(In *Mind and Language,* 2002, 17. 3-23)

Pragmatics, Modularity and Mind-reading

DAN SPERBER AND DEIRDRE WILSON

Abstract

The central problem for pragmatics is that sentence meaning vastly underdetermines speaker's meaning. The goal of pragmatics is to explain how the gap between sentence meaning and speaker's meaning is bridged. This paper defends the broadly Gricean view that pragmatic interpretation is ultimately an exercise in mind-reading, involving the inferential attribution of intentions. We argue, however, that the interpretation process does not simply consist in applying general mind-reading abilities to a particular (communicative) domain. Rather, it involves a dedicated comprehension module, with its own special principles and mechanisms. We show how such a metacommunicative module might have evolved, and what principles and mechanisms it might contain.

We would like to thank the participants in the *Mind & Language* Workshop on Pragmatics and Cognitive Science, and in particular Robyn Carston and Sam Guttenplan, for valuable comments and suggestions.

Address for correspondence: Deirdre Wilson, Department of Linguistics, University College London, Gower Street, London WC1E 6BT.

Email: deirdre@ling.ucl.ac.uk

1. Introduction

Pragmatic studies of verbal communication start from the assumption (first defended in detail by the philosopher Paul Grice), that an essential feature of most human communication, both verbal and non-verbal, is the expression and recognition of intentions (Grice, 1957; 1969; 1982; 1989a). On this approach, pragmatic interpretation is ultimately an exercise in metapsychology, in which the hearer infers the speaker's intended meaning from evidence she has provided for this purpose. An utterance is, of course, a linguistically-coded piece of evidence, so that verbal comprehension involves an element of decoding. However, the decoded linguistic meaning is merely the starting point for an inferential process that results in the attribution of a speaker's meaning.

The central problem for pragmatics is that the linguistic meaning recovered by decoding vastly underdetermines the speaker's meaning. There may be ambiguities and referential ambivalences to resolve, ellipses to interpret, and other indeterminacies of explicit content to deal with. There may be implicatures to identify, illocutionary indeterminacies to resolve, metaphors and ironies to interpret. All this requires an appropriate set of contextual assumptions, which the hearer must also supply. To illustrate, consider the examples in (1) and (2):

(1) (a) They gave him life.

(b) Everyone left.

(c) The school is close to the hospital.

(d) The road is flat.

(e) Coffee will be served in the lounge.

(2) (a) The lecture was as you would expect.

(b) Some of the students did well in the exam.

(c) Someone's forgotten to take out the rubbish.

(d) *Teacher:* Have you handed in your essay?

Student: I've had a lot to do recently.

(e) John is a soldier.

In order to decide what the speaker intended to assert, the hearer may have to disambiguate and assign reference, as in (1a), fix the scope of quantifiers, as in (1b), and assign appropriate interpretations to vague expressions or approximations, as in (1c-d). In order to decide what speech act the speaker intended to perform, he may have to resolve illocutionary indeterminacies, as in (1e) (which may be interpreted as an assertion, a request or a guess). Many utterances also convey implicit meaning (implicatures): for example, (2a) may implicate that the lecture was good (or bad), (2b) may implicate that not all the students did well in the exam, (2c) may convey an indirect request and (2d) an indirect answer, while (2e) may be literally, metaphorically or ironically intended. Pragmatic interpretation involves the resolution of such linguistic indeterminacies on the basis of contextual information. The hearer's task is to find the meaning the speaker intended to convey, and the goal of pragmatic theory is to explain how this is done.

Most pragmatists working today would agree with this characterisation of pragmatics. Most would also agree that pragmatic interpretation is ultimately a non-demonstrative inference process which takes place at a risk: there is no guarantee that the meaning constructed, even by a hearer correctly following the best possible procedure, is the one the speaker intended to convey. However, this picture may be fleshed out in several different ways, with different implications for the relation of pragmatics to other cognitive systems. On the one hand, there are those who argue that most, if not all, aspects of the process of constructing a hypothesis about the speaker's meaning are closely related to linguistic decoding. These code-like aspects of interpretation might be carried out within an extension of the language module, by non-metapsychological processes whose output might then be inferentially evaluated and attributed as a speaker's meaning. On the other hand, there are those who see pragmatic interpretation as metapsychological through and through. On this approach, both hypothesis construction and

hypothesis evaluation are seen as rational processes geared to the recognition of speakers' intentions, carried out by Fodorian central processes (Fodor, 1983), or by a 'theory of mind' module dedicated to the attribution of mental states on the basis of behaviour (Astington, Harris and Olson, 1988; Davies and Stone, 1995a; 1995b; Carruthers and Smith, 1996). Both positions are explored in the papers in this volume.

We want to defend a view of pragmatic interpretation as metapsychological through and through. However, departing from our earlier views (Sperber and Wilson, 1986/1995; Wilson and Sperber, 1986), we will argue that pragmatic interpretation is not simply a matter of applying Fodorian central systems or general mind-reading abilities to a particular (communicative) domain. Verbal comprehension presents special challenges, and exhibits certain regularities, not found in other domains. It therefore lends itself to the development of a dedicated comprehension module with its own particular principles and mechanisms. We will show how such a metacommunicative module might have evolved as a specialisation of a more general mind-reading module, and what principles and mechanisms it might contain; we will also indicate briefly how it might apply to the resolution of linguistic indeterminacies such as those in (1) and (2) (for fuller accounts, see Sperber and Wilson, 1986/1995; Carston, forthcoming; Wilson and Sperber, forthcoming).

2. Two Approaches to Communication

Before Grice's pioneering work, the only available theoretical model of communication was what we have called the classical code model (Sperber and Wilson, 1986/1995, chapter 1, sections 1-5; Wilson, 1998), which treats communication as involving a sender, a receiver, a set of observable signals, a set of unobservable messages, and a code that relates the two. The sender selects a message and transmits the corresponding signal, which is received and decoded at the other end; when all goes well, the result is the reproduction in the receiver of the original message. Coded communication need involve no metapsychological abilities. It clearly exists in nature, both in pure and mixed forms (in which coding and inference are combined). Much animal communication is purely coded: for example, the bee dance used to indicate the direction and distance of nectar (von Frisch, 1967; Hauser, 1996). It is arguable that some human non-

verbal communication is purely coded: for example, the interpretation by neonates of facial expressions of emotion (Fridlund, 1994; Sigman and Kasari, 1995; Wharton, 2001). Human verbal communication, by contrast, involves a mixture of coding and inference. As we have seen, it contains an element of inferential intention-attribution; but it is also partly coded, since the grammar of a language just is a code which pairs phonetic representations of sentences with semantic representations of sentences.

In studying such a mixed form of communication, there is room for debate about where the borderline between coding and inference should be drawn. One way of limiting the role of metapsychological processes in verbal comprehension would be to argue for an extension in the domain of grammar, and hence in the scope of (non-metapsychological) linguistic decoding processes. This is sometimes done by postulating hidden linguistic constituents or multiple ambiguities; approaches along these lines are suggested by Millikan (1984, 1988) and Stanley (this volume) (for discussion, see Origgi and Sperber, 2000; Carston, 2000; and Breheny, this volume). But however far the domain of grammar is expanded, there comes a point at which pragmatic choices – choices based on contextual information – must be made. An obvious example of a pragmatic process is reference resolution, where the hearer has to choose among a range of linguistically possible interpretations of a referential expression (e.g. 'I', 'now', 'this', 'they') on the basis of contextual information. Here, too, it is possible to argue that code-like procedures play a role in determining how pragmatic choices are made.

Many formal and computational approaches to linguistics suggest that certain aspects of pragmatic interpretation may be dealt with in code-like terms. One way of handling reference resolution along these lines is to set up contextual parameters for the speaker, hearer, time of utterance, place of utterance, and so on, and treat the interpretation of referential expressions such as 'I', 'you', 'here' and 'now' as initially determined by reference to these (e.g. Lewis, 1970; Kaplan, 1989). There are also code-like ('default-based') treatments of generalised conversational implicatures (e.g. the implicature regularly carried by (2b) above that not all the students passed the exam) (see for example Gazdar, 1979; Lascarides and Asher, 1993; Levinson, 2000). These formal accounts might be combined with an inferential approach by assuming that the output of these non-metapsychological pragmatic decoding processes is inferentially evaluated before being attributed as a speaker's meaning.

Grice himself seems to have seen explicit communication as largely a matter of linguistic and contextual decoding, and only implicit communication as properly inferential (Grice 1989: 25), and many pragmatists have followed him on this (Searle, 1969; Bach and Harnish, 1979; Levinson, 1983; Bach, 1994; for discussion, see the papers by Breheny; Carston; Recanati; and Stanley, this volume). However, the code-like pragmatic rules that have been proposed so far do not work particularly well. For example, even if 'now' refers to the time of utterance, it is still left to the hearer to decide whether the speaker, on a given occasion, meant *now this second, this minute, this hour, day, week,* year, etc. (Predelli, 1998). For other referential expressions (e.g. 'he', 'they', 'this', 'that'), and for disambiguation and the other aspects of explicit communication illustrated in (1) above, it is hard to think of a code-like treatment at all. Similarly, default-based accounts of generalised conversational implicatures typically over-generate (Carston, 1997), and it is widely acknowledged that particularised implicatures (which depend on special features of the context) are not amenable to code-like treatment at all (Levinson, 2000).

What the available psycholinguistic evidence shows is that, other things being equal, from a range of contextually-available interpretations, hearers tend to choose the most salient or accessible one, the one that costs the least processing effort to construct (Gernsbacher, 1995). This is also what many theoretical accounts of pragmatic interpretation (e.g. Lewis, 1979; Sperber and Wilson, 1986/1995) predict that hearers should do. The question is whether they do this because they are following a conventional, code-like procedure that children have to learn (as they have to learn that 'I' refers to the speaker, 'now' to the time of utterance, and so on), or because this is a sound way of inferring the speaker's intentions, independently of any convention. If it is such a rational procedure, then it falls outside the scope of a decoding model and inside an inferential account. We will argue that, within the specifically communicative domain, it is indeed rational for hearers to follow a path of least effort in constructing a hypothesis about the speaker's meaning, and that the pragmatic interpretation process is therefore genuinely inferential (for discussion, see Origgi and Sperber, 2000; Carston, this volume; Recanati, this volume).

Inferential comprehension, then, is ultimately a metapsychological process involving the construction and evaluation of a hypothesis about the communicator's meaning on the basis of evidence she has provided for this purpose. It clearly exists in humans, both in pure

and mixed forms. As we have seen, verbal communication involves a mixture of coding and inference, and there is room for debate about the relative contributions of each. By contrast, much non-verbal communication is purely inferential. For example, when I point to the clouds to indicate that I was right to predict that it would rain, or hold up my full glass to indicate that you need not open a new bottle on my account, there is no way for you to decode my behaviour, and no need for you to do so. You could work out what I intend to convey by a straightforward exercise in mind-reading, by attributing to me the intention that would best explain my behaviour in the situation (though if we are right, you can actually do it even more directly, via a dedicated comprehension procedure). Thus, metapsychological inference plays a central role in human communication, both verbal and non-verbal.

These theoretical arguments are confirmed by a wealth of experimental evidence linking the development and breakdown of general mind-reading abilities and communicative abilities, both verbal and non-verbal. In autism, both general mind-reading and non-verbal communication are impaired (Baron-Cohen, 1995; Perner, Frith, Leslie and Leekam, 1989; Sigman and Kasari 1995; see also Langdon, Davies and Coltheart, this volume). There are also links between the development and breakdown of general mind-reading and verbal communication (Happé 1993; Wilson 2000; and the papers in this volume by Bloom; Happé and Loth; Langdon, Davies and Coltheart; and Papafragou). For example, normal word learning involves the ability to track speakers' intentions, and correlates in interesting ways with the ability to pass the false-belief tasks used in the study of general mind-reading (Bloom, 2000, this volume; Happé and Loth, this volume). Reference resolution is another pragmatic ability that correlates in interesting ways with the ability to pass false-belief tasks (Mitchell, Robinson and Thompson, 1999); and there seems to be a well-established correlation between the interpretation of irony and second-order mind-reading abilities, (Happé, 1993; Langdon, Davies and Coltheart, this volume). However, there are different ways of analysing both general mind-reading abilities and their links to specifically communicative abilities. In the next section, we will consider some of these.

3. Two Approaches to Inferential Communication

Grice was rather non-committal on the source of pragmatic abilities and their place in the overall architecture of the mind. He wanted to be able to show that our communicative behaviour is rational:

I am enough of a rationalist to want to find a basis that underlies these facts, undeniable though they may be; I would like to be able to think of the standard type of conversational practice not merely as something that all or most do *in fact* follow but as something that it is reasonable for us to follow, that we *should not* abandon. (Grice, 1989b, p. 29)

However, he was prepared to retreat, if necessary, to the 'dull but, no doubt at a certain level, adequate answer' that 'it is just a well-recognized empirical fact that people do behave in these ways; they learned to do so in childhood and have not lost the habit of doing so' (Grice, 1989b, p. 28-9).

He was equally non-committal on the form of the comprehension process. What he clearly established was a link between pragmatic abilities and more general mind-reading abilities. But mind-reading itself can be analysed in rather different ways. It may be thought of as a conscious, reflective activity, involving Fodorian central processes, and many of Grice's remarks about the derivation of implicatures are consistent with this. For example, his rational reconstruction of how conversational implicatures might be derived is a straightforward exercise in 'belief-desire' psychology:

He said that P; he could not have done this unless he thought that Q; he knows (and knows that I know that he knows) that I will realise that it is necessary to suppose that Q; he has done nothing to stop me thinking that Q; so he intends me to think, or is at least willing for me to think, that Q. (Grice, 1989b, p. 30-31)

For Grice, calculability was an essential property of implicatures, and he gave several examples of how particular implicatures might be derived using a 'working-out schema' like the one given above. But there are several reasons for thinking that the actual comprehension process should not be modelled along these lines.

In the first place, it is hard to imagine even adults going through such lengthy chains of inference in the attribution of speaker meanings. Yet preverbal infants already appear to be heavily involved in inferential communication, and they are surely not using the form of conscious, discursive reasoning illustrated in Grice's 'working-out schema' (see the papers by Bloom; Happé and Loth; and Papafragou, this volume). In the second place, we have argued above that Grice substantially underestimated the amount of metapsychological inference involved in comprehension. Given the failure of the non-metapsychological pragmatic decoding account, his 'working-out schema' for implicatures would have to be supplemented with further schemas designed to deal with disambiguation, reference assignment, and other inferential aspects of explicit communication. While reflective inferences of this type do occur when spontaneous inference fails to yield a satisfactory interpretation, inferential comprehension is in general an intuitive, unreflective process which takes place below the level of consciousness.

All this is more consistent with a view of inferential comprehension as falling within the domain of an intuitive 'theory of mind' module. This view is tacitly adopted in much of the literature on mind-reading, and explicitly defended by Bloom (2000, this volume). Grice himself makes remarks indicating that he might not have been averse to a modularised implementation of his approach, in which the recovery of implicatures was treated as an intuitive rather than a reflective process:

> The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a conversational implicature; it will be a conventional implicature. (Grice, 1989b, p. 31)

There is thus no requirement in the Gricean framework that implicatures should actually be recovered by reflective reasoning. A modular view is also possible.

There has been a strong (though by no means unanimous) trend in the development of the cognitive sciences, and in particular in developmental and evolutionary psychology and in neuropsychology, towards a more modular view of the mind. (We use 'module' in a looser sense than the one suggested by Fodor, 1983, to mean a domain- or task-specific autonomous computational mechanism; see Sperber, 1996, chapter 6; forthcoming.) One reason for this trend is that a general-purpose inferential mechanism can only derive conclusions based on the formal (logical or statistical) properties of the input information it processes. By contrast, a dedicated inferential mechanism or module can take advantage of regularities in its specific domain, and use inferential procedures which are justified by these regularities, but only in this domain. Typically, dedicated modules exploit the relatively 'fast and frugal heuristic' (Gigerenzer et al., 1999) afforded by their special domain.

A cognitive ability may become modularised in the course of cognitive development, as in the case of reading or chess expertise. However, it is reasonable to assume that many modular structures have a strong genetic component. The selection pressures which lead to the emergence of cognitive systems over evolutionary time must also tend to make these systems more efficient, and in particular to attune them, via dedicated mechanisms, to the specific problems and opportunities it is their function to handle. Much developmental evidence also suggests that infants and young children come equipped with domain-specific cognitive mechanisms (Hirschfeld and Gelman, 1994; Barkow, Cosmides and Tooby, 1995). Mind-reading is one of the best-evidenced cases in this respect.

Most theories of mind-reading do assume that it is performed not by a general-purpose reasoning mechanism, which takes as premises a number of explicit hypotheses about the relationships between behaviour and mental states, but by a dedicated module. What is still open to debate is how this module exploits the regularities in intentional behaviour. According to the rationalisation (or 'theory-theory') account, the mind-reading module carries out a form of belief-desire reasoning which differs from the 'folk-psychology' of philosophers not so much in its logic as in the fact that it is modularised: that is, performed automatically, unconsciously, and so on. On this approach, mind-reading is a form of automatic inference

to the best rationalisation of behaviour. It involves, in particular, the attribution to the agent of beliefs and desires that would make her observed behaviour rational given its actual or likely effects. Another possibility (proposed by the 'simulation theory') is that mind-reading succeeds by exploiting similarities between the interpreter and the agent whose behavior is being interpreted, and amounts to a form of simulation. However, while it is true that an utterance is a type of action, and a speaker's meaning is a type of intention, we want to argue that neither the rationalisation nor the simulation view of mind-reading adequately accounts for the hearer's ability to retrieve the speaker's meaning.

According to the rationalisation account (e.g. Davies and Stone, 1995b; Carruthers and Smith, 1996), the procedure for inferring the intention behind an action should be as follows: first, decide what effect of the action the agent could have both predicted and desired; second, assume that this was the effect the agent intended to achieve. In most cases of utterance interpretation, this rationalisation procedure would not work, because the desired effect just *is* the recognition of the speaker's intention. As we have seen, the gap between sentence meaning and speaker's meaning is so great (going well beyond the standard ambiguities normally considered in the literature) that there may be no way of listing the possible speaker's meanings without some advance knowledge – however sketchy – of what she might want to convey. Moreover, the range of possible speaker's meanings that the hearer is able to reconstruct may include several candidates that, to the best of his knowledge, the speaker might have wanted to convey. In other words, only a hearer with some advance knowledge of at least the gist of what the speaker might have wanted to convey would find it relatively easy to reconstruct the intention behind her utterance using a rationalisation procedure. But we often say or write things that our hearers or readers did not anticipate, and we have no particular reason to doubt that we will be understood. In such cases, the standard procedures for inferring intentions do not help with identifying the speaker's meaning. Unlike what happens in regular cases of intention attribution, hearers cannot *first* identify a desirable effect of the utterance, and *then* infer that the speaker's intention was precisely to achieve this effect.

According to the simulation account (e.g. Davies and Stone, 1995a), we attribute intentions by imaginatively simulating the action we are interpreting, thus discovering in ourselves the intention that underlies it. As an account of comprehension, this is not too promising either.

Since the same sentence can be used to convey quite different meanings in different situations, a hearer who is simulating the speaker's linguistic action in order to retrieve her meaning must provide a considerable amount of contextualisation, based on particular hypotheses about the speaker's beliefs, preferences, and so on. Again, this would only work in cases where the hearer already has a fairly good idea of what the speaker is likely to mean. On this approach, the routine communication of genuinely unanticipated contents would be difficult or impossible to explain.

More generally, the problem of applying a general procedure for inferring intentions from actions to the special case of inferring speaker's meanings from utterances is that speaker's meanings typically carry a vastly greater amount of information than more ordinary intentions. This is true whether information is treated in quantitative probabilistic or qualitative semantic terms. In the repertoire of human actions, utterances are much more differentiated than other types of actions: many utterances are wholly new, whereas it is relatively rare to come across actions that are not reiterations of previous actions. While stereotypical utterances ('Nice day, isn't it?') make up a significant proportion of all uttered sentence *tokens*, they are only a minute proportion of all uttered sentence *types*. Leaving stereotypical utterances aside, the prior probability of most utterances ever occurring is close to zero, as Chomsky pointed out long ago. Semantically, the complexity of ordinary intentions is limited by the range of possible actions, which is in turn constrained by many practicalities. There are no such limitations on the semantic complexity of speaker's meanings. Quite simply, we can say so much more than we can do. Regular intention attribution, whether achieved via rationalisation or simulation, is greatly facilitated by the relatively narrow range of possible actions available to an agent at a time. There is no corresponding facilitation in the attribution of speaker's meanings. It is simply not clear how the standard procedures for intention attribution could yield attributions of speaker's meanings, except in easy and trivial cases.

Add to this the fact that, on both Gricean and relevance-theoretic accounts, there are always several levels of metarepresentation involved in inferential comprehension, while in regular mind-reading a single level is generally enough (Grice, 1989b; Sperber and Wilson, 1986/1995, chapter 1). It is hard to believe that two-year-old children, who fail for instance on regular first-order false-belief tasks, can recognise and understand the peculiar multi-level representations involved in communication, using nothing more than

a general ability to attribute intentions to agents in order to explain their behaviour. All this makes it worth exploring the possibility that, within the overall 'theory of mind' module, there has evolved a specialised sub-module dedicated to comprehension, with its own proprietary concepts and mechanisms (Sperber 1996, 2000).

Given the complexity of mind-reading, the variety of tasks it has to perform, and the particular regularities exhibited by some of these tasks, it is quite plausible to assume that it involves a variety of sub-modules. A likely candidate for one sub-module of the mind-reading mechanism is the ability, already present in infants, to infer what people are seeing or watching from the direction of their gaze. Presumably, the infant (or indeed the adult) who performs this sort of inference is not feeding a general-purpose inferential mechanism with, say, a conditional major premise of the form 'If the direction of gaze of a person P is towards an object O, then P is seeing O' and a minor premise of the form 'Mummy's direction of gaze is towards the cat' in order to derive the conclusion: 'Mummy is seeing the cat.' It is also unlikely that the infant (or the adult) rationalises or simulates the observed eye-movement behaviour. In other words, the inference involved is not just an application of a relatively general and internally undifferentiated mind-reading module to the specific problem of inferring perceptual state from direction of gaze. It is much more plausible that humans are equipped from infancy with a dedicated module, an Eye Direction Detector (Baron-Cohen, 1995), which exploits the de facto strong correlation between direction of gaze and visual perception, and directly attributes perceptual and attentional states on the basis of direction of gaze. This attribution may itself provide input for other dedicated devices, such as those involved in word learning (Bloom, 2000; this volume). In infants at least, such attributions need not be available at all for domain-general inference or verbal expression.

Similarly, for reasons given above, we doubt that normal verbal comprehension is achieved either by wondering what beliefs and desires would make it rational for the speaker to have produced a given utterance, or by simulating the state of mind that might have led her to produce it. The question is: Are there regularities specific to the production of utterances (or of communicative behaviour more generally) which might ground a more effective dedicated procedure for inferring a speaker's meaning from her utterance? If there are, they are not immediately obvious, unlike the strong and simple correlation between gaze direction and visual attention. Nevertheless, we have argued (Sperber and Wilson, 1986/1995; Sperber, 2000;

Wilson, 2000) that human communication exploits a tendency of human cognition to seek relevance in a way that narrowly constrains the interpretation of utterances, thus providing inferential comprehension with a strong regularity in the data which justifies a dedicated procedure. In the next section, we will outline these claims, adopting an evolutionary perspective.

4. Relevance, cognition and communication

Two kinds of evolutionary transformation may be distinguished. Some are continuous, and involve the gradual increase or decrease of a variable such as body size or visual acuity. Others are discrete, and involve the gradual emergence of a new trait or property, such as eyes or wings. We claim that relevance has been involved in two evolutionary transformations in human cognition: one continuous, and the other discrete. The continuous transformation has been an increasing tendency of the human cognitive system to maximise the relevance of the information it processes. The discrete transformation has been the emergence of a relevance-based comprehension module.

Cognitive efficiency, like any other kind of efficiency, is a matter of striking the best possible balance between costs and benefits. In the case of cognition, the cost is the mental effort required to construct representations of actual or desired states of affairs, to retrieve stored information from memory, and to draw inferences. The benefits are cognitive effects: that is, enrichments, revisions and reorganisations of existing beliefs and plans, which improve the organism's knowledge and capacity for successful action (Sperber and Wilson, 1986/1995).

In most animal species, the function of cognition is to monitor quite specific features of the environment (or of the organism itself) which enable it to exploit opportunities (for feeding, mating, and so on) and avoid dangers (from predators, poisonous food, and so on). For these animals, cognitive efficiency is a matter of achieving these benefits at the lowest possible cost. When the environment of such a species has remained stable enough for long enough, there is likely to have been a continuous transformation in the direction of greater efficiency, involving, in particular, a reduction in the costs required to achieve the given range of benefits. In some cases, this increase in efficiency may also have involved the emergence of cognitive mechanisms attuned to specific aspects of the environment, which

provide new cognitive benefits: this would be an example of a discrete transformation.

In humans, a considerable amount of cognitive activity is spent in processing information which has no immediate relevance to improving the organism's condition. Instead, a massive investment is made in developing a rich, well-organised data-base of information about a great many diverse aspects of the world. Some—though not all—of these data will turn out to be of practical use, perhaps in unforeseen ways. As a result, humans have an outstanding degree of adaptability to varied and changing environmental conditions, at the cost of a uniquely high investment in cognition.

Human cognition has three notable characteristics: it involves the constant monitoring of a wide variety of environmental features, the permanent availability (with varying degrees of accessibility) of a huge amount of memorised data, and a capacity for effortful attentional processing which can handle only a rather limited amount of information at any given time. The result is an attentional bottleneck: only a fraction of the monitored environmental information can be attentionally processed, and only a fraction of the memorised information can be brought to bear on it. Not all the monitored features of the environment are equally worth attending to, and not all the memorised data are equally helpful in processing a given piece of environmental information. Cognitive efficiency in humans is primarily a matter of being able to select, from the environment on the one hand, and from memory on the other, information which it is worth bringing together for joint – and costly – attentional processing.

What makes information worth attending to? There may be no general answer to this question, but merely a long list of properties – practical usefulness, importance to the goals of the individual, evocative power, and so forth – that provide partial answers. We have argued instead that all these partial answers are special cases of a truly general answer, based on a theoretical notion of relevance. Relevance, as we see it, is a potential property of external stimuli (e.g. utterances, actions) or internal representations (e.g. thoughts, memories) which provide input to cognitive processes. The relevance of an input for an individual at a given time is a positive function of the cognitive benefits that he would gain from processing it, and a negative function of the processing effort needed to achieve these benefits.

With relevance characterised in this way, it is easy to see that cognitive efficiency in humans is a matter of allocating the available attentional resources to the processing of the most relevant available inputs. We claim that in hominid evolution there has been a continuous pressure towards greater cognitive efficiency, so that human cognition is geared to the maximisation of relevance (we call this claim the First, or Cognitive, Principle of Relevance). This pressure has affected both the general organisation of the mind/brain and each of its components involved in perception, memory and inference. The result is not that humans invariably succeed in picking out the most relevant information available, but that they manage their cognitive resources in ways that are on the whole efficient and predictable.

The universal cognitive tendency to maximise relevance makes it possible, at least to some extent, to predict and manipulate the mental states of others. In particular, an individual A can often predict:

(a) which stimulus in an individual B's environment is likely to attract B's attention (i.e. the most relevant stimulus in that environment);

(b) which background information from B's memory is likely to be retrieved and used in processing this stimulus (i.e. the background information most relevant to processing it);

(c) which inferences B is likely to draw (i.e. those inferences which yield enough cognitive benefits for B's attentional resources to remain on the stimulus rather than being diverted to alternative potential inputs competing for those resources).

To illustrate: suppose that Peter and Mary are walking in the park. They are engaged in conversation; there are trees, flowers, birds and people all around them. Still, when Peter sees their acquaintance John in a group of people coming towards them, he correctly predicts that Mary will notice John, remember that he moved to Australia three months earlier, infer that there must be some reason why he is back in London, and conclude that it would be appropriate to ask him about this. Peter predicts Mary's train of thought so easily, and in such a familiar way, that it is not always appreciated how remarkable this is from a cognitive point of view. After all, there were lots of other stimuli that Mary might have noticed and paid attention to.

Even if she did pay attention to John, there were lots of other things she could have remembered about him. Even if she did remember that he had left for Australia, there were lots of other inferences she could have drawn (for example, that he had been on a plane at least twice in the past three months). So why should it be so easy for Peter to predict Mary's train of thought correctly? Our answer is that it is easy for two reasons: first, because attention, memory retrieval and inference are guided by considerations of relevance, and second, because this regularity in the data is built into our ability to read the minds of others.

Most studies of mind-reading have focused on the attribution of beliefs and desires. There has also been a lot of interest in joint attention, and particularly its role in early language acquisition. However, the understanding that we have of others routinely extends to an awareness of what they are attending to and thinking about even in situations where we ourselves are attending to and thinking about other things. There is no rich body of evidence on the development of these aspects of mind-reading. However, it would be possible to set up, as a counterpart to the famous false-belief task, a 'disjoint attention task' in which the participant has to infer what a certain character is paying attention to in a situation where there is a discrepancy between (a) what is relevant to the participant and (b) what is relevant to the character. We predict that children will succeed on well-designed tasks of this kind long before they succeed on false-belief tasks. After all, children try to manipulate the attention of others long before they try to manipulate their beliefs.

This ability to recognise what other people are attending to and thinking about, and to predict how their attention and train of thought are likely to shift when a new stimulus is presented, may be used in manipulating their mental states. An individual A may act on the mental states of another individual B by producing a stimulus which is likely:

(a) to attract B's attention;

(b) to prompt the retrieval of certain background information from B's memory;

(c) when jointly processed with the background information whose retrieval it has prompted, to lead B to draw certain inferences intended by A.

A great deal of human interaction takes this form. Individual A introduces into the environment of another individual B a stimulus which is relevant to B, and which provides evidence for certain intended conclusions. For example, Peter opens the current issue of *Time Out*, intending not only to see what films are on, but also to provide Mary with evidence that he would like to go out that evening. Mary chooses not to stifle a yawn, thereby providing Peter with evidence that she is tired. In this interaction, each participant produces a stimulus which is relevant to the other, but neither openly presents this stimulusas manifestly intended to attract the other's attention. These are covert – or at least not manifestly overt – attempts at influencing others.

However, many attempts to influence others are quite overtly made. For example, Peter may establish eye contact with Mary and tap the issue of *Time Out* before opening it, making it clear that he intends Mary to pay attention to what he is doing and draw some specific conclusion from it. Mary may not only choose not to stifle her yawn, she may openly and deliberately exaggerate it, with similar results. By engaging in such ostensive behaviour, a communicator provides evidence not only for the conclusion she intends the addressee to draw, but also of the fact that she intends him to draw this conclusion. This is 'ostensive-inferential' communication proper: that is, communication achieved by ostensively providing an addressee with evidence which enables him to infer the communicator's meaning.

Ostensive-inferential communication is not the only form of information transmission. A great deal of information is unintentionally transmitted and sub-attentively received. Some is covertly transmitted, particularly when it would be self-defeating to be open about the fact that one intends the other participant to come to a certain conclusion, as when wearing a disguise. However, ostensive-inferential communication is the most important form of information transmission among humans. In a wide range of cases, being open about one's intention to inform someone of something is the best way – or indeed the only way – of fulfilling this intention. For example, if Peter wants to go out with Mary, Mary will want to know about it; similarly, if Mary is too tired to go out, Peter will want to know about it. By being open about their intention to inform each other of something – that is, by drawing attention to their behaviour in a manifestly intentional way – each elicits the other's

co-operation, in the form of increased attention and a greater willingness to make the necessary effort to discover the intended conclusion.

Notice that ostensive-inferential communication may be achieved without the communicator providing any direct evidence for the intended conclusion. All she has to do is provide evidence of the fact that she intends the addressee to come to this conclusion. For example, Peter might just tap the cover of *Time Out* without even opening it. This is not normally part of the preparations for going out, and provides no direct evidence of his desire to go out. Still, by ostensively tapping the magazine, he does provide Mary with direct evidence that he intends her to come to the conclusion that he wants to go out. Similarly, when Mary ostensively imitates a yawn, this is not direct evidence that she is tired, but it is direct evidence that she intends Peter to come to the conclusion that she is tired. The same would be true if Peter said, 'Let's go out tonight!' and Mary replied, 'I'm tired.' Utterances do not provide direct evidence of the state of affairs they describe (notwithstanding some famous philosophical exceptions).

The fact that ostensive-inferential communication may be achieved simply by providing evidence about the communicator's intentions makes it possible to use symbolic behaviours as stimuli. These may be improvised, as when Peter taps the cover of the magazine, standardised, as in a fake yawn, or coded, as in an utterance. In each case, the symbolic stimulus provides evidence which, combined with the context, enables the audience to infer the communicator's meaning. How is this evidence used? How can it help the hearer discover the communicator's meaning when it never fully encodes it, and need not encode it at all? What procedure takes this evidence as input and delivers an interpretation of the communicator's meaning as output? This is where considerations of relevance come in.

5. Relevance and pragmatics

When it is manifest that individual A is producing an ostensive stimulus (e.g. an utterance) in order to communicate with another individual B, it is manifest that A intends B to find this stimulus worth his attention (or else, manifestly, communication would fail). Humans are good at predicting what will attract the attention of others. We have suggested that their success is based on a dedicated

inferential procedure geared to considerations of relevance. These considerations are not spelled out and used as explicit premises in the procedure, but are built into its functioning instead. So when B understands that A intends him to find her ostensive stimulus worth his attention, we can unpack his understanding in terms of the notion of relevance (terms which remain tacit in B's own understanding): A intends B to find the stimulus *relevant enough* to secure his attention.

Thus, every utterance (or other type of ostensive stimulus, though we will talk only of utterances from now on) conveys a presumption of its own relevance. We call this claim the Second, or Communicative, Principle of Relevance, and argue that it is the key to inferential comprehension (Sperber and Wilson 1986/1995, chapter 3). What exactly is the content of the presumption of relevance that every utterance conveys? In the first place, as we have already argued, the speaker manifestly intends the hearer to find the utterance at least relevant enough to be worth his attention. But the amount of attention paid to an utterance can vary: it may be light or concentrated, fleeting or lasting, and may be attracted away by alternative competing stimuli. It is therefore manifestly in the speaker's interest for the hearer to find her utterance as relevant as possible, so that he pays it due attention. However, in producing an utterance, the speaker is also manifestly limited by her abilities (to provide relevant information, and to formulate it in the best possible way) and her preferences (and in particular her goal of getting the hearer to draw not just some relevant conclusion, but a specifically intended one). So the exact content of the presumption of relevance is as follows:

Presumption of relevance

The utterance is presumed to be the most relevant one compatible with the speaker's abilities and preferences, and at least relevant enough to be worth the hearer's attention. (Sperber and Wilson, 1986/199595, p. 266-78)

The content of this presumption of relevance may be rationally reconstructed along the lines just shown, but there is no need to assume that hearers go through such a rational reconstruction process in interpreting utterances. Our suggestion is, rather, that the

presumption of relevance is built into their comprehension procedures.

The fact that every utterance conveys a presumption of its own relevance (i.e. the Communicative Principle of Relevance) motivates the use of the following comprehension procedure in interpreting the speaker's meaning:

Relevance-theoretic comprehension procedure

(a) Follow a path of least effort in computing cognitive effects. In particular, test interpretive hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility.

(b) Stop when your expectations of relevance are satisfied.

The hearer is justified in following a path of least effort because the speaker is expected (within the limits of her abilities and preferences) to make her utterance as relevant as possible, and hence as easy as possible to understand (since relevance and processing effort vary inversely). It follows that the plausibility of a particular hypothesis about the speaker's meaning depends not only on its content but also on its accessibility. In the absence of other evidence, the very fact that an interpretation is the first to come to mind lends it an initial degree of plausibility. It is therefore rational for hearers to follow a path of least effort in the particular communicative domain (though not, of course, in other domains).

The hearer is also justified in stopping at the first interpretation that satisfies his expectations of relevance because, if the speaker has succeeded in producing an utterance that satisfies the presumption of relevance it conveys, there should never be more than one such interpretation. A speaker who wants to make her utterance as easy as possible to understand should formulate it (within the limits of her abilities and preferences) in such a way that the first interpretation to satisfy the hearer's expectations of relevance is the one she intended to convey. It is not compatible with the presumption of relevance for an utterance to have two alternative co-occurring interpretations, either of which would be individually satisfactory, since this would put the hearer to the unnecessary extra effort of trying to choose

between them. Thus, when a hearer following the path of least effort arrives at an interpretation which satisfies his expectations of relevance and is compatible with what he knows of the speaker, this is the most plausible hypothesis about the speaker's meaning for him. Since comprehension is a non-demonstrative inference process, this hypothesis may well be false; but it is the best a rational hearer can produce. (Note, incidentally, that the hearer's expectations of relevance may be readjusted in the course of comprehension. For example, it may turn out that the effort of finding any interpretation at all would be too great: as a result, the hearer would disbelieve the presumption of relevance and terminate the process, with his now null expectation of relevance trivially satisfied.)

Here is a brief illustration of how the relevance-guided comprehension procedure applies to the resolution of linguistic indeterminacies such as those in (1) and (2) above. Consider the following dialogue, in which Mary's utterance 'John is a soldier' corresponds to (2e):

(3) *Peter:* Can we trust John to do as we tell him and defend the interests of the Linguistics Department in the University Council?

Mary: John is a soldier!

Peter's mentally represented concept of a soldier includes many attributes (e.g. patriotism, sense of duty, discipline) which are all activated to some extent by Mary's use of the word 'soldier'. However, they are not all activated to the same degree. Certain attributes also receive some activation from the context (and in particular from Peter's immediately preceding allusions to trust, doing as one is told, and defending interests), and these become the most accessible ones. These differences in accessibility of the various attributes of 'soldier' create corresponding differences in the accessibility of various possible implications of Mary's utterance, as shown in (4):

(4) (a) John is devoted to his duty

(b) John willingly follows orders

(c) John does not question authority

(d) John identifies with the goals of his team

(e) John is a patriot

(f) John earns a soldier's pay

(g) John is a member of the military

Following the relevance-theoretic comprehension procedure, Peter considers these implications in order of accessibility, arrives at an interpretation which satisfies his expectations of relevance at (4d), and stops there. He does not even consider further possible implications such as (4e)-(4g), let alone evaluate and reject them. In particular, he does not consider (4g), i.e. the literal interpretation of Mary's utterance (contrary to what is predicted by most pragmatic accounts, e.g. Grice, 1989b, p. 34).

Now consider dialogue (5):

(5) *Peter:* What does John do for a living?

Mary: John is a soldier!

Again, Mary's use of the word 'soldier' adds some degree of activation to all the attributes of Peter's mental concept of a soldier, but in this context, the degree of activation, and the order of accessibility of the corresponding implications, may be the reverse of what we found in (3): that is, (g) may now be the most accessible implication and (a) the least accessible one. Again following the relevance-theoretic comprehension procedure, Peter now accesses implications (g) and (f) and, with his expectations of relevance satisfied, stops there. Thus, by applying exactly the same comprehension procedure (i.e. following a path of least effort and stopping when his expectations of relevance are satisfied), Peter

arrives in the one case at a metaphorical interpretation, and in the other at a literal one. (For interesting experimental evidence on depth of processing in lexical comprehension, see Sanford, this volume. For a fuller relevance-theoretic account of lexical comprehension, and in particular of the relation between literal, loose and metaphorical uses, see Sperber and Wilson, 1998; Wilson and Sperber, 2000.)

6. Conclusion

We have considered two possibilities. First, comprehension might be an application of a general mind-reading module to the problem of identifying the speaker's meaning (a neo-Gricean view). Second, it might involve a sub-module of the mind-reading module, an automatic application of a relevance-based procedure to ostensive stimuli, and in particular to linguistic utterances. We have argued that, given the particular nature and difficulty of the task, the general mind-reading hypothesis is implausible. We have also argued that the tendency of humans to seek relevance, and the exploitation of this tendency in communication, provide the justification for a dedicated comprehension procedure. This procedure, although simple to use, is neither trivial nor easy to discover. So how can it be that people, including young children, spontaneously use it in communication and comprehension, and expect their audience to use it as a matter of course? Our suggestion has been that relevance-guided inferential comprehension of ostensive stimuli is a human adaptation, an evolved sub-module of the human mind-reading ability.

References

Astington, J., Harris, P. and Olson, D. (eds) 1988: *Developing Theories of Mind*. Cambridge: Cambridge University Press.

Bach, K. 1994: Conversational impliciture. *Mind and Language*, 9, 124-62.

Bach, K. and Harnish, R. M. 1979: *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

Barkow, J., Cosmides, L., and Tooby, J. 1995*: The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press.

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

Carruthers, P. and Smith, P. (eds) 1996: *Theories of Theories of Mind.* Cambridge: Cambridge University Press.

Carston, R. 1997: Informativeness, relevance and scalar implicature. In R. Carston and S. Uchida (eds) *Relevance Theory: Applications and Implications.* Amsterdam: John Benjamins, 179-236.

Carston, R. 2000: Explicature and semantics. *UCL Working Papers in Linguistics*, 12, 1-44. To appear in S. Davis and B. Gillon (eds) *Semantics: A Reader*. Oxford: Oxford University Press.

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

Davies, M. and Stone, T. (eds) 1995a: *Mental Simulation: Philosophical and Psychological Essays.* Oxford: Blackwell.

Davies, M. and Stone, T. (eds) 1995b: *Folk Psychology.* Oxford: Blackwell

Fodor, J. 1983: *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fridlund, A. 1994: *Human Facial Expression: An Evolutionary View*. New York: Academic Press.

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

Gazdar, G. 1979: *Pragmatics: Implicature, Presupposition and Logical Form.* New York: Academic Press.

Gernsbacher, M. 1995: *Handbook of Psycholinguistics.* New York: Academic Press.

Gigerenzer, G., Todd, P.M., and the ABC Research Group 1999: *Simple Heuristics that Make us Smart*. Oxford: Oxford University Press.

Grice, H. P. 1957: Meaning. *Philosophical Review*, 66, 377-388. Reprinted in Grice, 1989b.

Grice, H. P. 1969: Utterer's meaning and intentions. *Philosophical Review*, 78, 147-177. Reprinted in Grice, 1989b.

Grice, H. P. 1982: Meaning revisited. In N. V. Smith (ed.), *Mutual Knowledge.* London: Academic Press. Reprinted in Grice, 1989b.

Grice, H. P. 1989a: Retrospective Epilogue. In Grice, 1989b.

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

Happé, F. 1993: Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48, 101-19.

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

Hirschfeld, L. and Gelman, S. 1994: *Mapping the Mind: Domain Specificity in Cognition and Culture.* Cambridge: Cambridge University Press.

Kaplan, D. 1989: Demonstratives. In J. Almog, J. Perry and H. Wettstein (eds), *Themes from Kaplan*. Oxford: Oxford University Press.

Lascarides, A. and Asher, N. 1993: Temporal interpretation, discourse relations and common-sense entailment. *Linguistics and Philosophy*, 16, 437-493.

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

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

Lewis, D. 1970: General semantics. *Synthese,* 22, 18-67. Reprinted in D. Lewis, 1983, *Philosophical Papers.* Oxford: Oxford University Press.

Lewis, D. 1979: Scorekeeping in a language game. *Journal of Philosophical Logic,* 8, 339-59. Reprinted in D. Lewis, 1983, *Philosophical Papers.* Oxford: Oxford University Press.

Millikan, R. 1984: *Language, Thought and Other Biological Categories.* Cambridge, MA: MIT Press.

Millikan, R. 1998: Language conventions made simple. *Journal of Philosophy,* XCV, 161-180.

Mitchell, P., Robinson, E. and Thompson, D. 1999: Children's understanding that utterances emanate from minds: Using speaker belief to aid interpretation. *Cognition*, 72, 45-66.

Origgi, G. and Sperber, D. 2000: Evolution, communication and the proper function of language. In P. Carruthers and A. Chamberlain (eds) *Evolution and the Human Mind: Language, Modularity and Social Cognition*. Cambridge: Cambridge University Press, 140-169.

Perner, J. Frith, U., Leslie, A. and Leekam, S. 1989: Explorations of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, 60, 689-700.

Predelli, S. 1998: Utterance, interpretation and the logic of indexicals. *Mind and Language*, 13, 400-414.

Searle, J. 1969: *Speech Acts*. Cambridge: Cambridge University Press.

Sigman, M. and Kasari, C. 1995: Joint attention across contexts in normal and autistic children. In C. Moore and P. Dunham (eds) *Joint Attention: Its Origins and Role in Development.* Hillsdale, N.J.: Lawrence Erlbaum.

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

Sperber, D. 2000: Metarepresentations in an evolutionary perspective. In D. Sperber (ed.) *Metarepresentations: An Interdisciplinary Perspective*. New York: Oxford University Press.

Sperber, D. forthcoming: In defense of massive modularity. In E. Dupoux (ed.) *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler.* Cambridge, MA: MIT Press.

Sperber, D. and Wilson, D. 1986/95: *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sperber, D. and Wilson, D. 1998: The mapping between the mental and the public lexicon. In P. Carruthers and J. Boucher (eds) *Language and thought.* Cambridge: Cambridge University Press, 184-200.

Wharton, T. 2001: Natural pragmatics and natural codes. *UCL Working Papers in Linguistics*, 13, 109-158.

Wilson, D. 1998: Linguistic structure and inferential communication. In B. Caron (ed.) *Proceedings of the 16th International Congress of Linguists* (Paris, 20-25 July 1997). Oxford: Elsevier Sciences.

Wilson, D. 2000: Metarepresentation in linguistic communication. In D. Sperber (ed.) *Metarepresentations: An Interdisciplinary Perspective.* New York: Oxford University Press.

Wilson, D. and Sperber, D. 1986: Pragmatics and modularity. In *Chicago Linguistic Society Parasession on Pragmatics and Grammatical Theory* 22: 67-84. Reprinted in S. Davis (ed.) 1991: *Pragmatics: A Reader*. Oxford: Oxford University Press.

Wilson, D. and Sperber, D. 2000: Truthfulness and relevance. *UCL Working papers in linguistics,* 12, 215-254.

Wilson, D. and Sperber, D. forthcoming: Relevance theory. To appear in G. Ward and L. Horn (eds) *Handbook of Pragmatics*. Oxford: Blackwell.