

Relevance Theory – New Directions and Developments

Robyn Carston and George Powell

1. Introduction – an outline of Relevance Theory

As a post-Gricean pragmatic theory, Relevance Theory (RT) takes as its starting point the question of how hearers bridge the gap between sentence meaning and speaker meaning. That there is such a gap has been a given of linguistic philosophy since Grice's (1967) *Logic and Conversation*. But the account that relevance theory offers of how this gap is bridged, although originating as a development of Grice's co-operative principle and conversational maxims, differs from other broadly Gricean accounts in certain fundamental respects, and leads to a stance on the nature of language, meaning and communication which is at odds, not only with the view of Grice himself, but also with the view common to most post-Fregean philosophy of language.

Relevance theory grounds its account of utterance interpretation within a general claim about cognitive design, the claim that human cognition is geared towards the maximisation of *relevance*. For Sperber and Wilson (1986/95a), relevance is a potential property of inputs to cognitive processes. Any input may deliver a variety of different types of cognitive effect; it may, for instance, combine inferentially with existing assumptions to yield new conclusions (known as *contextual implications*), or it may provide evidence that strengthens existing beliefs, or it may contradict and eliminate already held information. At the same time, getting at the effects of a particular input demands processing effort. For Sperber and Wilson, relevance is, roughly speaking, a trade-off between cognitive effects and processing effort: the greater the ratio of effects to effort the greater the relevance of an input¹. Given this notion of relevance, to claim that humans are geared towards maximising relevance is to claim that we are designed to look for as many cognitive effects as possible for as little processing effort as possible. The idea is that, as a result of constant selection pressure towards increasing

¹ Sperber and Wilson distinguish between those effects which are beneficial to a cognitive agent (*positive cognitive effects*) and those which are not. Talking a little less loosely, therefore, relevance is a trade off between positive cognitive effects and processing effort. See Sperber and Wilson (1995b) and Wilson and Sperber (2004).

cognitive efficiency, we have evolved procedures to pick out potentially relevant inputs and to process them in the most cost-effective way (Sperber and Wilson 1995b).

All communication, and linguistic communication in particular, makes use of this cognitive drive for relevance. Taking the case of linguistic communication, Sperber and Wilson's claim is that an utterance raises quite specific expectations of relevance in its addressee, that is expectations about the effects it will yield and the mental effort it will cost. Quite generally, an utterance comes with a presumption of its own optimal relevance; that is, there is an implicit guarantee that the utterance is the most relevant one the speaker could have produced, given her abilities and her preferences, and that it is at least relevant enough to be worth processing. That utterances carry this presumption motivates a particular comprehension procedure, which, in successful communication, reduces the number of possible interpretations to one: in essence, it licenses a hearer to consider possible interpretations in order of their accessibility (that is, to follow a path of least effort) and to stop as soon as he reaches one that satisfies his expectation of relevance.

Sperber and Wilson thus posit a powerful cognitively-grounded machinery for the interpretation of utterances. Recent research within the relevance-theoretic framework suggests that the implications of such a machinery are far-reaching. As we discuss in section 2, this approach to utterance interpretation supports a view of language and meaning which differs fundamentally from that common in the contemporary philosophy of language. Furthermore, it raises important questions about the relationship between communication and mental architecture, which we outline in section 3.

Much work in relevance theory relies on the kinds of method and data familiar to linguistic philosophers: essentially introspection and native speaker intuitions on properties such as truth conditions, truth values, what is said, etc. Recently, however, relevance theorists have been at the forefront of a newly-emerging research field, experimental pragmatics, which aims to apply the empirical techniques of psycholinguistics to questions about utterance interpretation. Over the last few years, this new research methodology has thrown up interesting and sometimes surprising insights into the psychological processes underlying human communication and comprehension, some of which we discuss in section 4.

2. Relevance theory and the semantics-pragmatics distinction

Where should the line be drawn between semantics and pragmatics? On one familiar view, endorsed both by Grice and by most contemporary philosophers of language, the outline answer is clear: semantics is concerned with what is variously called *the proposition (semantically/literally) expressed, what is said* or simply *the truth-conditional content* of an utterance, whereas pragmatics is concerned with the implicatures of an utterance. There are, however, respects in which, for all but the most hard-line truth-conditionalist, context contributes to propositional content. Grice, for instance, accepted that what is said by an utterance is determined, not only by the ‘conventional meaning’ of the sentence uttered, but also by disambiguation and assignment of values to indexical expressions (Grice 1989: 25).

However, there has been much recent work within the relevance-theoretic framework arguing for the view that pragmatic contributions to propositional content go a great deal further than disambiguation and reference assignment. There are two key strands to this work: on the one hand, there has been research into lexical pragmatics which broadly defends the position that, not only those lexical items traditionally taken to be indexical, but more or less all lexical meaning is context-sensitive; on the other hand, there has been research into proposition-level context-sensitivity which supports the view that some pragmatic aspects of propositional content may not correspond to items present at any level of syntactic representation. This leads to a reassessment of the appropriate way to draw the distinction between semantics and pragmatics.

2.1 Relevance theory and lexical pragmatics

While indexicals, such as ‘she’, ‘those’, ‘here’, clearly require the pragmatic supplying of a contextual value, it might seem that what are often known as *content words*, - nouns, verbs, adjectives, etc. - come with a fully specified, context-invariant, conceptual content as a matter of their lexically encoded meaning. However, according to the relevance-theoretic view, such words, although linguistically unambiguous, may communicate a range of distinct (though related) meanings in different contexts. Consider the following examples (adapted from Searle 1983:145):

1. a. Pat *opened* the curtains.

- b. Bill *opened* his mouth.
- c. Sally *opened* her book to page 56.
- d. The child *opened* the package.
- e. The carpenter *opened* the wall.
- f. The surgeon *opened* the wound.

Although the lexically encoded meaning of the word ‘open’ is the same in these examples, it is understood differently in each case. As Searle points out, the contribution it makes to the truth conditions of quite literal utterances varies with the sentential context it occurs in. What constitutes opening a book is very different from what constitutes opening one’s mouth, which is quite different again from what constitutes opening a package, etc. Given that the concept expressed by the use of a word may also vary with extralinguistic context (for instance, in a scenario in which a person’s broken jaw has been wired together the process of opening the mouth is rather different from the usual one), a virtually indefinite range of different concepts can be communicated by uses of the verb ‘open’. The upshot of the pragmatic process at work here is a *narrowing* of the linguistically encoded meaning to a more specific concept.

The opposite result occurs too, that is, the concept communicated by the use of a word in context may be broader than the linguistically given concept. Consider the different interpretations of the adjective ‘flat’ in the following examples (adapted from Wilson 2004: 345):

- 2. a. This ironing board is *flat*.
- b. My back garden is *flat*.
- c. He had a *flat* face and sad eyes.
- d. Holland is *flat*.
- e. The sea was *flat*.

As with narrowing, different degrees and types of *broadening* (or *loosening*) are appropriate in different circumstances, so that, for instance, the departure from true flatness is greater in the case of a whole country than in the case of one’s back garden, and the flatness of a landscape is different in kind from the flatness of a face. Another variety of broadening involves what is often called ‘category extension’ and is typified by the use of salient brand

names or person names (e.g. ‘Hoover’, ‘Kleenex’, ‘Hitler’, ‘Chomsky’) to denote a broader category (vacuum cleaners in general, the class of megalomaniac leaders with inhuman policies, etc.). In some cases, the communicated meaning may involve both an element of narrowing and an element of broadening; for instance, consider a depressed woman who says of her irresponsible husband ‘Ken’s a bachelor’. The concept she communicates is both narrower than the encoded meaning of the word ‘bachelor’ since it is confined to the stereotype of an easy-going, promiscuous kind of bachelor, but it is clearly also broader since it includes in its denotation certain married men whose behaviour is like that of the stereotypic bachelor (Carston 2002: section 5.2).

Two important and distinctive characteristics of the RT approach to these phenomena are: (a) the claim that the pragmatic process involved is not a matter of implicature derivation but rather of conceptual adjustment which contributes to the proposition explicitly communicated (the truth-conditional content of the utterance), and (b) while most other pragmatic approaches assume that narrowing and broadening are to be treated as distinct processes, the RT view is that they are simply different possible outcomes of a single pragmatic process which fine-tunes the interpretation of virtually every word. The model of lexical semantics that we assume essentially follows Fodor (1998): lexical forms map to mentally-represented concepts, that is, elements of a conceptual representation system or ‘language of thought’ (leaving aside indexicals to which this does not obviously apply). These concepts constitute the meanings of words as expression-types in a linguistic system, so that words can be thought of as inheriting the denotational semantics of the Mentalese concept that they encode. So, when the outcome of the lexical pragmatic process of meaning adjustment is a narrowing, the denotation of the concept communicated by the use of a word is a proper subpart of the denotation of the lexically encoded concept, and when the outcome is a broadening, the opposite relation between the denotations of encoded concept and communicated concept holds. When the adjustment involves both outcomes, the relation between the denotations of encoded concept and communicated concept is one of mere overlap.

How does this single unified account work? Recall the relevance-based comprehension procedure mentioned in the introduction, according to which an addressee follows a track of least effort in trying out interpretations, stopping once he has one that meets his expectation of relevance (that is, of sufficient effects for no gratuitous processing effort). This procedure is automatically applied to the on-line processing of attended verbal utterances: taking the

schematic decoded linguistic meaning as input, processes of pragmatic enrichment at the explicit level occur in parallel with the derivation of the implications of the utterance. Central to the working of the procedure is a subprocess of ‘mutual adjustment’ of explicit content, contextual assumptions and contextual implications, a process guided and constrained by expectations of relevance. Here is a brief example involving the adjustment of explicit content in response to expected implications and where the outcome is a narrowing of a lexically encoded meaning:

3. A (to B): Be careful. The path is *uneven*.

Given that the first part of A’s utterance warns B to take care, B is very likely to expect the second part of the utterance to achieve relevance by explaining or elaborating on why, or in what way, he should take care. Now, virtually every path is, strictly speaking, uneven to some degree or other (i.e. not perfectly plane), but given that B is looking for a particular kind of implication, he will enrich the very general encoded concept UNEVEN so that the proposition explicitly communicated provides appropriate inferential warrant for such implications of the utterance as: B might trip over, B should take small steps, B should keep his eye on the path, etc. The result is a concept, which we can label UNEVEN*, whose denotation is a proper subset of the denotation of the lexical concept UNEVEN. (For much more detailed exemplification of the RT-based account of lexical adjustment, resulting in concept broadening, or narrowing, or a combination of the two, see Wilson and Sperber (2002).)

Finally, a distinctive RT claim in this context is that metaphorical and hyperbolic uses of words involve a kind of concept broadening (or loose use), so fall within this single process of lexical meaning adjustment. For instance, an utterance of the sentence in (4) could be taken as an ordinary broadening (if, say, it’s known that a particular run, referred to by ‘it’, was a little less than 26 miles) or as hyperbolic (if it was considerably less than the length of a marathon) or as metaphorical for a long, arduous, exhausting experience, whether physical or mental.

4. It was a *marathon*.

The idea is that there is no hard and fast distinction between these different degrees of loosening of the lexical concept MARATHON; rather, there is a continuum of cases from

ordinary approximations through to the more radical broadening involved in comprehending metaphors. (See Wilson 2004, Carston and Wilson forthcoming/2006.)

2.2 Relevance theory and reference

The primary domain of lexical pragmatics within RT has been the interpretation of predicate expressions. However, over recent years there has also been a certain amount of research from an RT perspective into the semantics and pragmatics of singular expressions (proper names, indexicals and demonstratives, both simple and complex) as well as definite descriptions.

While there is something very close to consensus among linguistic philosophers on the context-sensitivity of indexical and demonstrative expressions, there is much less agreement on how best to treat definite descriptions and proper names. As is well-known, the key question on definite descriptions over the last forty years has been how to analyse what Donnellan (1966) calls the referential-attributive distinction, i.e. how best to accommodate the apparent datum that a definite description ‘the F’ may be used either to talk about a particular antecedently-identified individual or to talk about *whatever* happens to be uniquely F. As regards proper names, there are broadly three positions: those who consider names to be the natural language equivalent of logical individual constants, those who take them to be descriptive, i.e. to contribute properties to truth conditions, and those who see them as closely related to indexicals.

Donnellan’s referential-attributive distinction has proved notoriously divisive within the philosophy of language. On the one hand, there seem to be good reasons to suppose that the distinction corresponds to a truth-conditional difference: the truth conditions of an utterance of a definite description sentence appear to alter according to whether the description is used referentially or attributively. On the other hand, there also seem to be good reasons to believe that definite descriptions are not ambiguous: although ‘the man drinking a martini’ may be used either referentially or attributively, it does not seem to be ambiguous in the way that, say, ‘bank’ or ‘coach’ is. But, of course, these two observations are hard to reconcile on standard philosophical assumptions: if (leaving aside indexical expressions) you identify the meaning of an expression with the contribution that expression makes to truth conditions, then it follows directly that an expression which is capable of making two different kinds of contribution to truth conditions has two different meanings, i.e. is ambiguous.

However, as a number of researchers working within the relevance-theoretic framework have pointed out (e.g. Rouchota 1992, Bezuidenhout 1997, Powell 2001), RT offers a natural way to reconcile these data. As discussed in the previous section, there is a key distinction drawn in Relevance Theory between, on the one hand, the linguistically encoded meaning of a particular expression and, on the other hand, the contribution that expression makes to truth-conditional content on an occasion of use. Given this distinction, the fact that a particular expression may make two distinct types of contribution to truth-conditional content is no evidence for its ambiguity at the level of linguistically encoded meaning. While differing in detail, all the above-mentioned RT accounts take the following position: that definite descriptions are linguistically univocal but truth-conditionally ambiguous. The gap between the encoded meaning of a definite description and what that description contributes to propositional content in a particular context is bridged by relevance-guided pragmatic inference.

Just as with definite descriptions, proper names have thrown up some notoriously thorny philosophical questions, of which the most attention has probably gone to Frege's puzzle on the informativeness of identity statements (for instance, 'Marilyn Monroe is Norma Jean Baker', 'Evan Hunter was Ed McBain'). There has been less work from a relevance-theoretic perspective on proper names than on definite descriptions, but Powell (1998, 2003) has addressed questions about proper names as part of a general RT-flavoured analysis of the semantics and pragmatics of singular expressions. On this analysis, all such expressions are profoundly context-sensitive: whether they make referential or descriptive contributions to truth conditions is not a matter of the encoded meanings of these expressions, but is rather a matter of broad context and pragmatic principles. Powell (2003) analyses the encoded meanings of singular expressions (including here definite descriptions) not in terms of their contribution to truth conditions, but rather in terms of their contribution to a hearer's mental representations. All these expressions, on this view, are marked as *individual concept communicators* by virtue of their linguistically encoded meaning. That is to say, they are marked as contributing to a hearer's mental representation a concept which, roughly speaking, is taken to be satisfied by a unique individual. Beyond that, the encoded meaning of these expressions is silent as to whether this concept should be *de re* (i.e. referential) or descriptive. Which constraints a particular singular expression lays on the concepts which may serve as its interpretation will vary according to the type of singular expression. In the case of a proper

name 'N', the constraint on interpretation is simply that the individual concept should be of a bearer of 'N'. Which concept that is on a particular occasion will be determined by context and pragmatic inference. A definite description 'the F', on the other hand, encodes a rather more complex condition: it constrains interpretation to an individual concept of a unique F in a salient context. Again, which is the salient context and which the intended individual concept (and whether it is referential or descriptive) on any given occasion is determined pragmatically.

On this analysis, traditional philosophical puzzles with proper names, such as the informativeness of identity statements, disappear. Consider an utterance of:

5. Evan Hunter was Ed McBain

The familiar problem is how such statements, which seem merely to predicate the identity of an individual with itself, can nevertheless be informative. On Powell's analysis, a hearer faced with the task of interpreting an utterance of (5) will access two individual concepts, one associated with the name 'Evan Hunter' and the other with the name 'Ed McBain'. So long as these two concepts are appropriately referentially anchored, they will be anchored to the same individual, since the names 'Evan Hunter' and 'Ed McBain' share a bearer. This does not, however, imply that the concepts share informational content. A person's 'Evan Hunter' concept might contain information such as *x is the author of 'The Blackboard Jungle'* while her 'Ed McBain' concept might contain information such as *x wrote the 87th Precinct novels*. Since these two concepts are associated with different information, when this person comprehends (5) she thereby gains access to new information, for instance, the information that the author of 'The Blackboard Jungle' also wrote the 87th Precinct novels. On this analysis, therefore, it is predicted that identity statements involving co-referring names are capable of being informative.²

2.3 Relevance theory and unarticulated constituents

² The informativeness puzzle is not limited to uses of different co-referring names; it can also arise for identity statements involving different uses of the same name, as demonstrated by Kripke's Paderewski example (Kripke 1979). It is interesting to note, however, that such cases only give rise to the puzzle in those instances where the one name is associated with two distinct individual concepts.

As discussed in the last two sections, recent research within relevance theory has supported the view that linguistic expressions of all sorts display profound context-sensitivity. But might context-sensitivity go even beyond this? Much attention has recently been paid to what Perry (1986) dubbed *unarticulated constituents*. The idea behind Perry's notion is that the proposition expressed by an utterance may contain constituents which do not correspond to anything in the syntax of the sentence uttered. Consider, for instance, the much-discussed sentence in (6):

6. It's raining

It seems that if I utter (6) in London then I have said that it is raining in London, whereas if I utter (6) in Paris I have said that it is raining in Paris; it seems, in other words, as if any utterance of (6) will be true iff it is raining *at a particular location* (and, in fact, the particular location is not always the place of utterance). Yet there is no constituent, at least in the overt syntax of (6), which corresponds to this location parameter.

How should one best account for this sort of datum? There are, broadly speaking, three types of response currently on the market. At one end of the spectrum are those who argue that features such as this location parameter form no part of the truth-conditional content of utterances. Cappelen and Lepore (2005), for instance, claim that the belief that such elements contribute to literal propositional content results from a confusion on the distinction between the proposition semantically expressed by an utterance and the speech acts the utterance is used to perform. On the middle path are those who accept that features such as the location parameter in (6) do indeed contribute to literal propositional content, but who argue that they must, therefore, be represented covertly in the syntax (e.g. Stanley 2000). Recent research within the relevance-theoretic framework, however, has defended a third position, that at least some of these features do contribute to truth-conditional content, while at the same time being genuinely unarticulated, i.e. unrepresented in the syntax at any level of representation. Carston (2004a) defends a position on which the retrieval of such constituents is a purely pragmatic process, the result of relevance-guided inference.

The argument is essentially two-fold. First, any theory of content which aims to play a serious role in a wider theory of interpretation and communication, i.e. which lays claim to any degree of psychological plausibility, must be answerable to native speaker intuitions on such

matters as truth conditions, truth values and what is said. This is taken to exclude positions such as that adopted by Cappelen and Lepore, on which the kind of minimalist propositions (indexical values being the only contextually-provided elements) which are taken to constitute truth-conditional content will be generally inaccessible to intuition. Once one takes intuition seriously, there seems little way of avoiding the conclusion that the sorts of constituent under discussion do genuinely contribute to truth-conditional content. Second, according to relevance theorists, any attempt to tread the middle path leads to very problematic results. Stanley (2000) finds support for his middle position from evidence that the postulated covert indexical elements, such as a location parameter in (6), can, like overt pronouns, enter into binding relations. Carston's (2004a) rejoinder starts from an argument first presented by Wilson and Sperber (2002), in which they pointed out that there is no principled limit to the number of covert elements that such a theory would have to posit. An utterance of (7), for instance, might express a proposition with a range of constituents corresponding to what is eaten, the time, place, manner of eating, and so on.

7. I've eaten

On Stanley's analysis, each of these would have to correspond to a variable or indexical at LF (linguistic logical form), a theoretical prediction which Wilson and Sperber take to be a *reductio* of Stanley's position. Carston develops this one step further by showing that, although all of these hidden indexical elements would have to be present at LF, there would be many instances on which some of these elements would receive no value. Consider an utterance of (7) in response to the question 'Would you like some dinner?'. While what was eaten and the time of eating might well be relevant (that the speaker has eaten a full meal and the eating took place in the recent past), the place and manner of eating would surely not be. It nevertheless seems that such an utterance would express a determinate proposition. This is not the sort of thing we expect of indexicals. Consider a standard use of an overt pronoun in an utterance of (8):

8. She put the book on the table.

If 'she' does not receive a value in context, then clearly (8) does not express a complete proposition. So Stanley's hidden indexicals are, at the least, a very different kind of thing

from the sort of indexicals we are used to³.

If relevance theorists are right, and (some of) these constituents do contribute to truth-conditional content while being genuinely unarticulated, this is another serious blow for hopes of building a truth-conditional theory of linguistic meaning, since it yet further breaks the link between sentence meaning and truth-conditional content. Rather, it fits with a view of the relation between encoded meaning and propositional content on which sentences encode not propositions but something more like propositional schemas or templates. These, then, must be pragmatically fleshed out in a context in order for the explicit content of the utterance to be recovered or, in many cases, in order that anything even minimally truth evaluable can be retrieved.

2.4 Conclusion

Where does the research discussed above leave us on the semantics-pragmatics distinction? There are two key elements to the relevance-theoretic view. First, relevance theorists take a view on pragmatic contributions to truth-conditional content which is fundamentally at odds with traditional views of the semantics-pragmatics distinction. On the RT view, there is, in principle, no limit to the effects of contextual information on propositional content: not only are all expressions context-sensitive, but context may also add constituents to propositional content which are entirely unrepresented in the syntax. Second, on the RT view, the processes which bridge the gap between linguistically encoded meaning and explicitly communicated meaning and those that bridge the gap between explicitly and implicitly communicated meaning (in Gricean terms, between what is said and what is implicated) are aspects of a single inferential process (seeking the optimally relevant interpretation); they occur in parallel and are subject to a mechanism of mutual adjustment as discussed in section 2.1. This distinguishes relevance theorists from others who, while broadly sympathetic to the strong contextualist stance of RT, take some version of a multi-phase view of pragmatics on which the processes that mediate linguistic meaning and explicit content may be different in kind from those responsible for implicatures (see, for instance, Asher 1999; Levinson 2000; Recanati 2002b, 2004).

³ For further arguments against the hidden indexical view, see Carston (2004b) and, from a different (non-RT) contextualist perspective, Recanati (2002a). Stanley's account has also come under some sustained fire from the opposite end of the spectrum, most recently from Cappelen and Lepore (2005).

Finally, given that the proposition explicitly expressed by an utterance is replete with pragmatically supplied content, only some of which is linguistically mandated, the rest being entirely pragmatically motivated, it is clearly not possible to draw a semantics-pragmatics distinction that coincides with the distinction between explicit utterance content and implicatures. According to the RT approach, the distinction has to be drawn between context-free linguistic expression-type meaning and what is communicated. That this is the only coherent way in which to draw the distinction is argued in more detail in Carston (forthcoming) (but see *THE DISTINCTION BETWEEN SEMANTICS AND PRAGMATICS* in this volume for a different view). It follows from this position that the long-standing Principles of Semantic Compositionality and Semantic Innocence hold, not at the level of the truth-conditional content of an utterance, but at the more schematic (often nonpropositional) level of linguistic expression-type meaning (see Powell 2002).

3. Relevance theory and mental architecture

While inferential pragmatics has its origins in the philosophy of language, the relevance-theoretic approach, on which it is construed as a mental processing system responsible for interpreting a kind of human behaviour (verbal and other ostensive communicative acts), sets it squarely within cognitive science. The result is an account which, while still very much concerned with the issue of the right distribution of labour between semantics and pragmatics in accounting for speaker meaning, is embedded in wider issues about human cognition. One of these is the question of cognitive architecture and the location of pragmatics within it: what sort of a system is responsible for pragmatic processing? is it task-dedicated or a more general problem-solving system? what other cognitive systems does it interact with? The account is receptive to considerations from evolutionary psychology concerning the kinds of cognitive systems that have been naturally selected to solve particular adaptive problems (is pragmatics one of these?). And it must answer to experimental findings about the nature and time-course of utterance comprehension; for example, results concerning which elements of conceptual information are activated at which points in the processing of ambiguous words, or metaphorical uses, or cases of implicature. Some of the ways in which relevance theory and such empirical work have begun to mutually inform each other are considered in the next section.

Probably the single most influential position on human cognitive architecture is that of Jerry Fodor (1983, 2000). On his view, the mind has a hybrid architecture: perceptual input systems, including language perception, and motor output systems are autonomous mental modules, while the central systems responsible for forming beliefs and making decisions are nonmodular. The processes of pragmatic inference are clearly a function of central systems: their goal is the fixing of a belief about a speaker's meaning (the content of her communicative intention) and they are highly context-sensitive. So the conclusion has to be that, while the phase of linguistically decoding an utterance may be carried out by a fast, automatic, informationally encapsulated system (a module), the inferential phase which bridges the gap between linguistic meaning and speaker meaning is nonmodular.

However, in recent years, there has been a shift towards a more modular (indeed a massively modular) view of the mind, albeit with a degree of relaxation of Fodor's criteria for what constitutes a modular system. This is largely a result of bringing evolutionary considerations to bear on hypotheses about the nature of mental architecture (see, for instance, Barkow, Cosmides and Tooby 1995, Sperber 1994b, Carruthers and Chamberlain 2000). Natural selection favours specific solutions for specific problems. A cognitive procedure dedicated to dealing with a particular recurrent environmental problem is very likely to outperform a more general process applied to the same problem because the computations of the more general process must effect a compromise in order to deal with several distinct types of problem. Thus an evolutionary perspective suggests the increasing and refining of mental modularity rather than any kind of merging into more general systems and the crucial property of a modular system on this view is that it is a special-purpose mechanism attuned to regularities within a particular problem domain.

If the Fodorian central interpretive systems are to be reconstrued as consisting of such modular mechanisms the issue becomes whether or not pragmatic processes are executed by such a system. Currently there are two main positions on this question. One is that attributing a meaning (a special kind of intention) to a speaker falls within our broader capacity to attribute intentions, beliefs and other mental states to each other on the basis of any kind of purposive behaviour, whether communicative or non-communicative, (variously known as a mind-reading ability or 'theory of mind') (see, for example, Bloom 2002). The current relevance-theoretic view, however, is that pragmatic processes are carried out by a dedicated, domain-specific comprehension module with its own special principles and

procedures. In support of this position, Sperber and Wilson (2002) point out some telling differences between general mind-reading and utterance interpretation.

First, while both of these kinds of mental state attribution clearly involve a metarepresentational capacity (the capacity to represent the mental representations of others), there is an important difference in the complexity of its application in the two cases. Ostensive communication involves an informative intention embedded in a communicative intention, so that a hearer has to recognise that the speaker *intends* him to *believe* that she *intends* him to *believe* a certain set of propositions (Sperber 1994a, 2000), that is, four levels of metarepresentation, while in understanding ordinary actions a single level of intention attribution is usually sufficient. Furthermore, there is an interesting disparity here in the abilities of three-year-old children, many of whom are quite competent linguistic communicators while nevertheless failing standard *false belief tasks* that require them to attribute only a single-level epistemic state to an agent (see, for instance, Baron-Cohen 1995, Scholl and Leslie 1999). This dissociation of capacities is difficult to explain if the attribution of a meaning to a speaker is simply a function of a general capacity to attribute intentional states.

Second, while the range of intentions that can be reasonably attributed to an agent on the basis of some noncommunicative behaviour in a particular situation (e.g. extending an arm into a cupboard, walking up a flight of stairs) is generally quite limited, the physical setting of an utterance places few restrictions on its content and, given the gap between linguistic meaning and speaker meaning, there is a vast range of possible meanings that a speaker could be communicating. The standard procedure for recognising and attributing an intention to someone on the basis of purposive behaviour (e.g. an intention to retrieve a bowl from the cupboard, an intention to reach the philosophy department on the third floor) involves observing the various effects of the behaviour, or consulting one's memory about the usual results of such a behaviour, and taking it that the desirable and predictable effects are the intended ones. This strategy would very seldom come up with the right result if applied to communicative behaviour because the desired effect just is the recognition of the communicator's intention: 'hearers cannot *first* identify a desirable effect of the utterance and *then* infer that the speaker's intention was precisely to achieve this effect' (Wilson 2003: 116). The claim, then, is that the strategy pursued in figuring out what a speaker means by her communicative behaviour is the one given in the introduction: the comprehension process

follows a path of least effort in accessing interpretations (at both the explicit and implicit levels) and it stops when the specific expectations of relevance raised by the particular utterance are satisfied. What underpins this strategy is the presumption of optimal relevance that accompanies all acts of ostensive communication and which is absent from other kinds of intentional behaviour. Thus distinct procedures are followed in the two kinds of intention-attribution.⁴ In fact, the RT view is that pragmatics is one of a cluster of modules that make up what could be broadly thought of as our social cognitive capacity.

Finally, suppose that it is true that pragmatics is a fast, automatic system with its own idiosyncratic relevance-based procedure for solving its own specific problem, the next question is what is the domain of this module? Given what has been said so far, one might think it is acts of linguistic communication (verbal utterances). However, the RT view is that the domain of pragmatics is quite a lot wider than this – it is *ostensive stimuli* and these comprise any and all human actions which come with a particular complex kind of intention, an informative intention embedded in a communicative intention, including, for instance, acts of mimicry and other bodily gestures whose primary purpose is communicative. At this stage, it might seem that the domain of the module corresponds closely with what Grice (1957) called cases of non-natural meaning (as opposed to natural meaning), but again RT departs somewhat from Grice. Wharton (2003) has pointed out that instances of natural human behaviours, such as spontaneous expressions of emotion - facial expressions, affective tone of voice - can be used by communicators as (or as components of) ostensive stimuli. For instance, a communicator conveying some positive news may openly let her audience see her spontaneous smile, or, in a different situation, may use a particular tone of voice which will calibrate the degree of anger her audience takes her to be conveying (both of these to be distinguished from the faking of a natural behaviour as a means of communication). As Wharton says, an ostensive stimulus is often a composite of verbal behaviour (non-natural) and natural behaviour, both of which provide rich clues to the addressee in recovering the speaker's meaning.

⁴ This aspect of the RT view stands in clear contrast with the position taken by Donald Davidson, according to which the interpretation of any rational intentional action, whether communicative or non-communicative, is a thoroughly non-modular, holistic matter of attributing to the agent those beliefs and desires that make best sense of his overall life and conduct (Davidson 1973, 1986).

4. Relevance theory and experimental pragmatics

So far we've been looking at the content of current theoretical research within the RT framework. In this section we turn to recent developments in research methodologies. The methods used by RT-oriented researchers have standardly been those employed by most philosophers of language: introspection, intuition, analysis and argument. Recently, however, a growing number of researchers have been approaching questions on the nature of pragmatic processing via experimental techniques familiar to psycholinguists, but less familiar to theoretical pragmatists.

Within the RT framework, a number of researchers have recently applied such techniques to questions surrounding so-called *scalar implicatures* (*scalar implications*, *scalar inferences* or sometimes just *scalars*). Very roughly, a scalar implicature arises when a speaker, by expressing a less informative proposition is taken to communicate the negation of a more informative proposition. Consider, for example, the following dialogue:

9. Peter: Do you like Woody Allen's films?
Jane: I like some of them.

It would seem that Jane's utterance communicates the proposition in (10):

10. Jane does not like all of Woody Allen's films.

Yet this proposition does not appear to be part of the meaning of the sentence she has uttered. After all, there is no incompatibility between the proposition explicitly expressed by Jane in (9) and the proposition in (11):

11. Jane likes all of Woody Allen's films.

Experimental techniques have been applied to a number of distinct questions about scalar inferences such as (10): whether they are genuinely pragmatic or are automatically triggered by elements in the grammar; whether they should be treated as implicatures or rather as elements of explicitly communicated content; at what stage in development children begin to draw scalar inferences; and, cross-cutting these, whether the set of phenomena which have

traditionally been treated as scalars form a coherent class, or whether different scalars should receive different types of analysis. Here we will focus on the first of these questions, since it has received particular attention in the recent literature.

Broadly speaking, there are two distinct approaches to the analysis of scalars currently on the market⁵. On the one hand, there are those who take scalar inferences to be triggered by elements in the grammar. Chierchia (2004), for instance, claims that the grammar delivers two distinct entries for each scalar term, with the logically weaker being filtered out according to linguistic context. Along similar lines, Levinson (2000) takes scalar implicatures to be default inferences, hence generated whenever a scalar term is used (with possible subsequent cancellation due to contextual incompatibility). On the other hand, there are those who think that scalar implicatures are purely pragmatic, i.e. that they are generated on a case-by-case basis according to context and pragmatic principles. This latter view, advocated by, for instance, Carston (1998) and Sperber and Wilson (1995b), has become the dominant relevance-theoretic position on scalars.

A number of researchers have turned to experimental psycholinguistic techniques in an attempt to adjudicate between these two positions. Noveck and colleagues (e.g. Noveck and Posada (2003), Bott and Noveck (2004)) have used both psycholinguistic and neuropsychological methods to explore hearers' reactions to underinformative statements such as those in (12) and (13):

12. Some cows are mammals
13. Some books have pages

Subjects were asked to judge these sentences as true or false while both their response time and neural activity were measured. The key results come from subjects who judged the sentences to be false, since this judgment depends on retrieving a scalar interpretation. That is to say, in order to judge (12) false you have to take it as conveying the proposition in (14):

14. Some but not all cows are mammals

⁵ We do not mean to suggest that adherents of each position would necessarily see themselves as fighting the same corner, merely that there are important similarities for our purposes.

Noveck and colleagues found evidence that those who gave such judgments took longer to reach an interpretation than those who did not. This is taken to favour the relevance-theoretic account of scalars, on which the retrieval of scalar interpretations is an effortful case-by-case matter, over default accounts, on which scalar interpretations are automatically triggered by the grammar. The relevance-theoretic view on scalars is also supported by a series of experiments conducted by Breheny and colleagues (Breheny, Katsos and Williams (in press), Katsos, Breheny and Williams (2005)), using a range of sophisticated techniques to investigate the role of context in scalar inference.

In section 2.1 above, we discussed ways in which the concept lexically encoded by a particular content word may be adjusted during interpretation. There has recently been interesting work conducted within the RT framework aimed at examining these processes from an experimental perspective and, in particular, at adjudicating between RT accounts and others currently available. Rubio (2005) uses on-line word-recognition tasks to show patterns of conceptual priming across time. Her results give preliminary support to the analyses of concept narrowing and concept loosening developed by Carston (2002). Beyond this, Rubio examines the time-course of activation and deactivation of a range of conceptual associates, i.e. concepts related to the concept lexically encoded by the test word. Her results point to the interesting conclusion that some conceptual correlates are so closely associated to a particular content word that they will remain active during interpretation regardless of their contextual irrelevance. For instance, in the case of understanding a metaphorical use of ‘John is a cactus’, she found that the concepts PLANT, a superordinate, and DRY, a strong associate of CACTUS, remain active even after the metaphorical interpretation (for which they are irrelevant) has been recovered. This sort of finding has to be accommodated by RT and any other pragmatic theory which aims to capture the actual on-line processes of comprehension.

5. Future directions for relevance theory

As discussed in the previous section, a growing number of researchers are committed to spelling out the empirical predictions of relevance theory and subjecting them to experimental testing. These include predictions that follow from the fundamental Cognitive Principle (Human cognition tends to be geared to the maximisation of relevance) and others flowing from the Communicative Principle (Ostensive stimuli can be presumed to be optimally relevant) (see Van der Henst and Sperber 2004). More generally, the emerging field of

psychopragmatics is being energetically developed by pragmatists working in several frameworks, with RT being strongly represented among them (see Noveck and Sperber 2004).

There are two other strands of empirical work in which ideas from RT are playing an increasing role. One is research into the development of communicative competence in children and its relation to their linguistic maturation, on the one hand, and to their developing mind-reading capacity, on the other. As touched on in section 3, ostensive communication emerges earlier (from 2 years old) than the less metarepresentationally complex ability to attribute false beliefs to others (maturing at around 4 years old). Happé and Loth (2002) take this as evidence in favour of Sperber and Wilson's view of pragmatics as a modular mental system distinct from general theory of mind. However, one question that arises here is whether the child is manipulating the kind of complex layered intentions generally assumed to characterise ostensive stimuli, or whether some other earlier emerging aspect of mind-reading such as *joint attention* (arising around 12 months) is sufficient to explain early communication and comprehension (see Tomasello and Rakoczy 2003, Breheny in press). Another interesting line of thought here concerns the degree of metarepresentational complexity of different kinds of *expectations of relevance* (crucial to the functioning of the RT comprehension procedure) that children and adults may have at different stages of development. It has been suggested that a young child may assume that any utterance directed at her just is optimally relevant to her, whereas more sophisticated expectations might make allowance for a speaker's fallibilities and/or ulterior motives (see Sperber 1994a, Wilson 2000). Clearly, the naïve expectation requires no consideration of the speaker's beliefs or desires whereas the more sophisticated ones do. There is potentially fruitful work to be done in deriving explicit predictions from these ideas and testing them on communicators at different stages of development.

The other area of empirical investigation concerns people with atypical or impaired communicative capacities. These are usually looked at alongside, or as an aspect of, atypical or impaired mind-reading capacities, autism being a much-studied case in point. Autistic people are widely seen as lacking certain mind-reading abilities (in particular, but not only, the capacity to attribute epistemic mental states (Leslie 1991, Baron-Cohen 1995)) and many also have difficulty understanding non-verbal communication, non-literal verbal communication and the various facial and prosodic expressions of affect that often accompany verbal acts. In an early test of the RT prediction that irony is more

metarepresentationally complex than metaphor, Happé (1993) showed that a group of autistic people who could understand similes had problems with both metaphor and irony, while another group who were able to handle metaphor could not grasp irony. She correlated this with the different levels of general mind-reading (in)capacities of the two groups as measured by performance on false belief tasks. More recently, Langdon et al. (2002) report similar results from studies of metaphor and irony understanding by people with right-hemisphere brain damage. However, on the basis of extensive testing of a group of schizophrenic people, their conclusion about the schizophrenic difficulty with metaphor is that it cannot be accounted for in terms of a theory of mind deficit. Their tentative suggestion here, based on the RT account briefly mentioned in section 2.1, is that the problem lies with disorganisation or degradation of the schizophrenic person's conceptual networks, which interferes with the kind of adjustment to the literal encoded concept that is necessary for metaphor understanding (Langdon et al. 2002: 98). People with Williams Syndrome are generally thought to have good mind-reading capacities but, while they are often volubly communicative, recent work indicates atypical lexical processing and difficulty with certain kinds of metaphor understanding (Thomas et al. unpublished data). This looks like another test-bed for RT ideas about lexical adjustment, including cases of metaphor.

On the one hand, RT has a wealth of ideas to offer to these various areas of empirical investigation; on the other, the theory itself has much to gain from the pressure for explicitness required in forming testable hypotheses and, of course, from the resulting evidence that may confirm or disconfirm its predictions. New directions for research within the relevance-theoretic framework will surely arise from this cross-fertilisation between the theoretical and the empirical.

References:

- Asher, N. 1999. Discourse structure and the logic of conversation. In: Turner, K. (ed.) *The Semantics/Pragmatics Interface from Different Points of View*. Oxford: Elsevier Science. 19-48.
- Barkow, J., Cosmides, L., & 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, Mass.: MIT Press.
- Bezuidenhout, A. 1997. Pragmatics, semantic underdetermination and the referential/attributive distinction. *Mind* 106: 375-409.
- Bloom, P. 2002. Mindreading, communication and the learning of names for things. *Mind and Language* 17: 37-54.
- Bott, L. and Noveck, I. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51: 437-457.
- Breheny, R. in press. Communication and folk psychology. *Mind and Language*.
- Breheny, R., Katsos, N. and Williams, J. in press. Are scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*.
- Cappelen, H. and E. Lepore. 2005. *Insensitive Semantics*. Oxford: Blackwell.
- Carruthers, P. & Chamberlain, A. (eds.) 2000. *Evolution and the Human Mind: Modularity, Language and Meta-Cognition*. Cambridge: Cambridge University Press.
- Carston, R. 1998. Informativeness, relevance and scalar implicature. In: Carston, R. and Uchida, S. (eds.) *Relevance Theory: Applications and Implications*. Amsterdam: John Benjamins. 179-236.
- Carston, R. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Carston, R. 2004a. Explicature and semantics. In: Davis, S. and Gillon, B. (eds.) *Semantics: A Reader*. Oxford: Oxford University Press. 817-845.
- Carston, R. 2004b. Truth-conditional content and conversational implicature. In: Bianchi, C. (ed.) *The Semantics/Pragmatics Distinction*. CSLI. Stanford University. 65-100.
- Carston, R. forthcoming. Pragmatics, semantics and semantics/pragmatics distinctions. *Synthèse*.

- Carston, R. and Wilson, D. forthcoming/2006. Lexical pragmatics and Relevance Theory. In: Burton-Roberts, N. (ed.) *Advances in Pragmatics*. Palgrave.
- Chierchia, G. 2004. Scalar implicatures, polarity phenomena and the syntax/pragmatic interface. In: Belletti, A. (ed.) *Structures and Beyond*. Oxford: Oxford University Press.
- Davidson, D. 1973. Radical interpretation. *Dialectica* 27, 313-328. Reprinted in Davidson, D. (1984) *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press. 125-139.
- Davidson, D. 1986. A nice derangement of epitaphs. In: Lepore, E. (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell. 433-446.
- Donnellan, K. 1966. Reference and definite descriptions. *Philosophical Review* 75: 281-304.
- Fodor, J. 1983. *Modularity of Mind*. Cambridge Mass.: MIT Press.
- Fodor, J. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.
- Fodor, J. 2000. *The Mind Doesn't Work That Way*. Cambridge, Mass.: MIT Press.
- Grice, H.P. 1957. Meaning. *The Philosophical Review* 66: 377-388. Reprinted in H.P. Grice 1989: 213-223.
- Grice, H. P. 1967. Logic and Conversation. William James Lectures. Reprinted in H.P. Grice 1989: 1-143.
- Grice, H.P. 1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.
- Happé, F. 1993. Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition* 48, 101-119.
- Happé, F. and Loth, E. 2002. 'Theory of mind' and tracking speakers' intentions. *Mind and Language* 17, 24-36.
- Katsos, N., Breheny, R. and Williams, J. 2005. Interaction of structural and contextual constraints during the on-line generation of scalar inferences. Proceedings of the 27th Annual Meeting of the Cognitive Science Society, Stresa, Italy.
- Kripke, S. 1979. A puzzle about belief. In: Margalit, A. (ed.) *Meaning and Use*. Dordrecht: Reidel. 239-283.
- Langdon, R., Davies, M. and Coltheart, M. 2002. Understanding minds and understanding communicated meanings in schizophrenia. *Mind and Language* 17, 68-104.
- Leslie, A. 1991. The theory of mind impairment in autism: Evidence for a modular mechanism of development? In: Whiten, A. (ed.) *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading*. Oxford: Blackwell.
- Levinson, S. 2000. *Presumptive Meanings*. Cambridge, Mass.: MIT Press.

- Noveck, I. and Posada, A. 2003. Characterising the time course of an implicature: an evoked potentials study. *Brain and Language* 85: 203-210.
- Noveck, I. and Sperber, D. (eds.) 2004. *Experimental Pragmatics*. Basingstoke: Palgrave.
- Perry, J. 1986. Thought without representation. *The Aristotelian Society Supplementary Volume LX*: 137-151.
- Powell, G. 1998. The deferred interpretation of indexicals and proper names. *UCL Working Papers in Linguistics* 10: 143-172.
- Powell, G. 2001. The referential-attributive distinction – a cognitive account. *Pragmatics and Cognition* 9(1): 69-98.
- Powell, G. 2002. Underdetermination and the principles of semantic theory. *Proceedings of the Aristotelian Society* 102(3): 271-278.
- Powell, G. 2003. *Language, Thought and Reference*. University of London PhD thesis.
- Recanati, F. 2002a. Unarticulated constituents. *Linguistics and Philosophy* 25: 299-345.
- Recanati, F. 2002b. Does linguistic communication rest on inference? *Mind and Language* 17: 105-126.
- Recanati, F. 2004. *Literal Meaning*. Cambridge: Cambridge University Press.
- Rouchota, V. 1992. On the referential/attributive distinction. *Lingua* 87: 137-167.
- Rubio, P. 2005. *Pragmatic Processes and Cognitive Mechanisms in Lexical Interpretation*. University of Cambridge PhD thesis.
- Scholl, B. and Leslie, A. 1999. Modularity, development and ‘theory of mind’. *Mind and Language* 14: 131-153.
- Searle, J. 1983. *Intentionality*. Cambridge: Cambridge University Press.
- Sperber, D. 1994a. Understanding verbal understanding. In: Khalfa, J. (ed.) *What is Intelligence?* Cambridge: Cambridge University Press. 179-198.
- Sperber, D. 1994b. The modularity of thought and the epidemiology of representations. In: Hirschfeld, L. and Gelman, S. (eds.) *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge: Cambridge University Press. 39-67.
- Sperber, D. 2000. Metarepresentations in an evolutionary perspective. In: Sperber, D. (ed.) *Metarepresentations: A Multidisciplinary Perspective*. Oxford University Press. 117-137.
- Sperber, D. and Wilson, D. 1986/95a. *Relevance: Communication and Cognition*. Oxford: Blackwell. Second edition 1995.
- Sperber, D. and Wilson, D. 1995b. Postface. In: Sperber, D. and Wilson D. 1995a: 255-279.

- Sperber, D. and Wilson, D. 2002. Pragmatics, modularity and mind-reading. *Mind and Language* 17: 3-23.
- Stanley, J. 2000. Context and logical form. *Linguistics and Philosophy* 23: 391-434.
- Thomas, M. S. C., van Duuren, M., Ansari, D., Parmigiani, C., and Karmiloff-Smith, A. unpublished data. The development of semantic categories and metaphor comprehension in Williams Syndrome.
- Tomasello, M. and Rakoczy, H. 2003. What makes human cognition unique? *Mind and Language* 18: 121-147.
- Van der Henst, J-B. and Sperber, D. 2004. Testing the cognitive and communicative principles of relevance. In: Noveck, I. and Sperber, D. (eds.) 2004: 141-171.
- Wharton, T. 2003. Natural pragmatics and natural codes. *Mind and Language* 18: 447-477.
- Wilson, D. 2000. Metarepresentation in linguistic communication. In: Sperber, D. (ed.) *Metarepresentations: A Multidisciplinary Perspective*. Oxford University Press. 411-448.
- Wilson, D. 2003. New directions for research on pragmatics and modularity. *UCL working Papers in Linguistics* 15: 105-127.
- Wilson, D. 2004. Relevance and lexical pragmatics. *UCL Working Papers in Linguistics* 16: 343-360.
- Wilson, D. and Sperber, D. 2002. Truthfulness and relevance. *Mind* 111: 583-632.
- Wilson, D. and Sperber, D. 2004. Relevance theory. In: Horn, L. and Ward, G. (eds.) *Handbook of Pragmatics*. Oxford: Blackwell. 607-632.

Robyn Carston
 Department of Phonetics and Linguistics
 University College London
 Gower Street
 London WC1E 6BT

Email: robyn@ling.ucl.ac.uk

George Powell
 Department of Phonetics and Linguistics
 University College London
 Gower Street
 London WC1E 6BT

Email: george@ling.ucl.ac.uk