguistic but cultural. See below for discussion.

[12] Gussenhoven and Chen (2003) show beautifully that the same pitch is interpreted as indicating different degrees of surprise by hearers from different languages. We do not think it follows (although of course it may be true) that such variations in pitch range have become grammaticalized, or properly linguistic. As suggested above, the same prosodic input may be more or less costly for hearers to process depending on what prosodic contours they are normally exposed to, and this may affect the comprehension process by 'natural' rather than coded means, via the automatic working of the comprehension procedure. Variations in phonological processing effort may therefore need to be taken into account in deciding between a 'natural' and a properly linguistic treatment.

In this section, we have tried to show how prosodic inputs may contribute to accidental, covert and overt information transmission by altering the salience of linguistically possible interpretations and/or creating non-propositional effects. We have suggested an automatic procedure for constructing a hypothesis about the speaker's meaning by following a path of least effort in looking for enough cognitive effects to satisfy the expectations of relevance raised by the utterance. The more salient the prosodic input, the more it will be expected to contribute to the speaker's meaning by achieving positive cognitive effects. A communicator who wants some prosodic feature of her utterance to be understood as contributing to her meaning should therefore do her best to make it salient enough, and rich enough in effects, to be picked out by the relevance-theoretic comprehension procedure and help make the utterance relevant in the expected way.

4 What does prosody encode?

In the last two sections we have made two apparently incompatible claims: first, that prosodic signals are naturally or linguistically coded, and second, that they often create a diffuse impression or communicate a wide array of weak implicatures rather than conveying a determinate message. A *code* is standardly seen as a set of rules or principles pairing signals with determinate messages. How is it possible to maintain both that prosodic signals are coded and that what they convey may be no more than a wide array of weak non-propositional effects?

In this section, we would like to pursue an idea proposed by Diane Blakemore (1987, 2002) and applied to different aspects of prosody by Vandepitte (1989), Clark and Lindsey (1990), House (1990, forthcoming), Escandell-Vidal (1998, 2002), Imai (1998), and Fretheim (2002) (see also König 1991). The idea is this: if linguistic communication typically involves a combination of decoding and inference, then linguistic signals might be expected to encode information of two distinct types. First, there is regular *conceptual* encoding, where a word (e.g. *boy*) encodes a concept (e.g. BOY) which figures as a constituent of the logical form of sentences in which that word occurs. Second, we might expect to find a form of *procedural* encoding, where a word (or other linguistic expression) encodes information specifically geared to guiding the hearer during the inferential phase of comprehension. The function of such 'procedural' expressions would be to facilitate the identification of the speaker's meaning by narrowing the search space for inferential comprehension, increasing the salience of some hypotheses and eliminating others, thus reducing the overall effort required. Following a suggestion of Dan Sperber (p.c.), such expressions might be described as encoding *meta-procedures,* which manage the accessibility or activation levels of the regular relevance-oriented procedures for perception, memory retrieval or inference, as

discussed in Section 3.1. We want to argue that both 'natural' and properly linguistic prosodic signals are procedural in this sense.

Properly linguistic expressions which have been analysed in procedural terms include discourse connectives, pronouns, mood indicators and discourse particles (cf. Blakemore 1987, 2002; König 1991; Wilson and Sperber 1993; Hall 2004). These are briefly illustrated below, with indications of their possible effects:

(a) Discourse connectives: e.g. *but* inhibits a conclusion that might otherwise be drawn.
(b) Pronouns: e.g. *she* facilitates the retrieval of female candidate referents.
(c) Mood indicators: e.g. imperative morphology facilitates the retrieval of a range of speech-act or propositional-attitude descriptions associated with imperatives.
(d) Discourse particles: e.g. *please* facilitates the retrieval of a range of speech-act or propositional-attitude descriptions associated with requests.

Properly linguistic prosodic signals (e.g. lexical stress, lexical tone and fully grammaticalized aspects of sentence stress and intonation) might be analysed on similar lines, as facilitating the retrieval of certain types of syntactic, semantic or conceptual representation. Thus, the notion of procedural encoding applies straightforwardly to properly linguistic prosodic elements.

There has been some debate about whether interjections such as *oh, ah* and *wow* are properly linguistic. Wharton (2003a) surveys the literature and concludes that interjections are best analysed as falling on the natural rather than the properly linguistic side. However, he also argues that interjections are natural signals rather than signs, and that they share with discourse connectives and discourse particles the property of encoding procedural rather than conceptual information. On this approach, the function of an interjection such as *wow* might be to facilitate the retrieval of a range of speech-act or propositional-attitude descriptions associated with expressions of surprise or delight, which might be narrowed in context by information derived from prosody, facial expressions, background assumptions, discourse context, etc., and contribute to the speaker's meaning in the regular way, by falling under the relevance-theoretic comprehension procedure.

The line of argument is taken further in Wharton 2003b, which proposes that natural signals such as smiles and other spontaneous facial expressions should also be analysed as encoding procedural rather than conceptual information. On this approach, the function of facial expressions of surprise or delight would be to facilitate the retrieval of similar propositional-attitude descriptions to those activated by the interjection *wow*. This approach makes it possible, on the one hand, to capture the fact that natural signals, interjections and properly linguistic signals such as mood indicators or discourse particles all have a coded element, and

on the other, to explain why what they communicate can sometimes be so nebulous, contextually shaded and hard to pin down in conceptual terms. It also makes it relatively easy to see how a given expression (e.g. an interjection) might move along the continuum from 'non-linguistic' to 'partly linguistic' to 'linguistic' without radically altering the type of information it conveys.

A spoken utterance is typically a composite of linguistic signals, natural signals and natural signs which interact in complex ways to yield a hypothesis about the speaker's meaning. Bolinger (who treats both facial expressions and intonation as signalling systems) describes their interaction in the following terms:

> If intonation is part of a gestural complex whose primitive and still surviving function is—however elaborated and refined—the signalling of emotions and their degrees of intensity, then there should be many obvious ways in which visible and audible gesture are coupled to produce similar and reinforcing effects. This kind of working parallel is easiest to demonstrate with exclamations. An *ah!* of surprise, with a high fall in pitch, is paralleled by a high fall on the part of the eyebrows… A similar coupling of pitch and head movement can be seen in the normal production of a conciliatory and acquiescent utterance such as "I will" with the accent at the lowest pitch—we call this a *bow* when it involves the head, but the intonation bows at the same time (1983: 98).

Dolan *et al.* (2001) provide experimental evidence of cross-modal priming between facial expression and emotional tone of voice: for instance, a facial expression of fear was more quickly identified when accompanied by a frightened tone of voice. We look at the implications of some of these experiments in more detail in the next section, where we consider how prosody fits into the architecture of the mind. Meanwhile, we note that affective tones of voice, like affective facial expressions, appear to be better analysed as natural signals rather than natural signs, and hence, on the approach developed in this section, as conveying information by procedural encoding rather than inference alone.

On the question of the precise relationship between linguistic and non-linguistic prosody, and where the line between them might be drawn, Bolinger inclines towards a largely 'natural' account:

> Intonation… assists grammar—in some instances may be indispensable to it—but it is not ultimately grammatical… If here and there it has entered the realm of the arbitrary, it has taken the precaution of blazing a trail back to where it came from (1983: 106-108).

However, as Ladd (1996), Gussenhoven (2002, 2004) and Wichmann (2002) suggest, there is a considerable cross-linguistic variation in the way these 'universal paralinguistic meanings' are realised, to a point where they may become heavily stereotyped or even fully grammaticalized and part of language proper. Our aim in this paper has not been to come down on one side or other of this debate, but merely to draw attention to two possibilities that have not been so widely considered in the literature: first, not all prosodic inputs are coded at all; and second, the fact that prosodic patterns and their interpretations become stereotyped or vary from language to language is not conclusive evidence that they are *linguistically* coded.

Commenting on an earlier version of this paper, Dan Sperber pointed out to us that some prosodic variation may be neither natural nor properly linguistic but cultural (Sperber 1996; Origgi and Sperber 2000; Sperber forthcoming a). Examples of cultural prosodic inputs might include the stylised intonation patterns or 'calling contours' discussed by Ladd (1978). To the extent that such inputs have a signalling function, they might be seen as falling into the category of what McNeill (1992: 32), discussing cultural signals such as the British two-fingered insult, calls *emblems*. Using evidence from Danish, Scheuer (1995: 446) suggests that 'culture-specific mechanisms' might also be at work in the stabilisation of a range of prosodic phenomena.[13] Just as emblems stabilise in a culture, so might certain prosodic patterns. In a fuller account, such cultural prosodic inputs would need to be factored into our figures 1-3 above.

Of course, in the case of prosody the task of teasing out these different distinctions is far from easy. At the end of a careful attempt to motivate the distinction between linguistic and paralinguistic intonation, Ladd (1996: 283) concludes: "But I concede that we must stop short of drawing a clear boundary between language and paralanguage. For now that question remains open." While leaving the question open, Ladd's fine-grained autosegmental analyses of intonational phonology shed considerable light on which parts of prosody are universal, and which are language-specific. We hope the distinctions drawn in this paper will make some contribution to this debate.

In this section, we have argued that both natural and properly linguistic prosodic signals might encode procedural information of a type shared by borderline

---

[13] Scheuer suggests that "in order to provide hypotheses about prosody in spoken Danish… it seems to be necessary to go… beyond the scope of universal pragmatics, i.e. pragmatics based on universal principles". While it is true that, so far, relevance theorists have been more concerned with psychological rather than sociological factors in communication, this has been more a matter of expedience than of principle. See Sperber and Wilson (1997) for discussion of possible interactions between cognitive and sociological factors; the theoretical notion of a non-natural, non-linguistic code may be useful in exploring these areas of interaction in more detail.

linguistic expressions such as interjections and properly linguistic expressions such as mood indicators, discourse connectives and discourse particles. This makes it possible to see how there could be a continuum of cases between purely natural prosodic signals and non-natural ones, whether cultural or properly linguistic. Studies of animal communication (e.g. Hauser 1996) suggest, moreover, that natural signals often evolve from natural signs, perhaps via an intermediate stage of stylisation, providing a continuum of cases along the way. While many empirical questions remain about where the borderlines between these different categories should be drawn, and how items move from one category to another, the theoretical position is fairly clear. Prosodic inputs come in several broad types, which convey information in different ways, all of which may be exploited in ostensive-inferential communication.

5 Prosody, pragmatics and mind-reading

There is evidence from the autobiographical writings of people with autism that they have particular problems with the interpretation of prosody. In her moving autobiography, Donna Williams reports:

> Speak to me through my words,' I asked Dr. Marek. I wanted to cut down the struggle in putting mental pictures into words. 'Can you take the dancing out of your voice and not pull faces so you don't distract me from what you're saying? (Williams 1994: 95).
>
> How do I stop people getting angry?' I asked. In effect, this meant how could I stop them from having any vocal variation whatsoever? I also wanted to know why they made faces and insisted on making their voices dance even though they could see it upset me. 'How do other people learn these things?', I wanted to know. 'They learn them naturally,' Dr Marek said (Williams 1994: 103).

Liane Holliday Willey, a writer with Asperger's syndrome, describes her problems with the interpretation of the adjective *disappointed* (discussed above in example (2)):

> If [my husband] were to tell me he was disappointed he had missed me at lunch, I would wonder if he meant to say he was sad – which is simply regretfully sorry; unhappy – which is somewhere between mad and sad; disheartened – which is a lonely sad; mad – which makes you want to argue with someone over what they had done; angry – which makes you

> want to ignore the person you are feeling this way towards; furious – which makes you want to spit; or none of the above. In order for me really to understand what people are saying I need much more than a few words mechanically placed together (Willey 1999: 63).

One explanation for this problem would be that she was unable to use information from prosody and facial expressions to narrow down the interpretation of the word *disappointed* along lines suggested above in Section 3.2.

As noted in Section 3.1, natural prosodic inputs appear to be interpreted by dedicated mind-reading mechanisms whose function is to attribute mental states on the basis of observable behaviour. Problems with the interpretation of prosody might therefore be seen as special cases of a more general problem with mind-reading that has been well established in people with autism and Asperger's syndrome (cf. Leslie 1987; Astington, Harris and Olson 1988; Perner, Frith, Leslie and Leekam 1989; Baron-Cohen 1995; Scholl and Leslie 1999). A question which has not to our knowledge been systematically investigated is whether these problems arise with all three types of prosodic input discussed in this paper (natural signs, natural signals and linguistic signals), and whether they arise equally in accidental, covert and overt information transmission.

Sabbagh (1999) suggests that prosodic difficulties in people with high-functioning autism are caused by damage to the "ability to perceive communicative intentions". In the framework outlined above (Section 3.2), this is a special-purpose ability, based on a relevance-theoretic comprehension procedure triggered by ostensive acts, which might be capable of being selectively impaired. If Sabbagh is right, people with autism should have prosodic difficulties only in the comprehension of overt communication, and not in interpreting accidental or covert forms of information transmission. According to a more standard view, utterance interpretation merely involves a combination of special-purpose linguistic abilities and the same general mind-reading abilities used to explain and predict the ordinary, non-communicative behaviour of others (cf. Bloom 2000, 2002; Sperber 2000, Sperber and Wilson 2002). On this approach, prosodic difficulties should be linked either to specific linguistic impairments or to impairments in general mind-reading abilities. A further possibility suggested by the arguments of this paper is that there may be at least three broad types of prosodic difficulty, linked to impairments in specifically linguistic abilities, general mind-reading abilities and the ability for overt communication. Clearly, all these possibilities are worth exploring. As a step in this direction, we will end by briefly surveying some experimental evidence on the localisation of prosodic impairments, and suggest some test cases which might shed further light on the relation between prosody, mind-reading and communication.

5.1 Experimental evidence

Sabbagh (1999) surveys a wide variety of prosodic impairments in autism and right hemisphere damage. These include problems with the interpretation of:

(a) emotional and attitudinal prosody
(b) 'inarticulate' prosody (e.g. grunts or sighs)
(c) contrastive stress
(d) 'intrinsic' prosody (e.g. declarative/interrogative intonation).

As noted above, he concludes that these difficulties are linked to a specifically communicative impairment. In a careful review of studies on prosody in autism, McCann and Peppé (2003) cite additional evidence showing difficulties with the interpretation of types of affective prosody, while the ability to distinguish a 'calm' from an 'agitated' attitude from prosodic clues remains intact. The framework we have outlined suggests two possible explanations for this difference: there may be selective impairment of the ability to interpret natural prosodic signals as opposed to natural prosodic signs; or there may be selective impairment of the ability to interpret natural prosodic indicators of mental rather than physical state. Both possibilities are worth investigating, and we will suggest some possible test cases below.

More generally, the right hemisphere appears to be relatively dominant in the interpretation of emotional prosody, and the left hemisphere in the interpretation of properly linguistic prosody (Ross et al. 1988, Baum and Pell 1999; Pell 2002b; though see Seddoh 2000 for scepticism about the distinction between affective and linguistic prosody). For instance, right hemisphere damage in Parkinsonism seems to affect the interpretation of emotional and sentential prosody but not lexical prosody (perhaps suggesting a universal non-linguistic basis for sentential prosody). However, while right hemisphere damage may cause difficulties with emotional prosody, it does not necessarily lead to problems with the identification of emotional facial expressions; in general, recognition of emotion from faces is reliably better than from voices (Sabbagh 1999; Pell 2000a,b).

Lieberman (2000) provides a fascinating account of the type of sub-attentive processes involved in the production and interpretation of unintentional prosodic signs and signals, which may be seen as contributing more to accidental (or covert) than to ostensive communication:

> At all times, we are communicating information about our emotional state, attitudes, and evaluations of whatever we are currently confronting ... Several of the nonverbal cues that reflect our internal state can be controlled consciously to some degree, but this will only occur if one

> directs one's conscious attention to the process of non-verbal encoding ... Additionally, there are other cues to one's internal state (e.g. tone of voice, blinking, posture) that the vast majority of us have little or no control over. We produce most of our nonverbal cues intuitively, without phenomenological awareness (Lieberman 2000: 111).

Lieberman further comments,

> The dance of nonverbal communication between two individuals often goes unnoticed by either participant ... It is, however, noticeable when it is absent or out of sync. The dance occurs intuitively, and when we get a sense of the other's state of mind as a result of the nonverbal cues the other has emitted, we often have nothing other than our intuition to justify our inferences (Lieberman 2000: 123).

According to Lieberman, sub-attentive impressions of this type are dealt with in the basal ganglia. This may be seen as providing some evidence of a neuropsychological basis for the distinction between accidental, covert and overt information transmission discussed in Sections 2-4. Further neuropsychological evidence may well shed light on the relations between prosody and linguistic, communicative and general mind-reading abilities.

5.2 Prosody, mind-reading and communication

In this paper, we have argued on theoretical grounds for two sets of distinctions: between natural signs, natural signals and non-natural signals (whether cultural or linguistic); and between accidental, covert and overt information transmission. Neither set of distinctions has been systematically applied to the literature on prosodic effects, and it is perhaps not even obvious that they could be.[14]

On the other hand, if we could find a way of applying them, it might yield useful insights into the nature of the impairments in autism and right hemisphere damage (do they affect general mind-reading abilities or specific communicative abilities?), and more generally, into the relation between prosody, mind-reading and communication. We would therefore like to end by suggesting a range of possible test cases which might be used in investigating the prosodic difficulties that arise in autism, Asperger's syndrome and right hemisphere damage, and might provide a first step towards a more systematic application of these distinctions.

---

[14] A further complication is that there is, of course, a *cultural* element to language itself, and the question of what differentiates the cultural from the linguistic is a complex one.

The first type of test case would consist of *natural prosodic signals* which are not overtly shown, and which would not normally be understood as contributing to a communicator's meaning. Examples might be someone trying to hide her anger while speaking, sighing while working alone in her room, or exclaiming with surprise when she drops something while no-one else is present. Comprehension of these non-ostensive signals in people with prosodic impairments might be compared with comprehension of cases where the same natural prosodic signal is ostensively used in addressing someone, and would normally be understood as contributing to the communicator's meaning.

The second type of test case would consist of *natural prosodic signs* which are not overtly shown, and which would not normally be understood as contributing to a communicator's meaning. Examples might be saying "The bus is coming" while sounding bored, tired, shaky or ill. Interpretation of these natural prosodic clues might be compared with comprehension of cases where the same natural prosodic sign is ostensively used, and would normally be understood as contributing to the communicator's meaning.

The third type of test case would involve natural prosodic signs of *physical* rather than mental state. Examples might be drunkenness, breathlessness, stammering, etc. These are interesting because they contribute to the speaker's meaning when ostensively used, but to neither speaker's meaning nor mind-reading when not ostensively used. Comprehension of these types of case might therefore yield particular insight into the relations between prosody, pragmatics and mind-reading, and a better understanding of these relations may suggest further directions for the study of prosody and pragmatics.

References

Astington, Janet, Harris, Paul, Olson, David, (Eds.), 1988. *Developing Theories of Mind*. Cambridge University Press, Cambridge.

Baron-Cohen, Simon, 1995. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, Cambridge MA.

Baum, S., Pell, Marc, 1999. The neural basis of prosody: insights from lesion studies and neuroimaging. *Aphasiology* 13, 581-608.

Blakemore, Diane, 1987. *Semantic Constraints on Relevance*. Blackwell, Oxford.

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

Bloom, Paul, 2002. Mindreading, communication and the learning of names for things. *Mind and Language* 17, 37-54.

Bolinger, Dwight, 1983. The inherent iconism of intonation. In Haiman, J, (Ed.), *Iconicity in Syntax*. John Benjamins, Amsterdam, 97-109.

Borod, Joan, Pick, Lawrence, Hall, Susan, Sliwinski, Martin, Madigan, Nancy, Obler, Loraine, Welkowitz, Joan, Canino, Elizabeth, Erhan, Hulyan, Goral, Mira, Morrison, Chris, Tabert,

Matthias, 2000. Relationships among facial, prosodic and lexical channels of emotional perceptual processing. *Cognition and Emotion* 14, 193-211.
Carpenter, Malinda, Pennington, Bruce, Rogers, Sally, 2001. Understanding of others' intentions in children with autism. *Journal of Autism and Developmental Disorders* 31, 589-599.
Carston, Robyn, 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Blackwell, Oxford.
Chen, Aoju, Gussenhoven, Carlos, 2003. Language-dependence in signalling of attitude in speech. In: Suzuki, N., Bartneck, C. (Eds.), *Proceedings of Workshop on the Subtle Expressivity of Emotion*, CHI 2003 Conference on Human and Computer Interaction.
Cheney, Dorothy, Seyfarth, Robert, 1990. *How Monkeys See the World*. University of Chicago Press, Chicago.
Cheney, Dorothy, Seyfarth, Robert, 1999, Mechanisms underlying the vocalizations of non-human primates. In Hauser, Marc, Konishi, Mark, (Eds), *The Design of Animal Communication*. Bradford/MIT Press, Cambridge, 629-644.
Clark, Billy, Lindsey, Geoff, 1990. Intonation, grammar and utterance interpretation. *UCL Working Papers in Linguistics* 2, 32-51.
Darwin, Charles, 1872. *The Expression of the Emotions in Man and Animals*. Edited with foreword by Ekman, P. 1998. Harper Collins, London.
Dolan, Raymond, Morris, John, de Gelder, Beatrice, 2001. Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences* USA, 98, 10006-10010.
Ekman, Paul, 1989. The argument and evidence about universals in facial expressions of emotion. In: Wagner, H., Manstead, A. (Eds.), *Handbook of Social Psychophysiology*. Wiley, New York, 143-164.
Ekman, Paul, 1993. Facial Expression and Emotion. *American Psychologist*, 48, 384-392.
Ekman, Paul, 1992. An argument for basic emotion. *Cognition and Emotion* 6, 169-200.
Ekman, Paul, 1999. Emotional and conversational nonverbal signals. In Messing, L, Campbell, Ruth (Eds) *Gesture, Speech and Sign*. Oxford, Oxford University Press, 45-57.
Escandell-Vidal, Victoria, 1998. Intonation and procedural encoding: The case of Spanish interrogatives. In: Rouchota, V., Jucker, A. (Eds), *Current Issues in Relevance Theory*. John Benjamins, Amsterdam, 169-203.
Escandell-Vidal, Victoria, 2002. Echo-syntax and metarepresentations. *Lingua* 112, 871-900.
Field, T., Woodson, R., Greenberg, R., Cohen, D., 1982. Discrimination and imitation of facial expressions by neonates. *Science* 218, 179-81.
Fretheim, Thorstein, 2002: Intonation as a constraint on inferential processing. In: Bel, B., Marlien, I. (Eds), *Proceedings of the Speech Prosody 2002 Conference*, 59-64.
Fridlund, Alan, 1994. *Human Facial Expression: An Evolutionary View*. Academic Press, San Diego.
Goffman, Erving, 1981. *Forms of Talk*. Blackwell, Oxford.
Grice, H. Paul, 1957. Meaning. *Philosophical Review* 66, 377-88. Reprinted in Grice 1989: 213-23.
Grice, H. Paul, 1969. Utterer's meaning and intentions. *Philosophical Review* 78, 147-77. Reprinted in Grice 1989, 86-116.
Grice, H. Paul, 1982. Meaning revisited. In: N. V. Smith (Ed.) *Mutual Knowledge*. Academic Press, London. Reprinted in Grice 1989, 283-303.
Grice, H. Paul, 1989. *Studies in the Way of Words*. Harvard University Press, Cambridge MA.
Gussenhoven, Carlos, 2002. Intonation and interpretation: phonetics and phonology. In: Bel, B, Marlien, I. (Eds) *Proceedings of the Speech Prosody 2002 Conference*, 47-57.

Gussenhoven, Carlos, 2004. *The Phonology of Tone and Intonation*. Cambridge University Press, Cambridge.

Hall, Alison, 2004. The meaning of 'but: a procedural reanalysis. *UCL Working Papers in Linguistics* 16, 199-236.

Happé, Francesca, Loth, Eva, 2002. 'Theory of mind' and tracking speakers' intentions. *Mind and Language* 17, 24-36.

Hauser, Marc, 1996. *The Evolution of Communication*. MIT Press, Cambridge MA.

Hauser, Marc, Konishi, Mark, (Eds), 1999. *The Design of Animal Communication*. Bradford/MIT Press, Cambridge.

House, Jill, 1990. Intonation structures and pragmatic interpretation. In: Ramsaran, S. (Ed.) *Studies in the Pronunciation of English*. Routledge, London, 38-57.

House, Jill, forthcoming. Constructing a context with intonation. Paper presented at the Prosody and Pragmatics Conference, Salford.

Imai, Kunihiko, 1998. Intonation and relevance. In: Carston, R, Uchida, S. (Eds) *Relevance Theory: Applications and Implications*. John Benjamins, Amsterdam, 69-86.

Kendon, Adam, 1988. How gestures can become like words. In: Poyotas, F. (Ed.) *Cross-cultural Perspectives in Nonverbal Communication*. Hogrefe, Toronto, 131-141.

König, Ekkehard, 1991. *The Meaning of Focus Particles: A Comparative Perspective*. Routledge, London.

Ladd, D. Robert, 1978, *The Structure of Intonational Meaning*. Indiana University Press, London.

Ladd, D. Robert, 1996. *Intonational Phonology*. Cambridge University Press, Cambridge.

Leslie, Alan, 1987. Pretense and representation: The origins of 'theory of mind'. *Psychological Review* 94, 412-426.

Lieberman, Matthew, 2000. Intuition: A social-cognitive neuroscience approach. *Psychological Bulletin* 126, 109-137.

McCann, Joanne, Peppé, Susan, 2003. Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders*, 38, 325-50.

Michelson, Axel, 1999. The dance language of honeybees: recent findings and problems. In Hauser, Marc, Konishi, Mark, (Eds), *The Design of Animal Communication*. Bradford/MIT Press, Cambridge, 111-132.

Millikan, Ruth, 1984. *Language, Thought, and other Biological Categories*. MIT Press, Cambridge MA.

Origgi, Gloria, Sperber, Dan, 2000. Evolution, communication and the proper function of language. In: Carruthers, P and Chamberlain, A. (Eds), *Evolution and The Human Mind: Modularity, Language and Meta-Cognition*. Cambridge University Press, Cambridge, 140-169.

Nelson, C.A., De Haan, M., 1996. Neural correlates of infants' visual responsiveness to facial expressions of emotion. *Developmental Psychobiology*, 29, 577-595.

Pell, Marc, 2002a, Surveying emotional prosody in the brain. In: Bel, B and Marlien, I. (Eds) *Proceedings of the Speech Prosody 2002 Conference*, 77-82.

Pell, Marc, 2002b. Evaluation of nonverbal emotion in face and voice: Some preliminary findings on a new battery of tests. *Brain and Cognition* 48, 499-514.

Perner, J, Frith, Ute, Leslie, Alan, Leekam, S 1989. Explorations of the autistic child's theory of mind: knowledge, belief and communication. *Child Development*, 689-700

Phillips, R. D., Wagner, S. H., Fells, C.A., Lynch, M. 1990. Do infants recognise emotion in facial expressions? Categorical and metaphorical evidence. *Infant Behaviour and Development* 13, 71-84.

Recanati, Francois, 1987. *Meaning and Force: The Pragmatics of Performative Utterances*. Cambridge University Press, Cambridge.

Ross, E., Edmonson, J., Seibert, G, Homan R, 1988. Acoustic analysis of affective prosody during right-sided Wada test: a within-subjects verification of the right hemisphere's role in language. *Brain and Language* 33, 128-145.

Sabbagh, Mark, 1999. Communicative intentions and language: Evidence from right-hemisphere damage and autism. *Brain and Language* 70, 29-69.

Scheuer, J., 1995. Relevance and prosody in spoken Danish. *Journal of Pragmatics* 23, 421-447.

Schiffer, Steven, 1972. *Meaning*. Clarendon Press, Oxford.

Scholl, Brian, Leslie, Alan 1999. Modularity, development and 'theory of mind'. *Mind and Language* 14, 131-153.

Seddoh, S. Amebu, A., 2002. How discrete or independent are 'affective' prosody and 'linguistic' prosody? *Aphasiology* 16, 683-692.

Sperber, Dan, 1994. The modularity of thought and the epidemiology of representations. In Hirschfeld, L., Gelman, S. (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge University Press, Cambridge, 39-67.

Sperber, Dan, 1996. *Explaining Culture: A Naturalistic Approach*. Blackwell, Oxford.

Sperber, Dan, 2000a. Metarepresentation. Lectures at the London School of Economics.

Sperber, Dan, 2000b. Metarepresentations in an evolutionary perspective. In: Sperber, D. (Ed.), *Metarepresentations: An Interdisciplinary Perspective*. Oxford University Press, New York, 117-137.

Sperber, Dan, forthcoming a. Seedless grapes. To appear in: Laurence, S., Margolis, E. (Eds), *Creations of the Mind: Theories of Artifacts and their Representation*. Oxford University Press, Oxford.

Sperber, Dan, forthcoming b. Modularity and relevance. To appear in: Carruthers, P., Laurence, S., Stich, S. (Eds), *The Innate Mind: Structure and Content*. Oxford University Press, New York.

Sperber, Dan, Cara, Francisco, Girotto, Vittorio 1995. Relevance theory explains the selection task. *Cognition*, 52, 3-39.

Sperber, Dan, Wilson, Deirdre, 1986/95. *Relevance: Communication and Cognition*. Blackwell, Oxford. (second edition 1995).

Sperber, Dan, Wilson, Deirdre, 1995. Postface to the second edition of *Relevance.* Blackwell, Oxford.

Sperber, Dan, Wilson, Deirdre, 1997. Remarks on relevance theory and the social sciences. *Multilingua* 16, 145-151.

Sperber, Dan, Wilson, Deirdre, 2002. Pragmatics, modularity and mindreading. *Mind and Language* 17, 3-23.

Vandepitte, Sonia, 1989. A pragmatic function of intonation. *Lingua* 79, 265-297.

von Frisch, Karl, 1967. *The Dance Language and Orientation of Bees*. Harvard University Press, Cambridge MA.

van Hooff, J., 1972. A comparative approach to the phylogeny of laughter and smiling. In Hinde, R., (Ed.), *Non-verbal Communication*. Cambridge University Press, Cambridge, 209-238.

Wharton, Tim, 2003a. Natural pragmatics and natural codes. *Mind and Language* 18, 447-477.

Wharton, Tim, 2003b. Interjections, language and the 'showing-saying' continuum. *Pragmatics and Cognition* 11, 39-91.

Wichmann, Anne, 2000. The attitudinal effects of prosody, and how they relate to emotion. In: *Proceedings of SpeechEmotion-2000*, 143-148.

Wichmann, Anne, 2002. Attitudinal intonation and the inferential process. In: Bel, B., Marlien, I. (Eds), *Proceedings of the Speech Prosody 2002 Conference*, 11-16.

Willey, Liane Holliday, 1999. *Pretending to be Normal*. Jessica Kingsley, London.

Williams, Donna, 1994. *Somebody Somewhere*. Jessica Kingsley, London.
Wilson, Deirdre, 2003. New directions for research on pragmatics and modularity. *UCL Working Papers in Linguistics* 15, 105-127. Revised version in *Lingua* 115(2005): 1129-1146.
Wilson, Deirdre, Sperber, Dan, 1993. Linguistic form and relevance. *Lingua* 90, 1-25.
Wilson, Deirdre, Sperber, Dan, 2002. Truthfulness and relevance. *Mind* 111, 583-632.
Wilson, Deirdre, Sperber, Dan, 2004. Relevance theory. In: Horn, L., Ward, G. (Eds), *The Handbook of Pragmatics.* Blackwell, Oxford, 607-632.