Speech Production, Dual-Process Theory, and the Attentive Addressee

Anna Jane Pollard

Thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

University College London
April 2012

I, Anna Jane Pollard, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: ……

Abstract

This thesis outlines a model of Speaker-Addressee interaction that suggests some answers to two linked problems current in speech production. The first concerns an under-researched issue in psycholinguistics: how are decisions about speech content – conceptualization – carried out? The second, a pragmatics problem, asks how Speakers, working under the heavy time pressures of normal dialogue, achieve optimal relevance often enough for successful communication to take place.

Links between these problems are discussed in Chapter 1; Chapter 2 reviews existing research on speech production and dialogue. Chapter 3 presents the central claim of my thesis: that the Addressee exerts a significant influence over the Speaker's decision-making at a level below the latter's consciousness. Using evidence drawn from psycholinguistics, developmental psychology and human-computer interaction, Chapter 4 presents evidence to support this claim, demonstrating that a Speaker's performance can be decisively affected at a preconscious level by the degree of attentiveness shown by the Addressee. Lack of attentiveness, in particular, appears to damage speech production at the conceptualization level. I suggest, therefore, that Speaker and Addressee are linked in a feedback loop: unless a Speaker achieves and maintains relevance to an Addressee, the Addressee's interest will be lost, and this will impair the Speaker's production abilities and hence the communication process itself.

Chapters 5 and 6 consider some automatic mechanisms that may help Speakers dovetail their productions to Addressee need. These include the neural mechanisms underlying face perception and social rejection; automatic aspects of theory of mind; intuitive memory and inference systems of the type being explored in dual-process theory; and connections between verbal performance and behavioural priming currently being investigated. Chapter 7 summarizes the complete argument, discusses its wider implications, and includes suggestions for further work.

Table of Contents

Acknowledgements

First and foremost, my limitless thanks to my supervisor Deirdre Wilson, without whom I would never have started this project, let alone finished it. Her guidance, her encouragement, her support during the difficult times, and her faith in my idea during the doubt-filled ones: these, together with her friendship and the delight that she communicated in research, have made my PhD years truly wonderful.

Many thanks, too, to UCL's linguistics, research, and support staff: for example, to Robyn Carston, who first alerted me to the problem of relevant speech production; to Neil Smith and John Harris, whose lectures I still remember with joy; to Richard Breheny and Paula Rubio-Fernandez, who introduced me to psycholinguistics; to Nick Allott, who gave me my first contact with Relevance Theory; and to Molly Bennett, whose calming presence was always an immense source of reassurance.

I also owe a big debt of gratitude to the UCL Institute of Cognitive Neuroscience, to their Monday seminars, and to all those whom I met at them and who stimulated my interest in the subject: to - for example - Chris and Uta Frith, to Vincent Walsh and, in particular, to Geoff Bird, whose enthusiasm and encouragement have been a source of real inspiration. And I want to say a very special thank you to Amanda d'Souza, with whom Monday evenings at the ICN became a regular occurrence, and who listened cheerfully and tirelessly to my accounts of research progress (or the lack of it). I also want to thank Alison Hall, Kate Scott, Ingrid Falkum, Hiroyuki Uchida, Marsha Hill, Eric Carlson and all my other fellow UCL students who encouraged and supported me, both during my MA and my research years. Further afield, I am more grateful than I can say to Clare Calder and to my 'friends in the Street' – Kate, Sam, and Sarah – for cheering me on and keeping me going throughout the whole period.

Finally, my deepest gratitude to my late grandmother and mother, Evelyn Harenc and Sara Sproule, to whose inspiration and encouragement I owe so much.

All mistakes in this thesis are my own.

Chapter 1

Introduction

Why do people say the things they do? How do they decide what they wish to achieve in speaking, work out how to reach their goals, shape their ideas into something that they think will convey their meaning to their audience?

As outlined in Levelt's seminal work on speaking, these are areas of decision-making that constitute a distinct stage – the message level, or Conceptualizer – in the speech production process. 'Talking as an intentional activity,' he writes, 'involves conceiving of an intention, selecting the relevant information to be expressed for the realization of this purpose, ordering this information for expression, keeping track of what was said before, and so on… The sum total of these mental activities will be called *conceptualizing*… The product of conceptualizing will be called the *preverbal message*' (Levelt, 1989: 9, author's emphases). And, although questions about these processes can be simply phrased, answers to them do not appear to be readily available.

In the speech production model Levelt describes, Conceptualizer output feeds down to a second level, where it is translated into linguistic form, and then to a third, where it becomes overt speech, and the mechanisms involved in these formulatory and articulatory stages of speech production have received enormous amounts of attention over the years. Conceptualization, however, has not. According to Harley (2001: 349), this process of interfacing both with the outside world and with the interior one of long-term memory is to some extent the 'forgotten level' of speech production, and very little is known about its format or about the processes of conceptualization as a whole. Moreover, although the notion of conceptualization was first used in psycholinguistics, the problems that it presents are not only psycholinguistic ones: broader pragmatic processes that apply in both verbal and non-verbal communication are also involved. As Sperber (2008) puts it, discussing

the current state of Relevance Theory, speech production is the 'next frontier … the biggest challenge we have got left.'

While Relevance Theory has provided important insights into the nature of relevance and its role in communication and cognition, it has focused mainly on inferential processes, and on inferential comprehension in particular. As Sperber & Wilson point out (1986/1995: 279), it has left largely unexplored the 'important and related domains' of the cognitive processes at work in the communicator, and the social character and context of communication. The question I address in this thesis brings together both these domains: how does a communicator succeed more often than not in helping an interlocutor cross the gap between sentence meaning and speaker's meaning, thus enabling the ready and accurate comprehension of the message conveyed?

According to relevance theory, communication succeeds by creating and satisfying expectations of relevance. Van der Henst & Sperber succinctly explain why, within this framework, Speakers must manage more often than not to make their utterances relevant to the Hearer:

> Speakers may fail to achieve relevance, or they may not even try, and in such cases the presumption of optimal relevance is unjustified. Justified or not, it is automatically conveyed by every utterance used in communication, and it guides the process of comprehension… The communicative principle could not be right – and relevance could not guide comprehension – if speakers were not, often enough, trying to be optimally relevant, *and successful at it*.
>
> (Van der Henst & Sperber, 2004: 268, my emphasis)

In other words, inferential communication would not work unless Speakers were somehow ensuring that it did, and doing so under the heavy and continual time pressures involved in normal dialogue. These are considerable – in Levelt's calculation (*op cit*: 22) speech is produced at the rate of two to three words per second – and time constraints are not the only ones. Under ordinary conversational circumstances, a Speaker has a mass of external stimuli and internal representations competing for his attention and processing resources: physical conditions, or preoccupations with past actions or future commitments, for example. Carrying on an 'ordinary' conversation, with all the decision-making involved, is therefore a far

more demanding task than it at first appears, not only at the levels of formulation and articulation, but also at the conceptual level.

It seems astonishing that relatively little attention has been paid to investigating the message-generating aspect of speech production and the cognitive and affective processes involved. One of the reasons may be that, as indicated above, the problem is a double one, in which the Speaker's conceptual ability and the Hearer's readiness to process the resulting output are closely intertwined. This makes the task of disentangling them particularly challenging, but it also suggests that insight into one should give insight into the other – something in fact acknowledged by Levelt himself. In a discussion of Speaker-Hearer interaction, he writes:

> It is often enough for a speaker to just hint at a certain piece of information; the addressee will interpret that information as relevant to the ongoing interaction and will infer the speaker's intention. Sperber and Wilson… argued that this facilitates processing for the listener. It probably also does so for the speaker. There is normally no need to formulate one's intentions in all detail for communication to be effective. The challenge for a psycholinguistic theory of the speaker is precisely to understand how speakers manage to select for expression just the relevant information.
>
> (*op cit*: 43)

More recently, Indefrey & Levelt also stress the inter-dependence of speech production and speech processing. In their meta-analysis of imaging literature on word production, from lexical selection through to the articulatory stage, they point out that, although theories of word production and word perception have developed within distinct research traditions, there cannot be any reasonable doubt that the two functional systems are very closely linked. Indeed, they note that a spoken language user 'constantly operates a dual system, perceiving and producing utterances. These systems not only alternate, but in many cases they partially or wholly operate in concert' (Indefrey & Levelt, 2004: 102).

Bavelas et al (2000: 941) make a similar point from a slightly different perspective. Introducing aspects of their work on the social nature of language use (see Chapter 4), they stress the need to consider communicative production and perception in terms, not just of systems, but of the people who operate them – and, in particular, in terms of how perceivers interact with producers. 'Listeners,' they comment, 'have at

best a tenuous foothold in most theories. At the most extreme, listeners are considered nonexistent or irrelevant because the theory either does not mention them or treats them as peripheral.' A slightly less extreme view treats the listener as a '"speaker-in-waiting"… present but not active during the other's speech'.

As Bavelas et al point out, the work of H.H. Clark and others on dialogue as a joint activity represents a major exception to this single-focus approach, and this might be taken to indicate an increasing acceptance of verbal communication as a single system, involving Speaker and Hearer as *simultaneously* active participants. However, it is still hard to trace any major impact of this approach on the double speech production problem described above: how does the Conceptualizer work, and how does it ensure that utterances tend to satisfy the hearer's expectations of relevance? Speculation on these problems raises a further question that also seems to have escaped widespread examination. Levelt opens his description of the Conceptualizer with the words, 'Talking as an intentional activity involves conceiving of an intention…'. But how far does talk – ordinary, everyday talk, carried out in the supermarket queue or when unexpectedly meeting a neighbour – really involve 'conceiving of an intention', in the sense that its output is the result of some deliberately-adopted strategy, selected from amongst many to meet the demands of the particular situation? More generally, given the rate at which normal speech takes place, should a Speaker be seen as making deliberate, conscious decisions in the course of the conceptualization process? Or are there circumstances in which decisions may be made automatically, below a Speaker's level of consciousness, so that fluent, context-appropriate communication can take place despite the time and other pressures?

It is the assumption that there may be an automatic aspect to the Conceptualizer that has guided me in approaching my research problem. My central claim is that a Speaker's decision-making can be influenced significantly at a preconscious level by the presence of a far-from-peripheral Addressee. Using evidence from psycholinguistics, developmental psychology and human-computer interaction, I argue that a Speaker's performance can be continuously and decisively affected by the attentiveness or otherwise of an Addressee, and that lack of such attentiveness, in particular, appears to do damage at the conceptualization level. As a result, Speaker

and Addressee are connected in a feedback loop: unless a Speaker *achieves and maintains relevance* to an Addressee, the latter's interest will be lost, which will in turn impair the Speaker's production abilities and hence the communication process itself.

Using further interdisciplinary evidence drawn from cognitive science and social cognitive neuroscience, I then consider some automatic mechanisms that may help the Speaker maintain successful rapport with the Addressee under the heavy pressures of normal dialogue. These include the neural mechanisms underlying face perception and social rejection; automatic aspects of theory of mind; intuitive memory and inference systems of the type being explored in dual-process theory; and the connections currently being investigated between verbal performance and behavioural priming. The combined effect of these mechanisms, I suggest, is to help create and maintain a continual dovetailing between Speakers' productions and Addressees' needs, thus shedding some light on why Speakers produce the utterances they do, and how they succeed in making them optimally relevant to their audience.

The multidisciplinary nature of the evidence I present has to a large extent been dictated by the complex nature of the problem itself. This complexity is summarised by Singer et al (2003/4: xvii): 'The study of social interaction involves by definition a bi-directional perspective and is concerned with the question of how two minds shape each other mutually through reciprocal interactions. To understand interactive minds we have to understand how thoughts, feelings, intentions, and beliefs can be transmitted from one mind to the other.' In order to address these various aspects of interpersonal communication, I have looked as widely as possible at different research fields for fragments of evidence that demonstrate the Addressee's influence on utterance production. The model of Speaker-Addressee interaction that I now present has been constructed from research conducted over a long span of time and using a wide variety of techniques and paradigms: a variety that, together with the quality of the evidence quoted, I hope contributes to the robustness of the argument that I present.

My argument opens (in Chapter 2) with an account of some of the main strands of psycholinguistics research that address issues of speech production and dialogue. I start with a detailed presentation of Levelt's speech production model, and go on to

discuss key features of both situation modelling and common ground theory. I then describe Pickering & Garrod's 'mechanistic psychology of dialogue', which treats linguistic priming as its driving mechanism and has played an important part in the development of my own account of Speaker-Addressee interaction.

In Chapter 3, using examples gathered informally from naturally-occurring dialogue, I outline my initial hypothesis: that there is a low-level, automatic system of interactions between Speaker and Addressee through which the Addressee can continually influence the Speaker's output, thus helping to ensure its relevance via continual adjustment to the readiness and capacity of the Addressee to process it.

Chapter 4 surveys experimental evidence to support this hypothesis, by identifying Addressee behaviour that promotes or discourages interaction. Evidence cited includes the 'still-face' effect in infants (cf Murray & Trevarthen , Striano); variations in adult gaze interaction (cf Kendon, Bavelas et al); and behavioural patterns in virtual reality environments (cf Pertaub et al, Garau et al, Boker, Cohn et al). Arising from this evidence, I extend the hypothesis in the following way: the Addressee's influence is such that a Speaker *must* achieve relevance – must engage and retain his interlocutor's attention – if he is to continue speaking competently.

In Chapter 5, I review the arguments that have led to this extended hypothesis of an Addressee-Speaker feedback loop, and go on to explore the means by which this feedback loop can be maintained, to the advantage of Speakers and communication in general, at the level of social interaction. Using as a background the basic assumptions of dual-process theory (cf Frankish & Evans, Smith & DeCoster, Mercier & Sperber), I consider the relevance of social cognitive neuroscience to my hypothesis, and survey neural evidence of reactions to social rejection (cf Eisenberger et al) and of responsiveness to facial movements (cf Baron-Cohen, Adolphs, Haxby et al). I end by considering potential contributions to Speaker social competence from mirror neuron theory and theory of mind, and potential connections between automatic neural processes and linguistic processing (cf Ullman, Adolphs).

In Chapter 6, I return to the question raised at the start of this Introduction: why do people say the things they do? Or, as re-worded in terms of my hypothesis, how far does a theory of automatic Addressee-Speaker rapport extend to the

conceptualization level of speech production, thus explaining how Speakers can work under the demanding cognitive and time constraints of normal dialogue, while at the same time retaining Addressee attention? I discuss two potential mechanisms for addressing this hitherto under-explored issue: verbal routines/routinisation (cf Pickering & Garrod, Kuiper), and the impact of behavioural – as opposed to linguistic – priming on verbal behaviour (cf Bargh, Kay et al, Catmur et al).

In Chapter 7, whose title is the same as that of the thesis, I present the conclusions drawn from my work.

CHAPTER 2

Speakers, audiences and models of dialogue

As explained in the previous chapter, the dual problem that is the subject of my thesis can be usefully addressed by an interdisciplinary approach that calls on insights and data from a wide variety of fields. These include pragmatics, psycholinguistics, discourse analysis, sociolinguistics, computer-mediated communication, cognitive and social psychology, and the rapidly-growing domain of social cognitive neuroscience. From within this broad spectrum, three main areas of inquiry emerge that, I suggest, contribute both to each other and to the development of a general theory of Speaker-Addressee interaction. They are speech production itself, the associated phenomenon of priming, and dual process theory, which distinguishes conscious from automatic types of cognitive activity.

In this chapter, I review research on the speech production aspects of the problem, with particular reference to the work of Levelt (speech production; Chapter 2.1), Zwaan & Radvansky (situation modelling; Chapter 2.2), Clark and others ('grounding' and audience modelling in general; Chapter 2.3), and Pickering & Garrod's model of dialogic interaction, which introduces the issue of priming (Chapter 2.4). The order in which I review them is more than purely chronological, although it does broadly follow a historical sequence. My intention is to show that there has been a progressive broadening in content, from the conception of Speaker as isolated entity, through a concern with the Speaker in a context, to the investigation of dialogue: a Speaker in interaction with a Hearer, who becomes a Speaker in his[1] turn.

[1] For reasons that will become obvious, the linguistics tradition of treating a Speaker as female and a Hearer as male is inapplicable in an investigation of dyadic communication, where the roles of Speaker and Hearer are exchanged continually but the identities of each participant remain constant. Where possible without sacrificing readability, I will deal with this by using the plural pronoun 'they' for both; in this sentence, the older practice of using the masculine pronoun to include both male and female referents seems appropriate. This usage will also be encountered in many of the passages quoted in the course of this thesis.

In later chapters, I will show how the issues of priming and automaticity make a major contribution to my account.

2.1 The Speaker in action: Levelt's model of speech production

The full version of Levelt's 1989 summary of the conceptualization process is as follows:

> Talking as an intentional activity involves conceiving of an intention, selecting the relevant information to be expressed for the realization of this purpose, ordering this information for expression, keeping track of what was said before, and so on. These activities require the speaker's constant attention. The speaker will, moreover, attend to his own productions, monitoring what he is saying and how… The sum total of these mental activities will be conceptualizing, and the subserving processing system will on occasion be called the Conceptualizer (in full awareness that this is a reification in need of further explanation – we are, of course, dealing with a highly open-ended system involving quite heterogeneous aspects of the speaker as an acting person). The product of conceptualizing will be called the preverbal message.
>
> (Levelt 1989: 9)[2]

The hypothesised mechanism for handling the collection of functions Levelt describes is a message generator, which has access to a speaker's knowledge of the world, and a monitoring unit, which can access both planned and overt speech. Together, they comprise the initialiser in a sequence of sub-system activities that culminates in an audible utterance. The output from each sub-system forms the input to the next: thus, the Conceptualizer passes on the message fragments it has developed to the Formulator, which uses two sub-systems of its own – Grammatical and Phonological Encoders – to turn a conceptual structure into a linguistic one. The resulting phonetic representation, or 'internal speech', is then passed for conversion into overt speech to the articulatory sub-system where, owing to asynchronies

[2] Levelt's later work – and, in particular, his 'weaver11' computational model of speech production – puts less emphasis on the conceptualization stage of speech production. Indeed, 'weaver11' specifically excludes it: 'From lifelong experience, speakers usually know what concepts are lexically expressible in their language. Our theory of lexical access is not well developed for this initial stage of conceptual preparation… In particular, the computational model does not cover this stage.' (Levelt et al, 1999: 8).

between cognitive modelling and physical movement, it is envisaged as being stored in a buffer before being processed by the Articulator itself.

Levelt argues that, owing to their specialised nature, most of the components in this speech production model can function in a highly automatic, reflex-like way: a characteristic that enables them to work in parallel, thus allowing the generation of uninterrupted fluent speech. In spite of this parallelism, however, and notwithstanding the model's feedback loops for speech monitoring, the flow of information through the system is unquestionably top-down. Speaking, as Levelt points out, is 'usually an intentional activity: it serves a purpose the speaker wants to realise' (*ibid*: 20), and it is in the Conceptualizer that the match between intention and planned action takes place. Input to this is also input to the whole system, and draws on two types of pre-existing Speaker knowledge: procedural and declarative. The former consists of action plans ('IF one intends to commit oneself to the truth of proposition *p*, THEN assert *p*.'). The latter consists of both the Speaker's encyclopaedic knowledge – his knowledge of the world – and his knowledge of the current discourse situation: 'The speaker can be aware of the interlocutors – where they are and who they are. The speaker, moreover, may be in the perceptual presence of a visual array, of acoustic information about the environment, and so forth. This *situational knowledge* may also be accessible as declarative knowledge, to be used in the encoding of messages' (*ibid*: 10). Additionally, Speakers will register facts about the discourse itself, such as its type (e.g. interview, lecture, informal conversation) and topics, and will also keep track of their own and their interlocutors' contributions to it: overall, they will construct a Speaker's discourse model.

In the present research context, the inclusion of such situational knowledge within the conceptualization framework is of considerable importance, as it includes, by definition, not just where and who the interlocutors are, but a further feature which varies from situation to situation: *what* they are doing. As I hope to show in the following chapters, one particular aspect of what a Hearer is doing has a major impact on what the Speaker himself does – or, perhaps, does not do.

2.1.1 Conceptualizer control…Conceptualizer automaticity?

As the above summary makes clear, Levelt himself leaves space for additions to his framework: as I also hope to show, the processes at work in a Speaker are indeed heterogeneous. But this is not the only point at which there is an overlap between Levelt's model and my own inquiry. The real point of intersection between the two occurs in the self-monitoring function, input to which comes from two sources. One, the Speech Comprehension System, is external to the Conceptualizer, and allows Speakers to monitor the output of both their Formulator (internal speech) and their Articulator (overt speech, audible to an interlocutor and themselves). The other is internal to the Conceptualizer, and is the Message Generator itself.

While Levelt acknowledges the theoretical existence of this intra-Conceptualizer source of input for the Monitor, he appears interested in it mainly in order to emphasise the proposed relationship between Conceptualizer and monitoring as a whole:

> When the speaker detects serious trouble with respect to the meaning or well-formedness of his own internal or overt speech, he may decide to halt further formulation of the present utterance. He may then rerun the same preverbal message or a fragment thereof, create a different or additional message, or just continue formulation without alteration. These processes are not of a different nature than what is going on in message construction anyhow.
>
> The speaker *no doubt* [my own emphasis; note the guardedness of Levelt's tone] monitors messages before they are sent into the Formulator… considering whether they will have the intended effect in view of the present state of the discourse and the knowledge shared with the interlocutor(s). Hence there is no good reason for distinguishing a relatively autonomous monitoring component in language production. The main work is done by the Conceptualizer, which can attend to internally generated messages and to the output of the Speech-Comprehension System (i.e. parsed internal and overt speech).
>
> (*ibid*: 14)

The importance of the intra-Conceptualizer monitoring loop is made clear in his detailed account of the whole self-monitoring process, where this loop warrants the following brief description:

> Before we turn to the connectionist accounts of self-monitoring, one additional remark should be made. Not all self-monitoring during speech is mediated by the perceptual loops. Speakers can also monitor their messages before they are formulated. They can attend to the appropriateness, the instrumentality, and the politeness of the speech act they are planning. This is an entirely conceptual activity; it need not involve any formulation or parsing.'
>
> (*ibid*: 474)

One of my aims in this thesis is to ascertain whether this rather sparse account of self-monitoring within the conceptualization process – which essentially amounts to a capacity to inhibit the Conceptualizer – can be amplified and, if so, what agents and processes are involved.

Another point of particular relevance to my inquiry is Levelt's stance on the automaticity of the speech production processes described in his model. Here, he draws a clear distinction between the Conceptualizer and the stages that follow. The latter, he says, are 'largely automatic'. Indeed, given the time pressures involved, they can hardly be otherwise: 'These things come automatically without any awareness. They also come with very high speed. Speech is normally produced at the rate of about two to three words a second. [Words] are selected at that rate from the many tens of thousands of words in the mental lexicon. There is just no time to consciously weigh the alternatives before deciding on a word. Articulation runs at a speed of about fifteen phonemes per second… Formulating and articulating are "underground processes"' (*ibid*: 22). As automatic processes, therefore, the hypothesised Formulator and Articulator sub-systems can function without awareness or conscious intention. Each draws on its specialised input, and produces an equally specialised output for processing at the next stage in the sequence. The Conceptualizer, however, is different.

Because speaking is usually an intentional activity, it is presented as being crucially under central or executive control. 'A speaker,' Levelt argues, 'can decide on one course of verbal action rather than another on the basis of practically any sort of information: his state of motivation, his obligations, his believing this rather than that, his previous speech acts or other actions, and so forth. The speaker will invest his attention on matters of this sort in planning what to say next' (*ibid*: 20). It is the Conceptualizer that handles such decisions and its workings, therefore, involve

highly controlled processing: an operation that, unlike the automatic processes, is affected by competition for the 'limited-capacity resource' of working memory, is usually serial rather than parallel, and is for this reason slow. Furthermore, if message construction is controlled processing, so is monitoring, with self-corrections hardly ever being made without 'a touch' of awareness.

However – and very importantly – Levelt acknowledges that this executive control of the Conceptualizer is not total: 'an adult's experience with speaking is so extensive that whole messages will be available in long-term memory and thus will be retrievable. Many conversation skills … have been acquired over the course of a lifetime and are quite directly available to the speaker. They are not invented time and again through conscious processing' (*ibid*: 21). As explained in Chapter 1, a key goal of my own inquiry is to assess how far the machinery of the Conceptualizer – the inaugural decision-maker of the speech production process – may be under pre-conscious, automatic control to an even greater extent than Levelt suggests. And, again, the question arises: who is the controller?

2.2 The Speaker in context: situational representations

Speakers do not work in isolation. As Levelt's model acknowledges, they operate in a context – a landscape populated by objects and people, and governed by considerations of space and time. Indeed, they operate in two overlapping landscapes simultaneously: the physical one, developing in real time around them, and the mental one, a landscape of the mind (featuring discourse models and much else) on which they draw to create utterances. Though one is physical and the other mental, the essence of the two is the same: both are situations which can be described or represented in conceptual terms, and both make essential contributions to the Speaker's declarative knowledge. In this section and the next, I review some of the issues raised by consideration of Speakers in situations – in contexts.

2.2.1 The 'way the world is': the role of the situation model

A theoretical construct that has been developed to analyse representational situations is that of the situation model, also known as the mental model: a 'multi-dimensional

representation of the situation under discussion … [which is] assumed to capture what people are "thinking about" while they understand a text' (Pickering & Garrod 2004: 172). As described by one of the originators of this approach, Johnson-Laird (1983: 2), a situation model is a theoretical entity, a mental representation of an element of external reality that forms a 'working model' of it and that allows the representer to execute mental operations on it. He continues:

> It is… plausible to suppose that mental models play a central and unifying role in representing objects, states of affairs, sequences of events, the way the world is, and the social and psychological actions of daily life. They enable individuals to make inferences and predictions, to understand phenomena, to decide what action to take and to control its execution, and above all to experience events by proxy; they allow language to be used to create representations comparable to those deriving from direct acquaintance with the world; and they relate words to the world by way of conception and perception.
> (*ibid*: 397)

Aspects of external reality represented in such a model include the temporal, spatial, and motivational aspects of the situation, together with the persons and objects involved. Thus, a model of a football game calls for a 'temporal sequence of events at various locations, for causal relations between the events, and for the representation of individuals, interacting physically and socially, governed by physical laws and constrained by the "laws" of the game and social conventions and motivated by various intentions' (*ibid*: 414). However, Johnson-Laird avoids the question of what mental models actually *are,* suggesting instead that, as they are supposed to be in people's heads, their exact constitution is 'an empirical question' (*ibid*: 398). Nor has this overall definition tightened over the years.

For instance, Zwaan & Radvansky, in their major review of the field, merely comment that mental/situation models are 'likely to be rather abstract representations in which, for example, tokens may represent protagonists or objects' (Zwaan & Radvansky, 1998: 31). It is also worth noting that, until recently, situation model theory as applied to communication appears to have been developed and discussed almost entirely with reference to comprehension of a verbally described situation, to Readers/Hearers rather than to Speakers (a point to which I will return). Nonetheless, there has been considerable research into the situation dimensions of

space, causation, intentionality, protagonists/objects and, to a lesser degree, time, and the resulting theoretical frameworks form an important underpinning to the work of Pickering & Garrod on dialogue, discussed in Chapter 2.4.

Zwaan & Radvansky describe a situation model as an 'integrated mental representation' of a described state of affairs, and their main concern is with the various dimensions (or 'indexes') involved in this integration. Distinguishing between a situation model – a representation of a token situation such as a specific visit to a restaurant – and a schema – a representation of a situation type such as a stereotypical visit to a restaurant – they note that schemata can be used as building blocks for situation model construction. However, the focus of their review is on the situation model itself, rather than the schemata which may play a role in its construction.

Findings that they survey include the suggestions that:

- Readers do not necessarily track spatial information spontaneously during comprehension, but they can do so if asked
- Memory performance is better when information can be easily integrated into a situation model than when it is stored across a number of representations
- Readers routinely keep track of causal information, whether indicated explicitly in text (eg by a causal connective such as 'because', to which they are highly sensitive), or inferred via a reader's knowledge of events
- Goal statements carry 'considerable weight' during text comprehension
- 'Very little' is known about the cognitive processing of temporal information in language, despite its importance and ubiquity
- In recall of a narrative, readers remember goal-related information about a protagonist better than other information not related to protagonist intentions

(Zwaan & Radvansky, 1998: 12-26)

As the last of the above points indicates, it is representations of the protagonists themselves that appear to lie at the core of the comprehension process, dominating

the landscape of recipients' minds and governing their comprehension of the whole story.

2.2.2 Situation models and the dominant protagonist

'Protagonists and objects' Zwaan & Radvansky say, 'form the "meat" of situation models… The general conclusion [from research] is that readers appear to be intensively engaged in keeping track of protagonists during comprehension whereas the amount of focus on objects appears to be more dependent on contextual cues' (*ibid*: 23). Thus, readers faced with a referentially ambivalent pronoun tend to resolve it in favour of a text's main protagonist (Morrow, 1985: 304). Reference resolution is also facilitated by the use of a protagonist's proper name rather than a role description, even if they denote the same person (Sanford et al 1988: 43)). Furthermore, some encyclopaedic properties of situation protagonists appear to be strongly associated with them, and therefore become highly activated when the protagonist is mentioned, whether they are explicitly introduced in the narrative or inferred by the readers themselves. For example Myers et al (1994: 876) find that, when a character initially presented with the behavioural trait of vegetarianism orders a cheeseburger, readers take longer to read the relevant sentence than one where this inconsistency is absent. In another study (Carreiras et al, l996: 639), readers ascribe stereotypical gender information to a story featuring a nurse or doctor, and similarly take longer to read a sentence which shows that their assumption is wrong (e.g. when a baby-sitter is referred to as male).

This dominance of the protagonist is reflected in Zwaan & Radvansky's own 1998 version of the situation model, which includes information about entities (protagonists and objects) and their properties (such as physical and mental attributes):

> Entities correspond to the people, animals, objects, and ideas that stand in relation to one another in a situation. These entities are represented by tokens in a situation model. Associated with each of these tokens are the properties of that entity. Typically, these properties are most relevant for understanding the situation. Properties can include such things as the entity's physical appearance or state, the intentions or goals of the entity and the emotions of the entity. Like relations, entities and properties are included in a situation

> model only when they are central to a person's understanding of the situation. However, the entity central to the situation model, the protagonists, is an obligatory part of the representation.
>
> (*op cit*: 36)[3]

Nor is the only function of a situation model to help with understanding a situation: the coherence it promotes arguably has a beneficial effect on memory performance. Enhanced understanding of an event, Zwaan & Radvansky suggest, leads to the improved encoding in memory of the information involved, which should therefore be easier to remember later on. For example, a number of studies have found that, when people were asked to remember a story they had read earlier, the goal-related information was recalled better than other information not related to the protagonist's intentions. 'In general, an increased number of such connections increases the probability of recall, except [in cases involving] very high levels of interconnectivity when recall may suffer because information is so interconnected that readers perform fewer elaborations on it and, therefore, remember it less well' (*ibid*: 23).

As mentioned earlier, most of the research on situation models has been carried out on comprehension rather than production. But it seems obvious that a theoretical structure devised to characterise a Hearer's internal representations of a situation should also be at least broadly applicable to those of a Speaker: both to his internal representations of the situation in which he and an interlocutor are taking part and, thence, to the representations that he consequently tries to create and convey to the Hearer. Although Zwaan & Radvansky do not stress the directly-experienced (as opposed to verbally communicated) type of situation, Johnson-Laird's original concept of situation/mental model certainly includes 'real world' phenomena in its scope as well as verbal descriptions of them:

> If you know what causes a phenomenon, what results from it, how to influence, control, initiate, or prevent it, how it relates to other states of affairs or how it resembles them…. then to some extent you understand it. The psychological core of understanding, I shall assume, consists in your having a "working model" of the phenomenon in your mind. If you understand inflation, a mathematical proof, the way a computer works, DNA or a divorce, you have a

[3] When associated with protagonist entities, the wide-ranging nature of such properties invites speculation on the considerable value of situation models as an aid to understanding situations and reacting appropriately to them at a pre-conscious level; see Section 6.5.

> mental representation that serves as a model of an entity in much the same way as, say, a clock functions as a model of the earth's rotation.
>
> (*op cit*: 2)

Similarly, Zwaan & Radvansky point out that some of the dimensions involved in the comprehender's mental model – time, space, and protagonist – are known to feature in autobiographical accounts of *directly* experienced events. (In one such, Wagenaar's 1986 account of his own recall of events over a period of six years, the dimensions used were protagonist, time, space, and the nature of the event itself. Varying these as retrieval cues produces varying results, with cuing along multiple dimensions usually producing better retrieval results than supplying cues singly. However, *what* as a single cue is very powerful, though *when* 'was almost useless' (Wagenaar 1986: 241).) They also briefly discuss the cueing choices made by writers and speakers to indicate to readers/hearers what to incorporate in a situation model, and point to foregrounding techniques such as clefting as offering some assistance here.[4]

Extending this existing theoretical framework to speech production could make a valuable contribution to theories of the Speaker. But such an extension would raise some important questions. For instance, are the situation dimensions that the theory posits (space, time, protagonists, causality, intentions) for comprehension the same as those used in production, and if not, how do they differ? How does the Speaker create the relationships between these dimensions so as to optimise his communicative success? How does he make sure that he and his interlocutor are, in Pickering & Garrod's words, 'thinking about' the same thing as they negotiate a spoken (or written) text? And, again, how does he do it all *so fast*?

Here it is worth considering the other sort of situation in which Speakers find themselves. This is the one in which they are pursuing their communication goals in

---

[4] While clefting is a useful foregrounding device, there are others: intonational prominence, for example, or word-order variation. The second of these is illustrated by the difference, described by Sperber & Wilson (1986/1995: 202), between 'It rained on Monday' and 'On Monday it rained'. Overall, as Sperber & Wilson's discussion suggests, three general strategies appear to be available to speakers who want to emphasize a particular point. The first is that it is 'natural' (*ibid*: 203) for given (uncontroversial) information to come before new, with focal stress falling towards the end of the utterance. The second is to use stress as a vocal equivalent of pointing: 'a natural means of drawing attention to one particular constituent in an utterance'. The third is to phrase the element of an utterance to be emphasised in a way that answers a *Wh-* question. Thus, 'It rained on Monday' would be an answer to the question 'When did it rain?', while 'On Monday it rained' would not.

the presence of their real-world Hearer and an array of information derived from the real-world environment, as opposed to a verbal description.

It is at this point that research explicitly focused on dialogue enters my narrative.

2.3 The Speaker in context: common ground and audience modelling

The inclusion of a real-world Hearer in a Speaker's situation brings with it a crucial factor not yet covered in my discussion: the question of that Hearer's requirements. In order to advance his own goals in a particular speech context, what does a Speaker need to tell the Hearer? What does the Hearer need to know – what, indeed, is he in a position to understand? To successfully answer these questions, Speakers must engage in a slightly different sort of modelling: one that processes, not the incoming information that presents itself to his mind, but the information he will transmit to his audience. In this section, I discuss some of the theoretical and empirical work carried out in the field of audience modeling.

2.3.1 Language use as a collective activity

The work of H. H. Clark and his associates is of major importance here. Indeed, its contribution to the study of communicative interaction can scarcely be overestimated, focusing as it does on the dynamic relationships between the interlocutors, their contexts, and their goals and intentions: the conception of language use as a human activity, a feature of linguistic performance, as opposed to the Chomskyan tradition of studying language as a product of human linguistic competence

As described in his seminal 1992 text, Clark bases his approach on the view that speaking and listening are collective activities: indeed, parts of the *same* activity. 'When two people use language, it is like shaking hands, playing a piano duet, or paddling a two-person canoe. It takes coordination, even collaboration, to achieve. Speaking and listening are two parts of a collective activity just as surely as playing the two parts of the piano duet are two parts of a collective activity' (Clark 1992:

xvi). And, he adds, many essential aspects of these two parts of the activity will emerge 'only when we study two people talking to each other' (*ibid*: xvi)

This approach to verbal interaction is inescapably linked to the notion of situation, conceived along the dimensions used in situation model research: who, when, where, why[5]. As he points out, it is precisely the speaker, time, place and circumstances of an utterance that distinguish the study of language use from the study of language structure. But one of the concepts most closely associated with this approach is that of 'common ground': a body of participant-shared information that acts as a dynamic background to a conversation, and without which language use as a collective activity could not take place.

According to Clark, the common ground between two interlocutors is the 'sum of their mutual knowledge, mutual beliefs, and mutual suppositions.' As the discourse proceeds, each utterance contributes further information to the common ground, and this is part of the process that Clark refers to as 'grounding', whereby speakers design their utterances so that their addressees can 'readily identify what is to be added to that common ground' (*ibid*: 3-5). Crucially in the context of my own inquiry, a feature that has become closely associated with this process is that of audience design, by which a Speaker takes a Hearer's capacities, interests and situational and encyclopaedic knowledge into account when constructing utterances.

2.3.2 Grounding: the heart of Speaker-Hearer collaboration

As an example of the grounding process at work, Schober & Clark (1989) describe experiments in which pairs of students take part in picture-matching games: one, the 'director', has to describe a set of abstract diagrams (tangrams) to the other, the 'matcher', who has to arrange them in a required sequence. During these conversations, the method of describing each diagram changes. The first time a diagram is referred to, the director describes it in some detail; later, they only use a condensed version of this description, so that the descriptions themselves get shorter

[5] This focus is, of course, essential in all professional and everyday communication, and therefore is by no means confined to the research lab. Nor is the suggestion that the crucial aspect of any situation model is the protagonist. As the author Tobias Hill comments (The Times, 3.10.09): 'Humanity is where the action is and …, while readers may enjoy plot or description, what people really want to read about is people.'

and shorter as the trials continue. At the same time, the number of exchanges needed to identify a diagram reduces considerably, as demonstrated by the following (asterisks show overlaps):

> The first time one pair saw one of the figures…, the two of them had this exchange:
>
> D: Then number 12 . is (laughs) looks like a, a dancer or something really weird. Um . and, has a square head . and um, there's like there's uh- the kinda this um .
> M: Which way is the head tilted?
> D: The head is . eh- towards the left, and then the-an arm could be like up towards the right?
> M: Mm-hm.
> D: *And . It's - *
> M: *an- . a big* fat leg? *You know that one?*
> D: *Yeah, a big* fat leg.
> M: and a little leg
> D: Right.
> M: Okay.
> D: Okay?
> M: Yeah.
>
> By the last trial, the reference was more compact, and the two of them took only one turn each:
>
> D: Um, 12 . the dancer with the big fat leg?
> M: Okay.
>
> (Schober & Clark, 1989: 216-7)

As Schober & Clark point out, this final reference combines the perspectives of both the director and the matcher. (It is of interest, too, that what turns out to be the definitive element in the description – 'big fat leg' – is offered by the experiment's formally-designated Hearer: the 'matcher'.) They contrast the process involved – a 'collaborative' one, in which speakers and hearers work together moment by moment to try to ensure that what is said is also understood – with what they term the 'autonomous' view of comprehension. According to this less dynamic model, speakers and hearers listen to the words uttered, decode them, and interpret them against what they understand to be the common ground of their interlocutors. : a

process that, though involving the concept of common ground, lacks the continuously cumulative element that is the essence of the grounding process.

For Schober & Clark, grounding is 'the heart of collaboration' and is basically an opportunistic process. It succeeds in part by exploiting adventitious commonalities between speakers and addressees: 'In our task, A offers one way of viewing a figure – say, as a whale – and if B happens to be able to see it that way, he accepts it, and they go on. If he cannot see it that way, the two of them try another perspective. The process is opportunistic in that it takes advantage of the first perspective A and B find they can agree on' (*ibid*: 229)[6]. Thus, they conclude, the social process of interacting in conversation plays a central role in the cognitive process of understanding it and, they suggest, these findings extend to other types of exchange, a point that is illustrated below.

As a theoretical construct, common ground – 'mutual knowledge, mutual beliefs, and mutual suppositions' – replaces the earlier notion of 'common knowledge', or 'mutual knowledge'. This had been rejected by Sperber & Wilson (1982, 1986/1995) as psychologically unrealistic since it leads to an infinite regress: Speaker and Hearer mutually know P iff S knows P, H knows P, S knows that H knows P, H knows that S knows P, S knows that H knows that S knows that P… and so on ad infinitum. However, the appeal to common ground does not in fact solve the problem of infinite regress, since it does not eliminate the appeal to mutual knowledge, and merely introduces further regressive notions such as 'mutual belief' and 'mutual supposition'. (Clark himself recognises the problem and attempts to avoid it by stressing that people's attributions of mutual knowledge or beliefs are 'simple inferences based on certain evidence and assumptions. They are *not*, repeat *not* [Clark's emphasis] an infinitely long list of statements' (*ibid*: 5-6).)[7] I will show in Chapter 2.4 that the very notion of grounding – together with that of explicit

[6] The formal goal of these experiments is to investigate differences in communicative success between listeners who are the speakers' intended targets and those who are not (i.e. who are 'overhearers'). The results appear to show that the overhearers, who are excluded from the collaboration permitted by the continuous grounding process, are at a disadvantage compared to the experimental 'matchers' , who can negotiate meanings and perspectives with their interlocutors. 'Understanding can only be guaranteed for listeners who actively participate in establishing these perspectives' (*ibid*: 230).

[7] The solution proposed by Sperber & Wilson (1995) was to introduce a notion of 'mutual manifestness' which they claim is weaker than 'common ground' or 'common knowledge' in just the required way.

audience design – has been challenged, because of the heavy cognitive load it is assumed to entail.

But whatever the controversy surrounding the theory of grounding itself, data that has been collected in its support throws helpful light on what interlocutors actually do in order to handle feedback and jointly negotiate their contexts with success; some examples are given in the following sub-section.

2.3.3 Fast, complex, effortless: audience modelling in action

The shared knowledge, beliefs and assumptions that interlocutors bring to their exchanges can be extremely varied. It may be based on extensive interpersonal history, or on very little; on detailed knowledge of individual interlocutors, or on stereotypes or general encyclopaedic knowledge. In conversation with strangers, stereotypes and general encyclopaedic knowledge naturally dominate, but communication may be no less effective for that. Thus, for example, Bostonians who in a 1968 study were asked, 'I'm from out of town, can you tell me how to get to Jordan Marsh', replied more fully – and presumably more helpfully – than those who were simply asked, 'Can you tell me how to get to Jordan Marsh?' (Kingsbury 1968, quoted by Krauss & Pardo, 2004: 203). Nor did the 'I am a stranger' message have to be explicitly communicated; Bostonians asked simply 'Can you tell me how to get to Jordan Marsh?' replied in greater detail if it had been asked in a non-local dialect.

Even such brief exchanges as these demonstrate that an obvious, but entirely fundamental, prerequisite for communication success is monitoring: not only the self-monitoring posited by unilateral accounts of speech production such as Levelt's, but the monitoring by one interlocutor of another, and of the world around them. As Clark and others have shown, such monitoring is carried out continually, and draws on signals from a variety of sources: voices, faces (especially eye gaze), bodies, the proximal environment (e.g. the interlocutors' joint workspaces), and the distal environment (e.g. a scene both interlocutors are watching). As I will show below, there is also evidence that the nature of the interaction dictates which of these sources take priority, revealing both the precision and the flexibility that Speakers bring to their role.

A detailed example of interaction between a physical situation and the communication that takes place in (and about) it is offered by Clark & Krych's (2004) workspace study. The procedure here follows a broadly similar protocol to the Schober & Clark experiments**:** participants are grouped in pairs – 'directors' and 'builders' – and directors are asked to tell builders how to construct 10 simple Lego models. The pairs work under four different sets of interactive conditions: participants can/cannot see into the builder's workspace (*workspace visible* vs *workspace hidden)*, and they can/cannot see each other's faces (*faces visible* vs *faces hidden*). In a fifth, non-interactive, condition, directors merely record their instructions and builders follow the recordings later.

The results show that participants complete their tasks more quickly when the workspace is visible than when it is hidden, and that they take longer, make more errors, or both, if the speakers cannot monitor their partners at all. When participants can see into the builder's workspace, mean building times are 'much shorter'. Conversely, where builders have only pre-recorded directions to follow, they have major difficulties: 'When monitoring was precluded, builders made eight times as many model errors [where the constructed model did not completely match its prototype] and 14 times as many block errors [where a Lego block was the wrong colour, size, or in the wrong place]… The most accurate pair in the non-interactive condition was only as good as the least accurate pair in the interactive condition' (Clark & Krych, 2004: 69).

In this case, monitoring each other's faces does not make any significant difference to the results, an interesting outcome that Clark & Krych ascribe to the fact that the interaction is task-oriented. 'Although people do make use of eye-gaze and head gestures when visible,' they conclude, '… this did not lead to measurably greater efficiency in our task. Monitoring the addressee's workspaces, on the other hand, *is* critical, and in our task, preventing it doubled the time needed. And preventing all monitoring of others led to eight times as many errors' (*ibid*: 76).

Participants' grounding techniques that account for these 'dramatic' differences in outcome include visual monitoring (mainly by directors), gesturing (particularly by builders, querying whether they have understood a director correctly), building

actions, and building 'postponements' where, say, a builder hesitantly holds a block over the model. In this last case, Clark & Krych say, 'builders appeared to use these to signal that they had too little information to proceed, and in every case, directors responded with more information' (*ibid*: 72). Meanwhile, as grounding theory predicts, both interlocutors show themselves skilled opportunists in their Speaker role, changing course in mid-speech to take advantage of openings offered by the Hearer's gestural acts and other visible actions and, typically, initiating such changes within half a second of the opportunities becoming available.

In another task-based example, by Kraut et al (2003), a similar pattern of sensitivity to the interlocutor's situation is demonstrated, but observed in terms of specifically linguistic performance. Here, the task is to repair a bicycle, and one focus of the experiment is the extent to which interlocutors use deictic expressions to refer to tools and other work objects during their activities[8]. Each repair session is carried out by two people, a 'helper' who guides operations, and a 'worker', who executes them. The participants operate under three sets of conditions: working side-by-side, in each other's physical presence; working in separate rooms, connected by an audio link; and working separately but connected by an audio-visual link. The results demonstrate that, in the side-by-side condition where participants can see both each other and the work objects, both use deixis frequently. In the audio condition, where helpers cannot see operations, neither interlocutor uses it. In the audio-visual condition, however, this symmetry does not obtain. Summarising this result in 2004, Fussell & Kraut explain that 'here, the helpers can see the workers and work space but cannot point to objects in it. Under these conditions, helpers rarely used deixis. However, workers can point to task objects and they know that helpers can see them do so through the video link. They use deixis instead of matching the helpers' nondeictic expressions… In short, the way workers referred to task objects and locations depended upon what their partners could see, not the language their partners previously used to refer to these same objects and locations' (Fussell & Kraut, 2004: 197).

In this experiment, it is the speakers who appraise their interlocutors' physical situations and modify their utterances accordingly. But the reverse – the partner-

[8] E.g. 'The derailleur is actually hanging down on this side… Right there', as opposed to 'The derailleur is hanging down by its cable… Off the left hand side of the bike' (Kraut et al 2003: 37).

specific sensitivity of hearers – has also been observed in a shared task setting where, as noted by Brennan & Metzing (2004), addressees interpret the same utterance differently when it is spoken by different speakers with whom different dialogue histories are shared. In Metzing & Brennan's experiments of 2001/3, the (confederate) Speaker tells an Addressee to reposition a large set of objects: a process that they complete several times, developing shared perspectives and terms for individual objects (e.g. *the shiny cylinder*). Metzing & Brennan propose that this lexical entrainment, or convergence on similar referring expressions, involves a 'conceptual pact' – a temporary, flexible agreement to view an object in a particular way (Metzing & Brennan 2003: 201). The Speaker then leaves the room and either returns or is replaced by a new Speaker, after which the repositioning task is undertaken for the last time:

> In the final trial, the new or old speaker used either the familiar term or a new, equally good term (e.g. *the silver pipe*) for the same critical object (amid many other references that did not use different terms). Addressees gazed immediately at the object when either speaker used the old term. However, when the old speaker used a new term (inexplicably breaking a conceptual pact), addressees experienced interference, delaying gazing at the target object. There was no such delay when the new speaker used the new term.
>
> (Brennan & Metzing, 2004: 192)

This hesitation, they suggest, indicates that the representations in memory from which entrainment emerges encode a cue that is partner-specific, which leads addressees to expect that a speaker should continue to use agreed expressions where no contrasting meaning is implicated.[9]

Overall, as Brennan & Metzing conclude (*ibid:* 192), such immediate effects provide evidence of 'impressive agility and potential for partner-specific processing in the language processing system'. And, notwithstanding the importance demonstrated above of visual monitoring during dialogue, evidence from dialogic situations where physical co-presence is by definition excluded also shows signs of such agility. For example, telephone interlocutors – deprived though these are of the visual cues that

[9] Relevance Theory would endorse this claim, though without appealing to 'pacts'. When the old speaker uses a new term, he is causing the hearer gratuitous processing effort unless he means to achieve extra, or different effects – hence the need for the hearer to search for these, and the delay in his reaction. In contrast, when a new speaker uses the same term, any extra effort that it placed on the hearer would be seen as accidental, and therefore no search for additional effects would take place.

participants in practical tasks find vital – nonetheless interact effectively through the systematic use of techniques such as back-channels (e.g. *Okay*, *All right*, *Uh-huh*). Results of research by Bangerter, Clark & Katz (2004) suggest that conversational 'project markers'[10] like these are employed in at least two distinct ways. The terms *Uh-huh*, *yeah*, or *right* are typically used to respond to new contributions to the body of a conversation, by acknowledging them or displaying agreement. In contrast, *okay* and *all right* are used primarily to enter and exit projects, such as moving into the body of a call or closing a side sequence. These words, they argue, are all specialized components of a system of contrasts that enables interlocutors to navigate joint discourse projects: '*Okay* and *all right* are specialized for vertical navigation – entering and exiting joint projects. *Yeah*, *uh-huh* and *right* are specialized for horizontal navigation – continuing within joint projects'. (Bangerter et al, 2004: 20)

These and other detailed accounts of dialogic situations give abundant evidence of how complex, how speedy, and yet how apparently effortless for the Speaker is the process of inferring common ground, and thus adapting speech to the needs of a specific situation, audience, or Hearer. But the evidence, as well as providing answers to some questions, raises others. True, it gives some insights into the mechanisms whereby a Speaker and a Hearer succeed in 'thinking about' the same things as they pursue their interaction, but the issues of speed and flexibility are not directly addressed. Perhaps more serious still, consideration of these issues raises an additional problem, that of cognitive load. It appears that interlocutors who are continually mindful of their opposite numbers' needs – and who are therefore successful communicators – have to deal with an enormous amount of data on a second-by-second basis. How do they do it? And, indeed, do they actually do it? It is at this point that the contribution made by Pickering & Garrod's 2004 proposal for a 'mechanistic psychology of dialogue' needs to be assessed.

[10] Bangerter et al equate conversation with project-oriented action: 'In conversation, the participants do not just speak – they do things together… To understand what people are doing in conversation, one must understand the joint activities [i.e. projects] they are engaged in' (*ibid*: 1). Furthermore, the social encounter is itself a sequence of joint projects, such as 'entry', 'body' and 'exit'.

2.4 Pickering & Garrod's model of dialogic interaction

Pickering & Garrod's model of dialogic interaction, like Levelt's speech production model before it, makes a major contribution to my own inquiry. It enables an examination of the role and capabilities of the Speaker to acquire a degree of structure that is in some ways comparable to that of Levelt's model, but which takes on an extra dimension: Speaker and Hearer are here considered, not separately, but together, as a single unit, thus replicating the physical reality of the dialogic situation.

For this reason, I will discuss Pickering & Garrod's major exposition of their theory (2004, with further exposition in 2007 and 2009) in some detail, starting with a discussion of their central claim.

2.4.1 Alignment of situation models: intentional or automatic?

Their central claim is that, in dialogue, the goal of the participants – even when engaged in irreconcilable argument – is to align their mental states, so that they are talking about the same things in the same way: 'Communication,' they suggest, 'is successful to the extent that communicators come to understand relevant aspects of the world in the same way as each other. In other words, they *align* [authors' emphasis] their representation of the situation under discussion… Dialogue is a form of joint action in which interlocutors have the goal of aligning their understanding of the situation (whether they agree about every detail or not)' (Garrod & Pickering 2009: 294-5).

This process of alignment is seen as taking place both intentionally (dialogue as a whole, they argue, 'constitutes a form of intentional joint action' (*ibid*: 295)) and automatically, below the level of consciousness, and it takes place at many levels of speech and comprehension. The topmost level – broadly corresponding to Levelt's Conceptualizer stage – is that of the situation model, as described by Zwaan & Radvansky (cf Chapter 2.2), and this, Pickering & Garrod suggest, can be applied just as readily to the case of the Speaker as of the Addressee. Thus, whatever role in a dialogue interlocutors are momentarily occupying, each is processing a similar mental representation that involves the same entities (the same people and objects)

and – providing these seem the 'most relevant for understanding the situation' (Zwaan & Radvansky, 1998: 36) – the same properties of entities, ie physical and mental attributes. As a consequence, they will also have the same idea of these entities' goals and intentions and (though this particular process still remains largely unexplained) they will place them in the same time frame. These similarities increase as a dialogue proceeds and more information is given, negotiated, and absorbed by both parties on the mental models they hold.

In this way, the situation models of the participants become metaphorically intertwined, so that the conceptual representations that they call on in their conversation become joint representations, in the sense that access to them is *shared*. Pickering & Garrod's claim is that it is this automatic alignment of situation models – rather than the computation-heavy process of grounding – that enables communicative success.

Describing this interaction model in their 2004 paper, Pickering & Garrod write:

> Most work on situation models has concentrated on comprehension of monologue (normally, written texts) but they can also be employed in accounts of dialogue, with interlocutors developing situation models as a result of their interaction… More specifically, we assumed that in successful dialogue, interlocutors develop aligned situation models. For example, in Garrod & Anderson [1987: research on dialogue models developed by players in a maze game; described in Chapter 2.4.3], players aligned on particular spatial models of the mazes being described. Some pairs of players came to refer to locations using expressions like *right turn indicator, upside down T shape, or L on its side*. These speakers represented the maze as an arrangement of patterns or *figures*. [Others] aligned on a spatial model in which the maze was represented as a network of *paths* linking the points they describe to prominent positions on the maze (e.g. the bottom left corner). Pairs often developed quite idiosyncratic spatial models, but both interlocutors developed the same model.
>
> (Pickering & Garrod, 2004: 172)

This account of communication as an inter-dependent, emergent activity, supported by automatic as well as conscious processes, represents a major advance in the development of theories of the successful speaker, and of successful interactive communication as a whole. The key difference between what Pickering & Garrod – like Schober & Clark earlier – term an 'autonomous' model of communication such

as Levelt's, and their own interactively aligned one lies in the process whereby an utterance is passed from its originator to the Hearer. In the Leveltian model, this takes place only when its final – phonetic – version crosses the gap between the Speaker and the Hearer, who processes it and uses the results to finally infer its originator's (presumed) communicative intention. As indicated by this description, the scope for error and ultimate misinterpretation is obvious. Pickering & Garrod's own model, in contrast, shows the interlocutors sharing their communications, not just at the phonetic level but at <u>all</u> levels simultaneously, including that of the situation model: what the pair are 'thinking about' as they successfully create and comprehend utterances. Here, in dialogue, production and comprehension processes are closely coupled in a largely automatic fashion:

> In formulating an utterance the speaker is guided by what has just been said to him and in comprehending the utterance the listener is constrained by what the speaker has just said… The interlocutors build up utterances as a joint activity (Clark 1996) with interlocutors often interleaving production and comprehension tightly. They also align at many different levels of representation… Thus, in dialogue *each* [my emphasis] level of representation is causally implicated in the process of communication and these intermediate representations are retained implicitly. Because alignment at one leads to alignment at others, the interlocutors come to align their situation models and hence are able to understand each other.
>
> (*ibid*: 176)

Pickering & Garrod base their argument on the following six propositions (*ibid*: 172), of which Numbers 2 and 3 introduce the issue of automaticity:

1. The alignment of situation models forms the basis of successful dialogue
2. Situation model alignment is achieved via a 'primitive and resource-free priming mechanism'
3. Alignment at other levels of representation – such as lexical and syntactic – is also achieved by this priming mechanism
4. Interconnections between the levels mean that alignment at one level leads to alignment at others
5. A further 'primitive mechanism' – verbalised reformulations, or clarification requests – allows interlocutors to repair representational misalignments interactively; and

6. It is only when these primitive mechanisms fail to operate correctly that there is a need for 'more sophisticated and potentially costly strategies that depend on modelling the interlocutor's mental state' – a reference to 'common ground' theory, which Pickering & Garrod term 'full common ground', and reject by reason of the cognitive costs involved.

An additional proposal concerns the connection between the interactive alignment process and the use of routine or semi-fixed expressions (e.g. 'break the ice') which, by short-circuiting language decision-making, simplify both production and comprehension. Pickering & Garrod dispute the assumption that speakers (unlike hearers, who have prior contexts to help them 'short-circuit' the comprehension process) have to work their way through every level of the production process with every utterance. 'It is logically just as possible to avoid levels of representation in production as in comprehension. Although we know that a complete output normally occurs in production, we do not know what has gone on at earlier stages. Thus, it is entirely possible, for example, that people do not always retrieve each lexical item as a result of converting an internally generated message into linguistic form …but rather that people draw upon representations that have been largely or entirely formed already' (*op cit*: 181). I will return later (see Chapter 6.2.1) to their discussion of routines – linguistic expressions that are more or less fixed – and routinisation, or routines that are set up 'on the fly' in the course of a dialogue, and therefore become fixed for the duration of the interchange.

The following subsections consider four important aspects of Pickering & Garrod's 2004 study: its methodological novelty; its experimental foundations; the nature of the psychological mechanism that it proposes; and its limitations, together with later modifications and developments.

2.4.2 Accounting for natural dialogue: a 'central goal of psycholinguistics'

Pickering & Garrod's approach, as described above, breaks significant new methodological ground on two main counts. The first is its focus, not merely on the psycholinguistics of dialogue, but on dialogue approached from a standpoint that combines two psycholinguistic traditions that have hitherto been separate.

The most natural and basic form of language use, Pickering & Garrod argue, is dialogue. Therefore, a 'central goal of psycholinguistics' should be to provide an account of the basic processing mechanisms used during natural dialogue – an account which, they say in their 2004 paper, has so far been lacking. One reason for this omission has simply been a practical one: naturalistic dialogue is assumed to be too hard to study, given the amount of experimental control needed. The other is deeper-seated and, they contend, lies in a theoretical gap between traditional psycholinguistics, derived from the decontextualised data used in generative linguistics, and the 'language-as-action' focus of Clark and others, in which speech is studied as the vehicle of interlocutors' goals and intentions. The first, almost by definition, deals with 'language-as-product' expressed in isolated sentences: in other words, with monologue.[11] The second, though it deals with dialogue, studies it within the context of examining the communicative and social strategies of the interlocutors, rather than the basic processing mechanisms involved in their linguistic interaction.

This dichotomy, Pickering & Garrod say, 'need not be. The goals of the language-as-product tradition are valid and important, but researchers concerned with mechanisms should investigate the use of contextualised language in dialogue' (*ibid*: 170). It is to bridge this theoretical gap that they propose an approach that considers from the viewpoint of psycholinguistics the process whereby interlocutors successfully communicate with each other, switching quickly and repeatedly between Speaker and Hearer roles as they do so.

A second important departure from previous practice consists of Pickering & Garrod's extension of situation model work. While, as stated above, most research on situation models has concentrated on monologue (and on the Hearer /Reader's role in it), Pickering and Garrod argue that the concept can also be applied to the development of shared mental models – and of the shared ideas of time, space, reference, causality and intention, of which these are composed – that take place when interlocutors engage in the role-exchanging, turn-and-turn-about practice of

[11] 'Dialogue', Pickering & Garrod comment, 'sits ill with the competence/performance distinction assumed by most generative linguistics …because it is hard to determine whether a particular utterance is "well-formed" or not (or even whether that notion is relevant to dialogue)' (*ibid: 170*).

dialogue. As a case in point, they cite Garrod & Anderson's 1987 maze game experiment, mentioned earlier.

2.4.3 Talking your way through the maze

Garrod & Anderson's experiment represents an attempt to explore naturalistic dialogue in a controlled way. The control lies in the context selected for the interchange: it is heavily restricted and consists of an electronic maze of boxes and pathways, viewed on Video Display Units (VDUs) by two players in separate rooms. The players have to move through the maze towards a goal and, as they do so, they have to describe their locations to their partners.

The record shows that the interlocutors start by using a variety of descriptive formats – for example, four ways of describing a particular box in the maze are:

a) 'See the bottom right, go two along and two up. That's where I am.'
b) 'I'm on the third row and fourth column.'
c) 'I'm at C4' [Here, the speaker, working from the right, has adopted the letters A-F to denote vertical node lines and the numbers 1-5 horizontal lines; perhaps surprisingly, this is a comparatively common solution.]
d) 'See the rectangle at the bottom right, I'm in the top left-hand corner.'

(Garrod & Anderson, 1987: 189-190)

As these examples show, participants initially conceive of the maze's spatial and functional organisation in different ways: they have different mental models of it. For some, it represents a set of links ('pathways') between nodes; for others, it is a collection of rows and columns. Others again convert it into a concept analogous to a map, capable of being described in an abstract code, or else see it as a set of discrete visual patterns, such as T-shapes, squares or rectangles. As the partnership develops, however, there is a tendency for the interlocutors to use the same descriptive format (e.g. both employ the 'path' conception used in Example (a), or the 'coordinates' conception in Example (c)), and this entrainment is progressive, increasing as the dialogue proceeds. These results, Garrod & Anderson conclude, suggest that 'speakers co-ordinate to establish a mutually acceptable form of description and that this process continues over some time, as a dialogue proceeds

from one game to the next (*ibid*: 196)'. In other words, the situation models of the two partners have converged: they are talking about the same thing in the same way because they are *thinking* about it in the same way.

Why, and how, does this happen? As I will explain below, the hypothesis advanced on these points represents a staging-post along the way to Pickering & Garrod's theory of dialogic interaction. Together with this theory itself, they also represent the formal starting point of my own inquiry.

Garrod & Anderson's 1987 suggestion derives from grounding/collaboration theory. Such convergence, they propose, is linked to the interlocutors' attempt to minimize collaborative effort, with a view to cutting down on the time and effort spent on ensuring mutual intelligibility. In the context of their maze game experiment, this

> … can be achieved by following a very simple interactional principle, which we believe may be the basis for much of the co-ordinated activity seen in dialogues in general. We will call this principle *output/input coordination*, and it may be simply stated as one of formulating your output (i.e. utterances) according to the same principles of interpretation (i.e. model and semantic rules) as those needed to interpret the most recent relevant input (i.e. utterance from the interlocutor). In effect, such a principle assumes that speakers should be locally consistent with each other, and so long as both speakers abide by the principle, then the chances are that they will quickly establish a mutually satisfactory description scheme with the minimum of collaborative effort.
>
> (*ibid*: 207)

Is this consistency the outcome of some agreement explicitly negotiated between the interlocutors? On the whole, Garrod & Anderson state, it is not – and, in cases where partners do explicitly negotiate a common conceptual/semantic scheme, they 'very often do not stick to it for long' (*ibid*: 206). The conclusion is, therefore, that the output/input principle ('Co-ordinate your output with the most recent relevant input') operates at some implicit and – given the speed of the exchanges, possibly automatic – level.

As Garrod & Anderson stress, the operation is both cost-effective and functionally simple: it absolves speakers from having to build up an explicit audience model, and minimises the linguistic resource pool that a Speaker/Hearer has to call on when performing his double role in a dialogue: 'When formulating an utterance the speaker

only has to refer to the same set of interpretation rules as those needed in understanding one on the same topic' (*ibid*: 208). However, little explanation is offered as to the cognitive processes that might support so parsimonious a reference strategy.

With Pickering & Garrod's theory, however, this changes. The core of their proposal for a 'mechanistic psychology of dialogue' is a cognitive mechanism, priming, that appears to satisfy the requirements outlined above for successful real-time dialogue: flexibility, data recall, data selection, overall speed of execution, and – enabling all these functions – automaticity. Citing the convergence of situation models in the maze game as evidence, they suggest that this works via a priming mechanism, whereby encountering an utterance that activates a particular representation 'makes it more likely that the person will subsequently produce an utterance that uses that representation' (*op cit:* 173). Thus, hearing an utterance that activates a particular aspect of a situation model will make it more likely that the interlocutor will produce an utterance consistent with that aspect of the model: an 'essentially resource-free and automatic' process. Nor, they continue, is the priming-to-alignment mechanism confined to semantic and pragmatic choices made by dialogue participants; it operates at the syntactic and articulatory levels as well.

2.4.4 Achieving 'implicit' common ground: alignment through priming

Some of the history and scope of research on the priming process will be examined in detail later in this inquiry (see Chapter 6.3). However, it is the work of Branigan, Pickering & Cleland (2000) that at this point plays an essential role in the narrative, by specifically assessing the ability of interlocutors to co-ordinate, not just their reference strategies, but also the syntactic structures they use during dialogue. Quoting the benefits to both Speaker and Hearer that result from semantic and lexical co-ordination, Branigan et al propose that the same advantages should also hold with respect to other linguistic levels, such as grammatical structure. Furthermore – and very importantly – they demonstrate that the linguistic priming mechanism is bi-directional: participants in a dialogue will display co-ordination of grammatical form by using a particular form if they have heard the other speaker use it. This is demonstrated in a series of experiments in which participants take turns to describe scenes that lend themselves to alternative syntactic constructions (e.g. double object

vs. prepositional object constructions, as in 'Give Bill the book' vs 'Give the book to Bill'). 'We found a dramatic tendency for speakers to produce a syntactic form that they had just heard the other participant use: When the verb remained the same, they produced 55% more syntactically co-ordinated responses than non-co-ordinated responses; when it differed, they produced 26% more co-ordinated than non-co-ordinated responses' (Branigan et al, 2000: B20).

Results such as these mark a significant advance on earlier investigations of syntactic priming, where the effect demonstrated is one-way only (i.e. production of a given grammatical form increased the probability of the spontaneous *production* of the same form; cf Bock, 1986; 360), and they lead to a significant conclusion: 'Our finding…' Branigan et al continue, 'is informative about the nature of the language processing system. It demonstrates that there are shared syntactic representations underlying comprehension and production, and moreover that these representations are activated during spontaneous dialogue' (*ibid*: B22)

Importantly in the context of my own inquiry, a suggested consequence of this sharing process is reduction of speaker effort. As Branigan et al point out (*ibid*: B15), speakers can, by coordinating grammatical form, reduce the computational load associated with the syntactic processing of their contributions: thus, 'when they have a choice between alternative grammatical structures to express a meaning, speakers should tend to use one or other form if the other speaker has just employed that form.' But this is not the only implication of Branigan et al's findings. Owing to the parity of representation that they indicate between the functions of comprehension and production, the way has been opened to devising a model of dialogue in which direct and automatic 'tight coupling' – enabled by priming – can be hypothesized at every level, from the articulatory to the situational, with the result that interlocutor misinterpretation rates are reduced. Furthermore, the difference observed by Branigan et al between the priming effect produced in same-verb and different-verb exchanges (i.e. the stronger syntactic repetition effect when prime and target utterances use the same verb) supports the idea that there is, not just horizontal alignment across the levels, but also vertical alignment between them. 'This thinking,' Pickering explains later, 'illustrates a principle of the interactive-alignment model: alignment at one level (in this case, lexical alignment) enhances alignment at

other levels (in this case, grammatic alignment)[12]. This means that alignment can "percolate up" through the different linguistic levels to the critical level of the situation model' (Pickering, 2006: 736). And this vertical 'percolation', linked with its accompanying horizontal process, can thus ensure that partners use similar language, articulated in similar fashion, to communicate with each other about the same thing, such as the same square in the Garrod & Anderson maze.

A major corollary of Pickering & Garrod's alignment-via-priming argument lies in its impact on the whole theoretical area of mutual understanding in its various aspects: common knowledge, mutual belief/grounding, audience modelling, and mutual manifestness. If, as their model suggests, alignment between two interlocutors takes place automatically, the need for deliberate, conscious audience modelling – a 'decision box' in which a Speaker decides what is most appropriate to say to a given Addressee – becomes substantially diminished. In its place, Pickering & Garrod propose a weaker concept, that of 'implicit' (rather than 'full', i.e. Clarkian) common ground. This, in fully interactive dialogue at least, gives both participants an extensive grasp of their interlocutor's knowledge state because – through the automatic alignment process – it makes information that was accessible to one party accessible to the other[13].

> [It] is effective because an interlocutor builds up a situation model that contains (or at least foregrounds) information that the interlocutor has processed (either by producing that information or comprehending it). But because the other interlocutor is also present, he comprehends what the first interlocutor produces and vice versa. This means that both interlocutors foreground the same information, and therefore tend to make the same additions to their situation models… As the conversation proceeds and more information is added, the amount of information that is not shared will be reduced. Hence, the implicit common ground will be extended. Notice that there is no need to infer the situation model of one's interlocutor.
>
> (*op cit*: 178)

---

[12] They point out that alignment also takes place at the articulation level, with speakers echoing each other's accents, speech rates, and the extent to which they shorten repeated expressions.

[13] Garrod & Pickering, in a later exposition of their 2004 alignment model, further clarify the distinction between the two 'grounding' concepts. Clark, they point out, argues that interlocutors use various *strategies* to accumulate common ground, which 'refers to all the information that both interlocutors believe to be shared by themselves and their conversational partner. This is actually a stricter notion than alignment, which merely refers to the information that happens to be shared' (Garrod & Pickering, 2007: 443).

Normal conversation, Pickering & Garrod continue, therefore does not routinely require modelling the interlocutor's mind, even though it may give the impression of doing so: an effect first noted, they point out, in Brown & Dell's 1987 study of production-comprehension relationships.[14] Instead, the overlap between participants' representations will be big enough for a speaker's specific contribution either to set off appropriate changes in the listener's representation, or to trigger the interactive repair process: 'The listener will retain an appropriate model of the speaker's mind, *because, in all essential respects, it is the listener's representation as well* [authors' emphasis]' (*ibid*: 180)

A further set of implications concerns the ease or otherwise with which Speakers fulfil their role. Such massive priming, Pickering & Garrod suggest, can greatly enhance language production, via the prior activation of relevant linguistic representations (e.g. lexical and syntactic representations). While speakers do not usually just aim to repeat what's been said to them, they will tend – thanks to the activation of their syntactic and lexical representations created by previous utterances – to repeat the syntactic and lexical forms used by an interlocutor, and thus become aligned with him. It follows from this that, in both sentence recall and dialogue, 'very much less choice' needs to be made than in monologue: an effect suggested earlier by Branigan et al. 'The decisions that occur in language production (e.g. choice of word or structure) are to a considerable extent driven by the context and do not need to be a burden for the speaker. Thus, they are at least partly stimulus-driven rather than entirely internally generated, in contrast to accounts like Levelt (1989)' (*ibid*: 183).

[14] In an investigation of how, or whether, speech production is adapted to comprehension, this concludes that, in the main, such adaptation takes place neither because speakers model their listeners' information needs nor because they remember their own comprehension experiences in the same linguistic context. Instead, Brown & Dell suggest, the key mechanism here is a process of conceptual constraint, whereby adaptation takes place because 'analogous processes underly [sic] translating a conceptual structure to a text and deriving a conceptual structure from a text' (Brown & Dell, 1987: 452). According to this listener-independent account, the way a Speaker processes a text as a 'Hearer' (actually, in this case, a 'Reader') influences the way he then reproduces it in speech. An unexpected, and therefore salient, element in the text will feature in the spoken summary, with the *incidental* result that the listener's comprehension may be enhanced by this inclusion.

2.4.5 Challenges and developments

Pickering & Garrod's 2004 model of priming-driven dialogue has not gone unchallenged. One interesting objection that can be raised concerns a potential conflict between Pickering & Garrod's treatment of dialogue production and comprehension as an automatic, low-level 'mechanistic' process involving no inferences about interlocutors' mental states, and a pragmatic approach such as Relevance Theory, which treats utterance comprehension as an inferential process resulting in the attribution of a complex mental state – a Gricean speaker's meaning. In the context of my own inquiry – which is broadly sympathetic to relevance theory but sees the alignment model as making a major contribution – this argument might seem to present a considerable problem; and indeed, Pickering & Garrod's search for an alternative to Clark's broadly Gricean approach was largely based on their desire to avoid what they saw as the costly modelling of interlocutors' mental states.

However, as Sperber & Wilson (2002) point out, the attribution of mental states ('theory of mind') itself can be conceived of in two ways. It might be thought of as a conscious, reflective reasoning process of the type described in Grice's famous 'working-out schema' for implicatures: '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, 1989: 30-31).

This seems to be the type of process that Pickering & Garrod have in mind in their debate with Clark. Alternatively, it might be an automatic, modularised inference process that takes place below the level of consciousness, and as such is relatively fast and effort-free. Sperber & Wilson argue that mental state attributions in general are carried out by an inferential 'mindreading' module with a variety of sub-modules, one of which is specifically dedicated to comprehension. The type of low-level priming processes discussed by Pickering & Garrod are quite compatible with this modular approach to inference, and I will argue that they can contribute to it in worthwhile ways.

The relevance-theoretic approach to comprehension is therefore both inferential *and* mechanistic, in the sense that the inferences carried out by the relevance-guided comprehension module are as spontaneous and automatic as the priming mechanism that supports the alignment model of dialogic interaction proposed by Pickering & Garrod. It is then an empirical question how much of utterance production and comprehension – which is treated on both sides as typically fast, automatic and unconscious – is purely a matter of activation (e.g. by priming), and how much is properly inferential. In what follows, I will focus mainly on the activation side, but I assume, unlike Pickering & Garrod (though in common with most people working in pragmatics), that a substantial element of inference is also necessarily involved.

Other grounds on which Pickering & Garrod's model has been criticised include its assumption of production/comprehension parity of representation, and its apparent rejection of an explicit audience modelling process as the basis for communication. One alternative proposed is that of a radically different method of expressing dialogic coordination, via a grammar formalism involving a common, parsing-based architecture for both comprehension and production (Kempson et al, 2001; Kempson, 2004: 202). And Pickering & Garrod themselves acknowledge limitations to their theory.

They accept, for instance, that alignment on the basis of 'implicit common ground' does not meet all communication requirements, simple and resource-effective though this is claimed to be; as a fail-safe device, they propose the use of full common ground, with concomitant inferencing about an audience, at moments when alignment has failed to clarify a reference. Use of full common ground is also likely when one interlocutor wants to deceive or withhold information from another, or when a Speaker wants deliberately to influence an Addressee in some way. 'It is important to stress that we are proposing interactive alignment as the primitive mechanism underlying dialogue, not a replacement for the more complicated strategies that conversationalists may employ on occasion' (*ibid*: 180).

Another disclaimer is perhaps more fundamental. Although Pickering & Garrod cite a wide range of examples of linguistic priming, they concede that 'fully specified theories of how much priming operates are not available for all levels' (*ibid:*

176). Indeed, the concept of automatic alignment through priming (like that of *automatic* audience modelling, to which it leads) has been another target of criticism. One issue of particular concern is an apparent methodological gap between priming at lower linguistic levels and priming at the level of the situation model. As explained by Warren & Rayner (2004), the objection here rests on the contention that phonological, lexical, and syntactic priming are similar in that representations primed are an 'inalienable part' of a message structure, while semantic priming, though less direct, can also draw on representations that are generally similar. Situation models, however, are different: there is no direct priming channel here as there is between the physical aspects of a message. Nor yet is the looser semantic priming pattern applicable, as 'individuals do not necessarily begin dialogue with similar situation representations. Therefore the priming link between individuals' situation models in the interactive alignment model must be of a different sort from the priming channels between other representations… More research into the details of situation model priming and the complex interplay of priming and inhibition between different levels of representation will be necessary in order to fully specify the operation of the model and to evaluate its ascribed simplicity' (Warren & Rayner, 2004: 211).

A further disclaimer made by Pickering & Garrod in 2004 concerns the upper limit of their model itself. Does the process of alignment that they hypothesize go beyond the level of the situation (e.g. who, what, why, when, etc) to that of content choice itself? Importantly in the context of my own inquiry, they here draw a dividing line:

> Our contention is that interactive alignment (and in particular the automatic alignment channels) affects the structures used in production and interpretation rather than directly determining the *content* [authors' emphasis] of production and interpretation. In other words, we assume that alignment provides an explanation of the *manner* [my emphasis] in which interlocutors produce and interpret contributions. So we propose that alignment channels only produce a direct link between the structures that the interlocutors use in language processing. Hence, the alignment process is automatic and direct, even though it does not determine exactly what the speaker produces (as this depends on his long-term knowledge) or how the addressee interprets what he hears "beyond" the level of the situation model.
>
> (*ibid*: 213)

In later work, Pickering & Garrod rise to these challenges. Interestingly, they later re-label the main mechanism supporting alignment as 'a process of automatic (nonconscious and effortless) *imitation* at different linguistic levels' (Garrod & Pickering, 2009: 296; authors' emphasis)[15]. However, they also suggest ways in which their automatic alignment model could be developed to include both linguistic and non-linguistic mechanisms, and review some accounts (Garrod & Pickering, 2007) of how these may operate. In addition to the priming-driven processes of linguistic alignment described in their 2004 model, they suggest that alignment may take place in a variety of ways: via beliefs about one's interlocutor; via imitation in non-linguistic dimensions (e.g. facial expressions); via tacit agreement between interlocutors (such as the 'conceptual pact' described by Brennan & Metzing; cf Chapter 2.3.3); via feedback; and via physical co-presence. They also explore the extent to which these may be either automatic or 'strategic': i.e. choices explicitly made on the basis of inferences about the extent of common ground existing between a speaker and his interlocutor.

Amongst these potential alignment mechanisms, the first – the appeal to beliefs about one's interlocutor – bridges the gap between automaticity and explicit audience modelling. It also marks a return on Garrod & Pickering's part to the notion of priming or, rather, of the processes that possibly underlie it. No doubt, they say, speakers make some reasoned decisions about how to produce descriptions on the basis of what they believe their audience to know. 'But their choices may also reflect automatic resonance processes … [which] could arise from the prior association of particular types of addressee with particular expressions, the activation of such types of addressee as a result of the interaction, and the subsequent activation of those expressions' (Garrod & Pickering, 2007: 444). Here they cite the work of Horton & Gerrig (2005) on the role played in achieving common ground, not by expressly-formulated communication strategies, but by ordinary memory processes acting on ordinary memory representations.

[15] A re-labelling that may seem to contain a contradiction in terms, as the usual interpretation of imitation involves adoption of the same goals – which is not the case in the 'nonconscious and effortless' process described here of copying form, such as grammatical choice or speech rate (cf also Garrod & Pickering, 2007).

Suggesting that the process of audience design, while important, remains 'woefully underspecified', Horton & Gerrig propose that the explanation of what appears to be common ground is the outcome of a memory process termed 'resonance': a 'fast, passive, and effort-free mechanism in which cues in working memory interact in parallel with information residing in long-term memory' (Horton & Gerrig, 2005a: 10) and which, by a process of association, brings a wide range of information within cognitive reach. And, they continue, where conversational situations are concerned, 'we suggest that other individuals function as highly salient cues to make information with which they are associated ready.'

Thus, apparent instances of calculated audience design on the part of a Speaker may in fact be the outcome of partner-specific memory processes that make information closely connected with an interlocutor accessible during language use. Horton & Gerrig test this claim by comparing the way in which different participants in a matching task are exposed to information about their partners, and their subsequent conversations with them. The results show that speakers were more likely to adjust their utterances to their partners' communicative needs when each addressee could be associated with a unique referential context and, therefore, those in the experimental group 'were able to exhibit more evidence of audience design because the memory representations required for audience design were more readily accessible during the period of time they allowed themselves to plan their utterances' (Horton & Gerrig, 2005b: 139). They also suggest that this resonance-based model can be extended to accommodate, not just the automatic referencing of information held in common by interlocutors, but also the automatic message formation that can take place between them – the words they use: 'We expect that addressees would serve as cues to increase the accessibility of particular formulations of concepts' (Horton & Gerrig, 2005a; 27). Garrod & Pickering (2007) make a connection between this last point and their own argument that references, when repeatedly made to the same objects, become routinized and represented as such in memory. 'It is quite possible that such routines become associated with the person with whom they were established, and that this leads to partner-specificity effects during comprehension and production' (*op cit* : 447). These are intriguing and important suggestions which invite more consideration than they so far appear to have received. One proposal for further developing them is given in Chapter 6.5.

While both Pickering & Garrod and Horton & Gerrig place their work within the context of the so-called 'autonomous' speech production model developed by Levelt, their ways of doing so differ considerably: a difference that is especially notable in the context of my own research. Horton & Gerrig see their work as contributing to the Leveltian model, not – as in Pickering & Garrod's interactive structure – standing in opposition to it. Additionally, and very importantly, they show themselves explicitly aware of the time element involved: of the need to answer the question 'How do speakers do it so *fast*?' 'We suggest,' they say,

> that commonality assessment and message formation are intrinsically involved in the generation of messages that reflect speakers' addressee-relevant knowledge… Our description of these processes is intended to accommodate the fact that many utterances are produced with a time course that makes effortful considerations of partner-specific information unlikely. The challenge, therefore, is to provide a model of conversational common ground that explains how relevant knowledge representations might become available within the time that speakers routinely allow themselves for utterance planning.
>
> (*op cit*: 4)

Their memory-based account, they continue, is offered in the hope of providing 'a starting point' for further specification of message planning more generally.

2.5 Agendas for the future

Horton & Gerrig's comments give one indication as to where future research would be valuable. Meanwhile, other pointers to future work have begun to accumulate as interest in dialogue – and, by association, in how Speakers fulfil their role – has increased. The work of Clark and of Pickering & Garrod has obviously provided a key impetus here, as demonstrated by the persistent recurrence of common ground and alignment as themes in later work. But stimulus is also coming from other aspects of linguistics and, indeed, from beyond it. One of the most notable contributors here is computer science and, in particular, the design and engineering of human-computer interfaces (HCI). As I will show in Chapter 4, work in one of the most advanced (and apparently unrelated) of the HCI areas – virtual

environments – has produced some findings of surprising relevance to the investigation of human-human interfaces. But areas closer to the human-human model, such as computerised natural language systems, are also generating insights and results. In one such study, for instance, Brockmann et al review how far the capacity for interlocutor alignment may differ from person to person, and conclude that 'some people are stronger aligners than others (Brockmann et al, 2005: 1). In another, presented in 2008 to the SIGdial special interest group of the Association for Computational Linguistics, Baker et al explore the role of redundancy in task-oriented dialogue. Their findings, echoing those of Kingsbury 40 years earlier (cf Chapter 2.3.3), show that redundancy increases when communicating with strangers: a result that, they suggest, can be used to improve the communicative efficacy of natural language generation systems.

Overall, trends in dialogue research are developing that may bring together computation-oriented topics (such as virtual environments and even robotics), psychology (with a particular focus on low-level interaction mechanisms, such as those investigated by Pickering & Garrod), and neuroscience. In a review of the field in 2006, Piwek suggests that work here will 'hopefully' provide new data for empirical researchers working on the *how* of human dialogue behaviour, with on-going technological developments throwing light on the way the brains of interlocutors process information (Piwek, 2006: 150). And, as I hope to show in later chapters, these hopes are already being fulfilled.

Piwek's review, which introduces a cross-disciplinary issue of *Research on Language and Computation* focusing exclusively on dialogue, acknowledges that it does not deal with computer-aided enhancements to human-human dialogue. It also points out that, from a computing perspective, the main concern in dialogue research is the investigation of the underlying language-independent mechanisms for communication, abstracting away from the details of specific languages (*ibid*: 147). As I argue above, however, the use to which such research is put obviously does not preclude its application to human-to-human interaction, and Piwek outlines a set of agendas for the future that applies to both linguistic and computing fields. First pointing to the inclusion of neuroscience in the researcher's equipment, he continues:

> The work by Pickering & Garrod on alignment as grounded in imitation strongly suggests that low-level non-verbal signals also play an important role in achieving alignment of situational representations. This requires further research into the integrated use in dialogue of verbal and non-verbal means… Finally, the emphasis on low-level processes in dialogue behaviour leads to models that are more concerned with the central place of emotion and feelings in dialogue behaviour.
>
> (*ibid*: 150-1)

As an example of what is already taking place along these lines, he quotes work by André et al and others that builds on a theory of dialogue that has its roots in sociolinguistics: Brown & Levinson's account of universal politeness patterns in language usage (1978/1987), based on the notion of appealing to interlocutor emotions by avoiding 'face-threatening acts' (FTAs) through a variety of interactional strategies that include indicating either approval of the hearer, or deference to his wishes. This affect-led approach, as André et al explain, can be readily applied to the development of user-computer interfaces that adjust their conversational behaviour to both user (i.e. Hearer) and ongoing context. Thus, a user who is already irritated due to communication mishaps will see a proposal by the (computerised) agent to input a long identification number as a threat, and will react accordingly. 'Knowledge about the causes for the user's emotions,' they continue, 'should guide the selection of politeness strategies. Consequently, the emotional state is a factor that emerges during the interaction and dynamically influences the ongoing dialogue' (Andre et al, 2004: 181).

A spoken language dialogue system devised by André et al combines Brown & Levinson's strategic politeness framework with a cognitive theory of emotions that takes user goals, standards and attitudes into account. They acknowledge, however, that their approach leaves plenty of room for extensions, amongst them the need to take account of the fact that face-threats are 'inherently multi-modal': that is, they involve more than just the spoken word. This is a point that, still within the field of human-computer interaction, is taken much further by Baylor (2009: 3560) in her description of an interface in which an error message is delivered either by a text box or by 'Survey Sam', a realistic-looking anthropomorphic agent[16]. Results of this comparison indicate that the agent-delivered message was significantly more

[16] The distinction between the IT terms 'agent' and 'avatar' is explained in Chapter 4.3

effective than the text box at reducing users' frustration, and also promoted a positive attitude to the experience. 'This suggests,' Baylor comments, 'that delivery of a frustration-mitigating message via an anthropomorphic agent may be more effective than simple text-based feedback'; in other words, the deciding factor in a successful Speaker-Hearer interaction may not be its verbal content at all, but its visual one.

As will be seen, my thesis attempts to develop a corresponding investigation within the field of human-human interaction of these issues of low-level processing, multi-modal effects, and affective as well as cognitive relevance. The role of affect, in particular, is explored in Chapter 3, while Chapters 4 and 5 focus on the part played by non-verbal stimuli and low-level processing in enabling dialogue to be successfully initiated and continued. Chapter 6 returns to the issue of verbal interaction itself.

2.6 Summary and conclusions

This chapter reviews work that has contributed over the last quarter-century to the development of a theory of the Speaker. Its starting-point is Levelt's 1989 analysis of the Speaker as a speech-producing system, in which the information to be communicated is passed through one specialist sub-system to another on its way to becoming articulated output, subject to a continuous process of self-monitoring. I then amplify this initial picture by considering the circumstances surrounding a Speaker: now seen, not as a stand-alone entity, but as someone operating in a situational context. How this context may be structured, both for external analysis and in the Speaker's own mind, is suggested by the situation model work of Johnson-Laird (1983) and Zwaan & Radvansky (1998), discussed in Chapter 2.2. An influential further development of the theme of 'Speaker in context' is contributed by the notion of common ground/grounding – of the accumulation of shared information by interlocutors, to the benefit of their mutual understanding – developed by H.H. Clark and others and described with examples in Chapter 2.3. As I have tried to show, the significance of this and of the notion of audience modelling, with which it is closely associated, lies above all in its focus on verbal communication as an activity that essentially involves, not just a Speaker, but a Speaker and a Hearer: 'Speaking and listening are two parts of a collective activity' (Clark, 1992: xvi).

With this acknowledgement that speaking and hearing are two parts of a single system, the need to consider speech production not as a stand-alone phenomenon but as one that is tightly tied into its most naturally-occurring context – dialogue – is shown to be of very considerable importance.

This need forms the base on which Pickering & Garrod (2004 and later; see Chapter 2.4) construct their own highly influential model of the Speaker-Hearer communication dynamic in dialogue. According to this, interlocutors achieve successful communication by 'aligning' their representations of the situation under discussion, mainly in an automatic, pre-conscious manner: a claim that makes a dramatic extension to audience modelling theory by offering an explanation of how a Speaker can 'model' his audience's needs without the cognitive strain that intentional, strategic modelling seems to imply. The automatic mechanism that supports such alignment, Pickering & Garrod claim, is priming, which operates at all levels (e.g. lexical, syntactic) right up to that of the situation model. They later draw on the work of Horton & Gerrig (2005) to suggest that the potential scope of this mechanism might be extended to include a Speaker's representation of individual types of Hearer as distinctive entities in their own right, each one associated in the Speaker's memory with particular linguistic expressions that the Hearer's very presence activates in the Speaker. Horton & Gerrig themselves suggest that this memory-based approach may be capable of addressing the problem – until this point, to a large degree overlooked – of speaker fluency: producing coherent speech at the rate of two to three words a second (Levelt's figure). They see this work as a 'starting point' for further investigation into the Speaker's message-planning processes more generally.

Further suggestions for future work are discussed in the final section of the chapter, which briefly looks at the recent impetus given to dialogue research by computer science, particularly in areas such as computerised natural language systems. The influence of Pickering & Garrod's priming-based (later modified to imitation-based) model is strongly apparent here, and its focus on low-level interaction mechanisms is acknowledged as an important basis for further investigation. It is also suggested that dialogue modelling could be extended from there to cover the 'central place' occupied by emotion and feelings in dialogue behaviour (Piwek, 2006).

What picture of the Speaker does the above body of work draw? In one sense, it is an expanded version of Levelt's 1989 model, with the additions mainly falling in the model's input zone: the knowledge store comprising the Speaker's 'discourse model, situation knowledge, encyclopaedia etc'. The research that I have described has added considerably to our understanding both of the types of information that are involved here and – crucially – of the processes by which they enable a Speaker, not merely to formulate his communication intentions in audible speech, but to do so in a way that is readily understood by his interlocutor. Thus, the continual accumulation of shared information postulated by common ground theory helps both parties to a dialogue to keep track of their discourse models. The low-level alignment processes described by Pickering & Garrod take this idea further by suggesting how shared access to situation knowledge may be achieved at little cognitive cost to the interlocutors. Meanwhile, the development of the notion of a situation model has clarified how that information may be structured at a conceptual level, and which aspects of it may be most or least salient to its producers and recipients.

At a higher level of generality, the expanded model shows an interesting tension between the idea that Speakers use strategies (as required by the notion of intentional audience design) and that Speakers effortlessly adjust their output to their audience (as achieved via the priming-driven alignment mechanism proposed by Pickering & Garrod). This tension, in turn, broadens the whole debate, as it introduces the last of the three major areas of inquiry mentioned at the beginning of the chapter: the issue of conscious control vs automaticity.

Overall, and at whatever level of generality this revised model of the Speaker is considered, it represents a major advance in our understanding of why people say the things they do. However, it still appears to fall some way short of offering a comprehensive theory of the Speaker – so which way should the narrative turn now? As acknowledged above, obvious areas for further investigation include the nature of the automatic, low-level mechanisms (including affective ones) that support interlocutor alignment, and the relationship between these mechanisms and decision-making operations at a more conscious, strategic level. While the main focus of work here would naturally be on the nature of the alignment operation, it would also

increase insight into the more or less conscious nature of the self-monitoring function that Levelt envisages as being within his Conceptualizer.

My main goal, however, is to find an approach to the alignment issue that goes beyond the limits acknowledged by Pickering & Garrod, and considers the mechanisms underlying speech decision-making at the message level. In the following chapter, I begin this task by considering an aspect of dialogue that seems to have received scant attention in the literature to date: its initiation. What makes one person start talking to another? How much of a conscious decision is involved? And, conversely, why do people sometimes find speech difficult? As I will argue, an examination of these questions leads to a conclusion that puts affect into the foreground of the dialogic alignment model. It also envisages the Hearer as playing, not a passive, 'audience' role in speech production, but a highly active one: at some levels at least, even a controlling one.

Chapter 3

Speech eclipse, speech elicitation: the power of the Addressee

Introduction

According to Pickering & Garrod, dialogue is the most 'natural and basic form of language use' and an account of the mechanisms employed during natural dialogue should be a central goal of psycholinguistics. I therefore begin my own inquiry by discussing examples of naturally-occurring dialogue in action and considering whether they show any indications of low-level, non-verbal mechanisms that contribute to interlocutor alignment at message level. From this initial – informal – survey, one finding in particular emerges: the apparent power, not of the Speaker, but of the Addressee to promote or suppress interaction.

I advance a hypothesis that may account for this effect.

3.1 A universal mode of dialogue

The data discussed in this chapter have, with a few exceptions, been collected in naturally-occurring circumstances by myself. As will be seen, all apart from Example 2 fall mainly or entirely into a category of communication whose nature and boundaries are the subject of some discussion, but which is generally known as small talk, or phatic communion. My reasons for this will be discussed in detail later (see Chapter 3.4), but the main – and most obvious – one is that this category is crucially dialogic. According to Bickmore & Cassell (1999: 1), small talk is 'non-task-oriented conversation about neutral topics whose primary functions are to mitigate face threats, provide an initial time interval in which the interlocutors can "size each other up", establish an interactional style, and establish some degree of

mutual trust and rapport'. While most forms and functions of verbal communication – instructions, explanations, commands, persuasion and so on – are frequently used in talking to oneself in the absence of any other interlocutor, small talk as defined above is not. There is simply no point in using it on oneself.

There are two further initial reasons for taking phatic communion as a starting point. The first is its apparently universal nature and the motivation behind it: as summarised by Abercrombie (1994; 3006) it 'enters the everyday experience of everybody, from the most highly civilized to the most primitive, and far from being useless, this small talk is essential to human beings getting along together at all.' The second is its relatively standardised nature and the fact that certain types of content – typically unchallenging and low in intrinsic relevance – are particularly suited to this use. The phatic use of language, as Žegarac & Clark explain, is possible on any given occasion 'to the extent that the interlocutors' mutual cognitive environment includes some assumptions about the way conversations are usually conducted: how are certain topics usually relevant? What is the social relationship between the communicators? What are the social norms for appropriate linguistic behaviour?' (Žegarac & Clark, 1999: 336) On this account, it is the situation model – the interlocutors' *joint* situation model – that determines whether a certain use is 'phatic', and the parameters of this model that determine its standardised nature.

Furthermore, although the skills and sensitivities involved in phatic communion vary, as a general case it needs no special aptitude or training, no prerequisite level of education[1]. (It seems to be age-proof as well, in that adults of all ages do it; it would be interesting to investigate the age at which children start small-talking, a query raised with some immediacy by Example 4.) And this apparent ease of delivery is arguably a product of its semi-ritualised linguistic and conceptual content. Again as will be described in more detail below, what people say in a phatic situation, and how they say it, both fall within certain boundaries of likelihood; thus, phatic dialogue, though it occurs in a naturalistic setting, contains an element of standardisation that readily allows re-use in a variety of circumstances.

[1] This last is a variable that, in linguistic studies, could have some importance, given that so many experimental subjects may – by practical necessity – be higher education students and therefore not necessarily representative of the population at large in terms of verbal skills.

The standardised nature of phatic communication also makes it relatively easy to spot deviations from the norm, thereby providing information about the norm itself. Perhaps the most extreme examples of such deviation come from attempts at phatic communion with an interlocutor with communication difficulties, as in autism. Example 1, collected by Uta Frith (Frith, 1989.2003: 116-7), illustrates the conversational problems experienced by a teenager with autism, and also – by its absence – demonstrates the mutuality of assumptions and effort on which successful phatic communion depends.

Example 1

Context: Uta, an academic, has just tested the reading skills of Ruth, a 17-year-old student. She now tries to engage the student in conversation:

a) Uta: Now you live in that lovely flat upstairs?
b) Ruth: Yes.
c) Uta: Is that really good?
d) Ruth: It is.
e) Uta: Do you do some cooking there?
f) Ruth: Yes, I do
g) Uta: What kinds of thing do you cook?
h) Ruth: Anything.

As Frith notes later, Ruth 'had a good vocabulary, excellent grammar, and was a superb reader. Yet she was an abysmal partner for small talk' (*ibid*: 128). Not only is the reciprocity required for phatic communion missing – Frith has to develop all the topics herself – but so, apparently, is any sign of the motivation that should underlie it. Ruth's terseness seems to indicate a lack, not of the cognitive or linguistic skills required for phatic language use, but of the affective need that drives it: she appears to have no feelings about either her own activities or the effect she is making on her conversational partner.

An indication of how profound the effect of such interlocutor disengagement can be amongst neuronormal interactants – and for how little cause – is given in an early informal experiment carried out in a standard office setting by Ferguson and described in his 1976 discussion of politeness formulas. Recalling how he

deliberately failed to reply to his secretary's *Good morning* on two consecutive days, he found that the second day 'was full of tension. I got strange looks not only from the secretary but from several others on the staff, and there was a definite air of "What's the matter with Ferguson?"' (Ferguson, 1976: 140). A more complex deviation from the norm – in which it is the subjects themselves who are encouraged to reject the phatic style in favour of an information-oriented one – is examined in an investigation by the Couplands & Robinson, which studies replies from 40 elderly people at a day centre to the routine enquiry *How are you?* The deviation here rests on the fact that the respondents are as likely, on the grounds of their age and consequent health difficulties, to reply to the question literally rather than phatically. However, this encouragement is not enough to make the neuronormal subjects completely abandon their need to size up their interlocutor and establish a comfortable interactional style: their responses emerge as interestingly subtle, showing too much variation and ambiguity to be classified outright as phatic, but characterised in many examples by a 'systematically' phatic orientation. 'Interviewees find many overlapping strategic means to hold back, at least initially, the full force of their negative health experiences… they tend to opt for multiply qualified statements and hedges and rapidly shifting judgments of their own well-being' (Couplands & Robinson, 1992: 225-6).

It is in pursuit of some further types of subtle variation that this chapter's instances of naturally-occurring and informally-captured speech are examined.

3.2 Differing dialogues: observations and intuitions

In this sub-section, I will compare and contrast three examples of naturally-occurring conversation, two of which are phatic, while one is not. In Example 2, which is clearly not phatic, two speakers bring opposing goals to the discussion of a topic of common relevance.

Example 2

Context: A dinner party where four people are present. This exchange takes place at the same time as conversation elsewhere round the table.

a) Speaker 1: Where's the cigarettes, [name]?
b) Speaker 2: Sorry, [name]. I've cut you off. You said you'd had the last one. You promised me the last one was the last one.
c) Speaker 1: Well I want to have one more.
d) Speaker 2: Cost you a buck.
e) Speaker 1: Oh give me a break, [name]!
(collected by Eggins; quoted in Eggins & Slade, 1997: 9)

Example 3 is a typical passage of phatic communication, where the relevance to the two partners lies 'not in the information (linguistically-encoded or otherwise) that the utterance carries, but in the act of the utterance's creation' (Pollard, 2005a: 10). It clearly illustrates natural dialogue's incremental quality: each interactant builds neatly on what the previous one has said, and this pattern continues over a considerable number of exchanges (not given here).

Example 3

Context: AJP, on her way through a public garden, stops to admire the flowers. Across the flower bed, she notices another passer-by, doing the same thing. They look at each other. Their paths through the garden separate, then the passer-by returns into view and approaches AJP to say:

a) Passer-by: Those flowers – they're so pretty there…
b) AJP: Yes, aren't they.
c) Passer-by: Really lovely...
d) AJP: Yes, and those yellow ones at the top of the steps – they smell beautiful, don't they…

(Pollard, *ibid*: 15)

Example 4, also collected in a garden, is far from standard. There are two interactants, but only one Speaker:

Example 4

Context: A group of people are outdoors at night, waiting to see the full moon rise. Among them, standing side by side, are AJP and a 9-year-old boy, who is interested in stars, star maps, and similar subjects.

a) AJP: We're facing north – look, there's the Great Bear. You can use it to find the Pole Star…

(The boy turns his back on her, and wanders away.)

b) AJP *(initially lost for words):* Oh, what's the point…

(collected by Pollard)

These dialogues appear to differ, not in one, but in several ways: indeed, none of them appears directly comparable to the other two. Examples 2 and 4 involve people who know each other; in Example 3, the interactants are strangers. The relevance of the exchanges in Example 2, which is clearly non-phatic, is largely a function of the propositions expressed; in Examples 3 and 4, both of which are phatic in nature, the main relevance lies in the act – or rejection – of communication itself. Meanwhile, measured by standards of effective interaction, the one-sided Example 4 is a failure and the other two are successes.

Despite these differences, however, all three have one characteristic in common. In each example, Hearers – from now on called Addressees, for reasons that will hopefully become clear – make their presence and requirements plainly felt, with results that guide the exchange. In Example 2(b), Speaker Two's protestations are a summary of what s/he thinks Speaker One, now the Addressee, must expect, given their joint past history. In Example 3, the passer-by's choice of topic is a natural response to the situation, given what she has just observed her Addressee doing. In Example 4, AJP draws on her knowledge of her interlocutor to reinforce sociable relations: an error in relevance assessment that is quickly made clear to her.

What connects these examples, therefore, is the decisive impact, not of the Speaker on the Addressee, but the reverse. Counter-intuitive though this conclusion seems, given the natural assumption that the Speaker plays the dominant role in dialogue, it nonetheless appears to be inescapable if the excerpts are considered within the framework of situation and audience modelling outlined in the previous chapter. In each case, the Speaker is *reacting* to a particular situation or audience, although – on the face of it, at least – the impact of the audience on the Speaker is not direct, but mediated through the Speaker's mental processes. Thus, it is not the nine-year-old that directly prompts my abortive attempt at conversation, but my internal

representation of him. In the same way, the unrepentant smoker of Example 2 does not directly prompt Speaker Two's mix of self-justification and authority; instead, this has its source in Speaker Two's mental model of Speaker One's habits. Both are the product – accurate in the latter case, wrong in the former – of audience modelling, in Brennan & Metzing's definition as 'pragmatic and partner-specific knowledge… implemented by basic mechanisms of memory' (*op cit*: 192), with the process being fed in turn by situation modelling with greater or lesser degrees of accuracy.

Example 2 is a particularly useful illustration of audience/situation modelling in operation. Speaker Two's assumptions about Speaker 1, mistaken though they are, clearly fit Johnson-Laird's definition of a [mental] representation of 'states of affairs…, the way the world is, and the social and psychological actions of daily life … [enabling] individuals to make inferences and predictions' (*op cit*: 397). Zwaan & Radvansky's components of a situation model can also be identified: the protagonist (Speaker 1) is obviously present in Speaker 2's representation, but so are details about Speaker 1's 'intentions… goals…emotions' (*op cit*: 36). Additionally, the updating of this model resulting from Speaker 1's reneging on an earlier promise illustrates, in a forcibly explicit way, the re-establishment of common ground between the couple over the issue: the re-alignment of their mental states, even when they are in disagreement. Speaker Two's nimble re-adjustment to the changed situation ('Cost you a buck') demonstrates the efficiency and power of the re-alignment process.

Re-alignment can also be seen taking place in Example 3, although in a more subtle fashion. Here, too, situation models are playing their part in the interchange (limited though the interlocutors' knowledge of each other is), and these are updated over both time and, interestingly, space. As this is a phatic exchange, the key element of the situation is the attitude of the speakers to each other – is this person, at even a superficial level, friendly/trustworthy? – and, with every utterance, the interaction confirms and re-confirms such friendliness. But it does so with reference to a continually-changing exterior reality: the interlocutors are moving along past the plants and talking as they go, with the deictic referents changing from plant to plant. Here is Clark's representational common ground in action, constructed from scratch

on the sum of what Clark calls the interlocutors' 'mutual knowledge, mutual beliefs, and mutual suppositions' (*op cit*: 3), and developing – accumulating – with every step through a real-world landscape.

Example 3 provides a further illustration of the theoretical approaches outlined in the previous chapter in that, like phatic utterances in general, it is largely strung together from the whole-message 'chunks' to which Levelt refers in his description of the outcome of long-term adult experience with speaking. *So pretty*, *really lovely, aren't they/don't they?…*: these are stock fragments from the speech repertoire of many adult English speakers, retrievable effortlessly from long-term memory rather than being 'invented time and time again through conscious processing'(*op cit*: 21). However, the most striking illustration of how Levelt's production model is assumed to operate occurs, not in the successfully-conducted interchanges in Examples 2 and 3, but in the abortive Example 4. The model component at work here seems to be Levelt's monitoring unit, sited inside the Conceptualizer and receiving input from outside this – from both the overt (audible) and internal speech that a Speaker has formulated – and from within it. This second aspect of the monitor's functioning is the one that checks, not whether the utterance being prepared is well-formed or comprehensible, but whether one should produce it at all. Is it appropriate? Is it polite? Will it have the intended effect 'in view of the present state of the discourse knowledge shared with the interlocutor(s)?' *(ibid*: 14) In Example 4, the appropriateness and effectiveness of any second utterance AJP may make to her audience are quite blatantly lacking. It would seem to be Levelt's intra-Conceptualizer monitor that has brought her to this conclusion, and done so extremely quickly.

At this point, however, a difficulty arises in the application of Levelt's theory. As explained earlier (Chapter 2.1.1), Levelt draws a distinction between the Conceptualizer with its monitoring sub-unit, and the formulatory and articulatory functions of his model. The latter, Levelt suggests, work automatically, 'without any awareness', and also at very high speed. The Conceptualizer, though, is presented as being under executive – discretionary – control: 'Speaking is usually an intentional activity…An intentional activity is, by definition, under central control…A speaker can decide on one course of verbal action rather than another on the basis of

practically any sort of information: his state of motivation, his obligations, his believing this rather than that, his previous speech acts or other actions, and so forth' (*op cit*: 20). And, Levelt continues, both message construction and monitoring are controlled processing, with self-corrections 'hardly ever' being made without some degree of awareness. 'The speaker can *attend* [Levelt's emphasis] to his own internal or overt speech' (*ibid*: 21).

This characterisation of AJP's initial loss for words as being the outcome of informed decision does not correspond to the subjectively-experienced reality itself, both the speed and the impact of which seemed dramatically out of proportion to the context. So is there more automaticity in the Speaker's self-monitoring process than Levelt suggests? Is this mechanism under dual control: that of the conscious mind and of an automatic system, described by Levelt himself as 'executed without intention or conscious awareness… usually quick, even reflex-like; the structure of the process [being] 'wired in', either genetically or by learning (or both)' (*ibid*: 20)?

As I have tried to show in both this chapter and the previous one, there is ample evidence for the impact made by an Addressee on a Speaker via audience modelling, carried out with varying degrees of conscious strategic purpose. However, it seems plausible that this memory-fuelled sensitivity to Addressees' needs, encapsulated in situation/audience models, may account for no more than a part, though an important one, of the Addressee's influence on the messages a Speaker produces. I would like to suggest that a *direct*, low-level, automatic system of interactions may also exist by which an Addressee can influence a Speaker on a moment-by-moment basis during an interchange and, in this way, ensure that what the Speaker produces is something readily processable at the Addressee end.

My goal in the research presented in this thesis is to investigate the existence and workings of such a mechanism, which can be conceived of to some extent as a cognitive system: a system whose function is to help ensure that there is a 'mesh' between Addressee needs and Speaker productions, and that the Speaker's output is adjusted from the start to the readiness of the Addressee to receive it. As hypothesised, this system draws on several aspects of investigations already undertaken into the capacities of the Speaker – e.g. Levelt's speech production model, the continuous observation of interlocutors revealed by 'common ground'

research, Pickering & Garrod's focus on low-level priming mechanisms, and (in their later work) the extension of this to include 'automatic resonance processes' (*op cit*, 2007: 444). My suggestion that there may be a dual-control process at work in self-monitoring also calls into play the dual-process accounts of cognition that have been put forward by researchers in thinking, reasoning and memory, which essentially 'posit two minds in one brain' (Evans, 2003: 454). I will develop the dual-process theme in Chapter 5.

At the same time, I also ask whether the Speaker-Addressee adjustment system that I propose is purely cognitive in nature. It seems plausible that, as Piwek speculates, the adjustments are significantly influenced by affect: by the 'central place' played by emotion and feelings in dialogue behaviour (*op cit*: 151).

Overall, I would like to suggest that, as hypothesized, such a system could go some way to accounting for Speaker success in achieving optimal relevance within the extraordinarily tight time restrictions inherent to naturalistic conversation. The remainder of this chapter describes some initial indications that such a system exists – mainly drawn from informal personal observation, particularly of dialogue onsets. Chapter 4 will look at experimental support for the hypothesis that these practical observations have led me to propose.

3.3 The power of the Addressee

The Speaker emotion expressed in Example 4b, which follows on my initial sensation of being dumbstruck by my audience's defection, is one indication that a low-level, automatic self-regulatory mechanism may exist alongside a consciously-operated one within the speech production system. And this reaction is far from unusual. The experience of being – quite literally – bereft of words in the face of Addressee non-attentiveness is a common one: the presenter whose voice falters into silence before indifferent listeners is experiencing it, as is the academic who, noticing a student sleeping through his lecture, feels his throat choke up at the sight. So is the student who no longer frequents a common room because of the 'deafening

silence' that rules in it.[2] And so is any interlocutor, actual or potential, who finds their audience is unwilling to maintain an acceptable degree of responsiveness during communication.[3]

As is readily observable, what counts as 'acceptable' varies from situation to situation and, indeed, from person to person. However, closed eyes that should be open are an obvious indicator of Addressee apathy, as is an averted gaze (or even a turned back), and there are others. Additionally, a Speaker whose output is greeted by an Addressee's blank, statue-like stare (usually followed by gaze aversion) can also find his interactive attempts negatively affected and, indeed, does so at a very early stage of human development. As Striano (2004: 468) points out, 'One of the most robust phenomena suggesting that young human infants expect people to behave in certain ways is the negative response caused by the sudden still face adopted by a social partner in a face-to-face interaction.' (This 'still-face effect' will be discussed in detail in Chapter 4.)

Nor do such negative reactions appear to be confined to the visual modality: a response in an abrupt or bored tone of voice will have similar effects to the ones just described. Suppression by an Addressee of back-channelling utterances on the telephone can have a noticeably disturbing effect on the Speaker, while lack of heard Addressee responsiveness can have an impact on a Speaker even in situations where both are in each other's presence. The health and safety sector provides an illustration here: commenting on evidence that hands-free mobile phones compromise driver safety, a spokesman for the Royal Society for the Prevention of Accidents is reported as saying that 'talking to someone in the vehicle is less distracting because that person can sense hazardous situations and modulate the

---

[2] Sources are personal communications from the three Speakers involved.

[3] A 'potential' interlocutor – Speaker or Hearer – is one who has not yet become engaged in interpersonal interaction, but who, given the physical constraints of the situation, could become so. Thus, the person standing next to me on the Underground is a potential Hearer although, in social terms, it would be unadvisable to try turning him into an actual one. Schober & Clark (1989) follow earlier practice by dividing the Hearer's role into two further categories, Addressees and Overhearers: 'Addressees are participants in the conversation at the moment, and overhearers are not' (*op cit*: 212). By definition, however, this classification refers to interactions that have actually started, not to ones that may or may not take place, and which I include in my discussion here (cf Example 4). For this reason, I would like to broaden the use of the term 'Addressee' in an attempt to cover both active and potential interlocutors. Hopefully, this will be a helpful generalisation, given the important role the not-yet (and possibly never-to-be) Hearer plays in my thesis. If my hypothesis is correct, a potential interlocutor has as much power as an actual one to impair Speaker fluency, just through the clear intimation of non-attentiveness.

conversation [i.e. stop talking, thereby halting the driver's own conversation]' (Webster 2007: 38). However, the research area that I cover here is limited to discussion of interactions that take place – either literally or virtually – face-to-face.

Overall, whatever the modality used and whatever the age of the interlocutors, the power exerted by Addressee non-attentiveness can have an impact that is both physical and immediate. Between one sentence and the next, an adult Speaker thus influenced feels the throat close and fluency deteriorate, and this deterioration is not confined to the operational processes of speech production. Conceptualization is also affected: the Speaker's grasp of what to say weakens, ideas do not come, and the drive to say anything at all loses strength. Speech can, of course, be maintained, but the process has now lost its normal fluent quality and become, as a frustrated Speaker will confess afterwards, 'hard work'.

This raises an interesting question. If reluctant Addressees can decisively reduce the efficiency of an interlocutor's speech production system, with damage being done right up to and including the message level, do willing Addressees have the power to do the reverse? Can they, by virtue of their evident attentiveness, actually *enhance* production efficiency, eliciting fluent speech from an interactant as opposed to damaging it? It is at this point in my argument that I return to considering the forms and functions of phatic communion.

3.4 Small talk: 'speech on autopilot'

Universal though it is, the phenomenon of small talk has received, at best, sporadic treatment in the linguistics literature. For both practical and theoretical reasons, it seems to lie outside the mainstream of academic interest: it is too spontaneous, too embedded in social context to be experimentally investigated, and too lacking in explicitly communicated propositional content to be of enduring interest to theorists. However, technological developments are now close to resolving some of the practical difficulties involved in experimental work (see Chapter 4), while, as I argued earlier, phatic communion already has qualities that make it a valuable proving-ground for dialogic investigation. But the most important of these qualities

lies at a deeper level than the standardisation of form and content mentioned earlier, and – crucially – involves motivational and affective factors as well as cognitive ones. It lies in the nature of the phatic communicator's goal, which involves establishing a positive connection between oneself and another and which, I would argue, demonstrates the elicitatory power of the Addressee at its most recognisable.

This goal of making interpersonal connections has been seen as the defining characteristic of phatic communion from its earliest appearance in the literature, in the famous description by anthropologist Bronislaw Malinowski. Eighty-eight years after it was first published, his analysis of the function and nature of the 'language used in free, aimless, social intercourse' (Malinowski 1923: 476) remains highly relevant:

> In discussing the function of Speech in mere sociabilities, we come to one of the bedrock aspects of man's nature in society. There is in all human beings the well-known tendency to congregate, to be together, to enjoy each other's company… Now speech is the intimate correlate of this tendency, for, to a natural man, another man's silence is not a reassuring factor, but, on the contrary, something alarming and dangerous…The modern English expression, "Nice day to-day" or the Melanesian phrase, "Whence comest thou?" are needed to get over the strange and unpleasant tension which men *feel* [my emphasis] when facing each other in silence.
>
> After the first formula, there comes a flow of language, purposeless expressions of preference or aversion, accounts of irrelevant happenings, comments on what is perfectly obvious. Such gossip, as found in Primitive Societies, differs only a little from our own. Always the same emphasis of affirmation and consent, mixed perhaps with an incidental disagreement which creates the bonds of antipathy… There can be no doubt that we have here a new type of linguistic use – *phatic communion* I am tempted to call it… – a type of speech in which ties of union are created by a mere exchange of words.'
>
> (*ibid*: 477-8)

Perhaps surprisingly, the topic receives only scant explicit mention in Brown and Levinson's seminal work on politeness (1987 [1978]: cf 109, 117-8). However, a set

of politeness techniques that they identify for claiming 'common ground[4]' with interlocutors covers some utterance goals that overlap with Malinowski's observations. These goals include seeking agreement with Addressees, avoiding disagreement with them, presupposing/raising/asserting common ground with them, showing exaggerated interest, approval, or sympathy, using in-group identity markers, and making jokes (*ibid*: 102). All of these are readily found in everyday phatic usage, along with the phatic mode's heavy reliance on context. Some examples are given below:

Example 5

Context: 'Several people have been waiting at a bus stop in North London for about twenty minutes. One of them walks some way up the road to see if a bus is coming. He then returns to the others and says (facing another person who is also waiting):

a) A: No sign of a bus. I suppose they'll all come together.
b) B: Oh yes. They travel in convoys.'

(Žegarac & Clark, 1999b:567)

Example 6

Context: Midwinter; AJP buys a paper

a) Newsvendor *(giving back change):* There you are, darling.
b) AJP: Thanks. Cold, isn't it?
c) Newsvendor: Cold ? This is a heatwave – I'm dreading winter.

(Pollard, 2005a: 3)

Example 7

Context: In the street, AJP suddenly hears someone close behind her, sighing:

a) Passer-by *(loudly)*: Oohhh dear!
b) AJP*(startled, looks round and catches his eye)*

[4] Note that the meaning they give here to the term is more general than Clark's: it merely indicates areas of common experience or interest that will allow a Speaker to convey fellow-feeling to an Addressee, and thus respect for the Addressee's needs, values, and self-image.

c) Passer-by: It's been a long day so far.

d) AJP I do know what you mean.

(Both smile; the passer-by moves on)

(*ibid:* 15)

Example 8

Context: In a train, AJP climbs on the seat to get her bag from the luggage rack. Another passenger stands up to help her:

a) Passenger: You should have worn your heels.

b) AJP *(who seldom wears high heels but doesn't want to kill the joke):*
Oh yes.

(*ibid:* 19)

Example 9

Context: AJP waits on the pavement to cross a busy road. She is carrying a loudly mewing cat in a basket. Another pedestrian glances at the cat, then away. He glances again, then looks up to meet AJP's gaze:

a) Pedestrian: They don't like it, do they?

b)AJP: No – he's very patient, but he's not very keen.

(collected by Pollard)

Here are Malinowski's comments 'on what is perfectly obvious' (Examples 6 and 9). Here, too, as described by Brown & Levinson, is avoidance of disagreement (Example 8), joking (Examples 5 and 6), and exaggerated sympathy (Example 7). And they give plentiful illustrations of whole-message utterances and linguistic routines that can skip conscious selection (cf Levelt) or the full production process (cf Pickering & Garrod) and come to the utterer's lips ready-made; even the 'convoy' reference in Example 5 is now so old that it has lost its freshness as a joke and become a verbal ritual – a routine. This is not speech production as hard work; it is, instead, speech on something close to autopilot, speech where utterances geared to

the Addressee's requirements are produced quickly, continuously, and with minimum apparent recourse to decision-making. How does this happen?

The answer, I would argue, overlaps with that to the question I raise at the end of the previous section: does the influence of an Addressee over a Speaker extend to eliciting fluent speech, as well as inhibiting it? The next section explores this connection and also considers an aspect of dialogic interaction that Pickering & Garrod tend to bypass**,** apart from some isolated references (eg 'People use beliefs about their interlocutors to start the process of alignment', Garrod & Pickering, 2007: 445). This is the nature of the procedures employed when a dialogic episode is opened. What happens in your Conceptualizer – your decision-maker for speech intentions – to cause you and your interlocutor to first start talking? And is it the potential Addressee who, here too, at a low, non-verbal level, controls the exchange?

3.5: Affective discomfort and the need to speak

After Malinowski himself, the leading theorist on phatic communion has been John Laver, who in 1974-5 suggested that the 'fundamental social function of the multistranded communicative behaviour that accompanies and includes phatic communion is the detailed management of interpersonal relationships during the psychologically crucial margins of interaction' (Laver, 1975: 217), and outlined a seven-stage process – culminating in speech – by which such interactions are opened and (in reverse order) closed. The first/last stage in these marginal interaction phases where 'psychological comfort is most at risk' (*ibid*: 236), is identified as establishing mutual eye contact.

The accuracy of his observation – which is anticipated in Malinowski's reference to potential interlocutors 'facing each other' – is confirmed by all the examples given above of successful phatic interchanges. In Examples 5 and 7, the references to eye-contact or 'facing each other' are explicit; in others (eg Examples 3 and 8), it is inherent in the situation, even if the verbal part of the interaction in Example 3 is slow to develop. Example 9 is particularly interesting, especially if it is considered in association with Example 10 (given below). In the former, AJP's interlocutor

initially feels no compulsion to comment on the cat's howls, impossible though these are to ignore. It is only when mutual awareness is established – when he sees that I am looking at him – that a typical phatic exchange is embarked upon. (Example 7 gives a less marked illustration of the same phenomenon.) However, no such access of mutual awareness takes place in the situation described in Example 10 where, in spite of a promising context, communication simply fails to get off the ground:

Example 10

Context: The Welsh rugby team has just scored a dramatic victory against England. The following day, AJP (a Wales supporter) sees a man wearing a WRU jersey, waiting beside her to get off an intercity train:

a) AJP: Grand game, wasn't it?

b) WRU supporter: *(looks round, nods slightly, looks away again)*

(Pollard, *op cit*: 16)

Not only is the opening stage of Laver's phatic interaction process here ignored but the Addressee, when accosted, refuses even to sustain eye-contact – thus signalling that interaction is not welcome.

From its beginnings, the phatic communion literature has emphasized the joint social and psychological benefits of the phenomenon. The Couplands & Robinson summarise them: 'A mode of discourse through which we are sanctioned to behave formulaically and yet meaningfully, veridically and yet not absolutely so, will have unique bridging potential – relationally and interactionally' (*op cit*: 226). And, also from the start, the literature has unhesitatingly pointed to the negative consequences of refusing such interaction: to the 'strange and unpleasant tension' involved in a silent face-to-face encounter; to the anxiety experienced at the openings and endings of encounters; to the threats to (metaphorical) face. 'Most peoples', as Abercrombie comments (*ibid*: 3006) 'have a feeling that a silent man is a dangerous man. Even if there is nothing to say, one must talk'.

It is in the compulsion to which Abercrombie refers that some answers may perhaps be found to the main question posed above: does the influence of an Addressee over a Speaker extend to eliciting fluent speech, as well as inhibiting it? I would suggest

that it is precisely this influence that triggers this compulsion: that it is the Addressee-to-be who, by his or her silent but attentive presence, elicits the stream of noncontroversial, non-informative, semi-ritualised speech that the initial Speaker in a phatic exchange produces so readily and easily. Further, it is not an interactant's need to produce speech that creates the tension that precedes the interchange; as Examples 4 and 10 demonstrate, a unilaterally-taken decision to speak does not, on its own, guarantee successful communication. Instead, it may be the need – felt by both interactants – to *be* an Addressee that finally sets the interaction going.

At one level, the outcome of this shared impulse is simply a moment's social pleasure: a tiny unexpected bonus in the course of a humdrum day. I would argue, however, that its real value to the participants lies deeper, at the level of affective need. To be in eye contact with someone and yet find no reflection there of one's reality is indeed an alarming experience. It threatens, not merely one's self-esteem, or 'face', but the very sense of one's existence. And it is this existential threat that creates the need, not so much to take action, as to produce a *reaction* – to receive proof, via some attempt at interpersonal connectivity from the other, that you really have being.

Sensitivity to such discomfort obviously varies, both between cultures and between individuals, and so does the effect produced by the temporary social context on how the potential interlocutors deal with the problem. In situations where social connectivity is taboo, such as the London Underground, the solution is to avoid eye contact altogether (failure to do so carries a notorious risk of adverse reaction). But, where there is a social presumption that connectivity is appropriate or even obligatory – as between colleagues who see each other daily – a stare that loses responsiveness and becomes blank is scarcely an option, putting as it does a sharp brake on the other's wish to connect. (The verb 'to blank', as in 'Then she blanked both of us', is a recent and expressive addition to informal vocabulary in UK English.) Instead, a range of solutions is on offer, following along the lines of Laver's seven-part organisational structure of a phatic encounter's opening/closing stages: the exchange of non-verbal gestures of acknowledgement, assuming facial

expressions of attentiveness, moving close enough for audible talk and so on[5]. And, if circumstances permit, the rule that 'one must talk' will sooner or later come into play. One of the interactants – perhaps the one with the greater affective need, the one who most urgently requires the reassurance of the Other's acknowledgement of his or her presence – will give way, and prompt this acknowledgement… by starting to talk.

The pressure exerted by face-to-face silence is not, of course, confined to the phatic environment; Levelt himself, in his account of a highly information-charged interview between two academics and a student, acknowledges that a long silence on the student's part 'made it important to do something' (*op cit*: 5). Indeed, amongst the six basic aspects of language listed by Jakobson (1960/1995:73), it seems plausible that only the 'expressive' function, which 'aims a direct expression of the speaker's attitude towards what he is speaking about', exempts a Speaker from the need to negotiate the affectively threatening margins of interaction. As Jakobson himself points out, it would be hard to find verbal messages that fulfil only one function. It follows from this that an element of phatic concern – the need to ease the anxiety attendant on establishing recognition by, and maintaining contact with, the interlocutor – appears to be present in the great majority of dialogues. I would speculate that it is ultimately this anxiety, which is the product of a potential or actual Addressee's influence, that elicits fluent, easily put-together speech from a Speaker.

A further question that arises from the above discussion is this: can it be that much of the effort expended by a Speaker on speech production is in a relationship of *inverse ratio* to the attentiveness of an Addressee? The commonly-experienced behaviour typified in Examples 3 and 4 – taken from each end of the effort continuum – seems to give intuitive support to this idea, with indications of the mechanism involved appearing in Examples 9 and 10.[6] In between the extremes lies a range of Addressee/Speaker reciprocal relationships in which the impact of the one on the

[5] Ferguson, in his account described above, got no further through this repertoire than 'a friendly smile'. As we have seen, the imbalance between even this positive phatic signal and that of his interlocutor was enough to cause pronounced reactions.

[6] It is interesting here to note that this impact is routinely created artificially in the context of film production: in a studio, it is common practice for an actor to be positioned behind the camera, to act as a responsive 'Addressee' to his colleague on the set. As a television professional points out, 'For people who act, you have got to have an audience' (Calder, 2006).

other may be more or less marked but, at the extremes themselves, some very powerful force is certainly at work.

This is a different type of audience modelling from that described in the previous chapter. It is powered, not by the Speaker's strategically-deployed memory of situational factors, but directly, by the audience itself. At a fundamental level, where a Speaker can be compelled into speech or completely eclipsed, the Addressee is in control.

Such a conclusion, in turn, arguably demonstrates an interaction between perception, the affect system, and speech production (notably at message generation level). The connection between perception and the affect system is scarcely in doubt: as Cowie (2009: 3522) comments, 'human emotional engagement depends on perceiving not only what the other party's emotional state is but also that the other party is engaging emotionally. When agents are unable to give cues that signify engagement, emotion can be identified but emotional rapport cannot be built.' It is the further connection between these systems and verbal interaction that, as Piwek suggests, requires investigation. This chapter discusses informally-gathered gathered indications that such a connection may be present; the following one considers experimental evidence for its existence.

3.6 Summary and conclusions

This chapter presents the central hypothesis examined in this thesis: that a low-level, automatic system of interactions exists between a Speaker and an Addressee whereby an Addressee can influence the Speaker's output on a moment-by-moment basis, thus ensuring its continuous adjustment to the readiness and capacity of the Addressee to process it.

As initial support for this proposal, I provide naturally-occurring, informally-captured examples of phatic communion: a mode of communication that, although often ignored in research, is well-suited to this particular research topic for several reasons, among which are its apparent universality and its obligatorily dialogic nature. It is, however, the underlying intention behind phatic communion – the

establishment of interpersonal rapport via speech – that most clearly demonstrates the relevance of such a communication mode to my hypothesis.

There is a general assumption that the dominant partner in a dialogue is the Speaker; it is the Speaker, so this assumption runs, that initiates exchanges, dictates their progress and, dependent on his verbal skills or the lack of them, establishes rapport. However, the examples given here seem to show the reverse. In these, it is the Addressee who appears as the main player, either promoting interconnectivity or abruptly suppressing it. And, although the channel of overt communication between the two is a verbal one, the Addressee's power to elicit or extinguish speech from an interlocutor seems to derives from a process operating at a lower, non-verbal, level.

The hypothesis offered in this thesis is that it is Addressee attentiveness – evinced, especially, by eye-contact, though other methods are available – that has this elicitatory power over an interlocutor's speech intentions and abilities; conversely, it is Addressee indifference, often demonstrated by discontinued or occluded eye-contact, that sharply brings even a willing Speaker to a halt. In the case of the elicitatory function, I argue that this is driven, not by intentional decision-making on the part of the Speaker, but by low-level affective requirements: by an interactant's emotional need, shown particularly clearly in the phatic situation, to achieve existential acknowledgement from his interlocutor. It seems reasonable to suggest that the silencing effect similarly has its origins in affective need – with, in this case, speech production being negatively influenced by the shock of having one's existence denied, through an interlocutor's inattention.

Although the inquiries described in this chapter are both informal and limited in scope, they can already claim to add to the picture of the Speaker that emerged from the review presented in the previous chapter. For example, the speed with which Addressee indifference takes effect appears to extend Levelt's model of speech production beyond the limits that he himself suggested; as subjectively experienced, 'eclipse by Addressee' seems to take place extremely fast, and independently of any conscious decision on the Speaker's part,. Is it possible, then, that this part of Levelt's Conceptualizer – its internal Monitor, responsible for inhibiting the formation of inappropriate or ineffective messages, and theoretically under the

Speaker's conscious control – is capable at times of functioning automatically as well?

The capacity of a Speaker's conceptual decision-maker to thus act as a dual system – working both within his awareness and below it – gives added focus to the whole issue of conscious control vs automaticity raised at the end of Chapter 2, for it places it within the context of dual-process theory currently being investigated in several fields, including cognitive and social psychology. The potential addition of speech production to the range of this theory seems well worth examining and, in Chapter 5, I consider this in detail. However, my first priority is now to seek firmer empirical support for the suggestion that speech production can be dramatically influenced by its hearer's attentiveness, or the lack of it. Does the interested gaze of an Addressee have a reliable, positive impact on the fluency of a Speaker? Does an occluded gaze, or an abstracted one, have the opposite effect? And what about other non-verbal signals of attentiveness, such as nodding or smiling? Empirical evidence on these points is reviewed in the next chapter.

CHAPTER 4

The eyes of indifference: how the Addressee exerts control

Introduction

The previous chapter presented the hypothesis that, by indicating attentiveness or indifference, Addressees can control Speakers' output at a pre-conscious, non-verbal level. The initial mechanism by which face-to-face Addressees exert their influence also appears reasonably well-defined: as shown by the examples given, the evidence seems to point to the face itself – and the eyes in particular – as the indicator of attentiveness or the reverse.

To investigate both my main hypothesis and the mechanisms that may support it, this chapter considers data from three different research contexts – child development, sociolinguistics, and software engineering (virtual environment construction) – on the effects of gaze direction and eye-contact. I argue that this evidence identifies communicative/interactive behaviour that promotes or discourages interaction, including verbal interaction, and that it therefore offers experimental confirmation of the hypothesis proposed in the previous chapter: that the Addressee can control the Speaker in a low-level, non-verbal way, thus ensuring that Speaker productions meet Addressee needs on a moment-by-moment basis.

This evidence also supports claims of the importance of affect in dialogue behaviour, and adds to the Leveltian model of the Speaker by demonstrating a relationship that is automatic as well as voluntary between the knowledge store representing the outside world (the Addressee's behaviour included) and the intra-Conceptualizer monitor. Additionally, the chapter illustrates the development of experimental methods of capturing naturalistic dialogue, from the 1960s to the virtual reality applications of Boker, Cohn et al (2009), together with the considerable promise that

these advances offer for future research. At the same time, it outlines the development of interest in what dialogic interactants actually say, and the relationship between this and their non-verbal behaviour.

Principal sources are: infant development (Chapter 4.1), Murray & Trevarthen, Adamson & Frick, Delgado et al, and Striano; adult gaze interaction (Chapter 4.2), Kendon, Argyle & Dean, Kleinke, Bavelas et al; virtual reality environments (Chapter 4.3) Pertaub et al, Garau et al, and Boker, Cohn et al.

4.1 'A deep-rooted drive to communicate'

In the previous chapter, informally-captured examples were given of Addressees whose reluctance to engage with a Speaker had an immediately suppressive impact on the latter. While, until recently, little formal investigation appears to have been carried out into the effect of this 'blanking' on adult communicators, the same is not true of infants, among whom such treatment has been claimed to produce a reliably negative effect. Findings such as these support the argument that a connection exists between indications of attentiveness and the human drive to communicate. Moreover, they support the suggestion that this connection starts so early in development that the capacity for making it may be hard-wired into the human brain: that, in the words of Tricia Striano (2004; 478), there exists in the human infant 'a deep-rooted drive to communicate with others and not be ignored.' I would suggest, further, that this deep-rooted drive and its mechanisms lie at the core of the human adult's communicative behaviour, and that the reaction of the infant to being blanked by its mother is a precursor to that of the adult whose verbal competence is eclipsed by a stony-faced Addressee. Evidence of this second phenomenon will be presented later in the chapter (see Chapter 4.2.5)

This section traces the development of the argument concerning infant responsiveness, and discusses some of the findings on which it is based. It shows how a non-responsive interactant can close down interaction even in the earliest stages of child development, and also how this can be discouraged by responses that are inappropriately timed. In addition, it outlines discussions that have taken place

on the possible roles played by affect and cognition at this foundation stage of communication.

4.1.1 The still-face effect: withdrawal and eclipse

Several of these issues are raised in one of the earliest and best-known demonstrations of the reaction to 'blanking' in very young children: Murray & Trevarthen's 1985 study of the emotional regulation of interactions between 6-12-week-old infants, undertaken to investigate connections between affective expression and the development of interpersonal communication. Building on ground-breaking earlier work by Tronick, Adamson et al (1975/1978) the investigation focuses on infant behaviour when confronted with precisely the phenomenon discussed in Chapter 3: an expected interactant – here, the infant's mother – who refuses to interact[1]. Additionally, the infant subjects are given two possible 'reasons' for such cessation. In one condition, interaction between mother and baby is interrupted when the experimenter enters the room and briefly asks the mother a few questions. In the other, no such intervention from outside occurs; instead, the mother simply looks at her child while keeping her face expressionless and unresponsive (the 'blank-face' or 'still-face' condition). Both conditions alternate with periods of 'normal' face-to-face interaction, with one of these starting the experimental session. All the experimental interaction periods are brief: 30 seconds in the case of the 'interruption' break, 45 seconds in the case of 'still-face'. In presenting these results, I will try to distinguish clearly between observed behaviour, on the one hand, and functional interpretations in terms of attributed mental or affective states, on the other.

Infant reactions studied include gaze direction; wide open mouth and tongue protrusion (which Murray & Trevarthen interpret as 'communicative effort'); smiling and raised eyebrows (interpreted as evidence of positive affect); and frowning and drooping head (interpreted as negative affect, as are behaviours such as yawning,

[1] According to Csibra (2010: 145), the power to distinguish between eye contact and the lack of it is evident from the earliest moment possible: humans, he says, are sensitive to the presence of eyes from birth. He goes on to cite evidence (Farroni et al, 2002) that, even at the age of three days, infants shown photographs of faces looking directly at them or looking away prefer to look at the face that appears to make eye contact with them: a 'very robust effect, unusually strong among studies with neonates.'

fingering the face, grimacing, and chewing the lower lip, all categorised as 'displacement activities'). As the results of the procedure show, even the 'normal' periods in the experimental sessions provide valuable data on the interactive behaviour of very young infants. During these periods, for example, they look at their mother's face most (90%) of the time, they display raised or relaxed eyebrows, smile frequently, seldom frown, and make few 'displacement actions' apart from occasionally fingering their clothes: all actions that Murray & Trevarthen take as suggesting the presence of a deep-seated drive to interact, and to derive enjoyment from the interaction. Data is also provided on length of gaze periods: most of the infants' looks at their mothers were 'long', lasting over 5 seconds, while a smaller number were short glances (under 2.5 seconds) and a few more fell between the two.

However, the most striking results are produced by the infants' reactions to their mother's cessation of interaction – and, in particular, their reaction to their mother's suddenly still face. Murray & Trevarthen's account of the still-face effect obtained goes into considerable detail:

> Within a few seconds the infant showed distress in peculiar, sneering grimaces of the mouth, increased handling of the clothes, touching the face, sucking the fingers, and frowning. Efforts at communication, defined by a wide open mouth and tonguing postures were sustained, and even intensified at first, the whole sequence being toned with negative affect and accompanied by active gesturing of the limbs. This initial reaction gave the impression of protesting or straining to reinstate interaction with the mother. This phase was followed by withdrawal: the infant averted his gaze downward from the mother's face, looking to her overall only 34% of the time. The relaxed expression faded and the infant almost never smiled.'
>
> (*op cit*: 186)

Although the impressionistic basis of Murray & Trevarthen's interpretations has to be acknowledged, it is perhaps not fanciful to see early evidence in the first phase of this behavioural pattern of Malinowski's 'strange and unpleasant tension' in operation, evoked – as in adult life – by the presence of an aware but non-responsive Other. Here, it is the 'blanked'infant who tries to resolve the situation by increasing the expressiveness of its behaviour, in the possible hope of eliciting a response: an effect that has been replicated recently amongst adults (see below, Chapter 4.3.3).

When this fails, eclipse follows, with the baby's normal gaze at its mother much reduced.

The interruption condition, in which the mother looks away from her infant to interact with the experimenter, does not produce such an extreme reaction. While mouthing and tonguing movements decrease, together with the incidence of smiling, the babies remain relaxed overall, and even appear to show a 'quiet interest' (*ibid*: 192) in the experimenter's entrance.

The *length* of the babies' gaze at their mothers also changes during the two experimental conditions: dramatically, in the case of the still-face one. Here, the number of long continuous looks is halved, the number of short glances is doubled, and the intermediate looks slightly increase. (The interruption condition shows similar but less marked changes.) These results, Murray & Trevarthen suggest, parallel the descriptions of presumed affective change produced by broken mother-infant contact, and show that it is not simply the overall amount of looking that changes across conditions. Gaze patterns become organised differently, with long, sustained, continuous looks taking place during normal face-to-face communication in contrast to the many short glances evoked during the blank-face episode. They conclude: 'In all conditions, gazing to the mother's face is accompanied by communicative behavior (tonguing and wide open shaping of the mouth) and signs of positive affect. When the infant is looking away from the mother no such active efforts are shown. In the Blank-face condition the infant showed a tendency to display negative affect when looking to the mother' (*ibid*: 188).

This analysis, Murray & Trevarthen argue, shows 'coherently organized and complex' expressions of infant affect and attention that are systematically geared to changes in maternal behaviour. In a supplementary study with four children of the same age, they investigate the impact of disturbing, not the nature of the mother's interaction with her baby, but its timing. Here, mother and infant respond to each other's gaze and vocalisations over a double video system while in separate rooms. During the control episode, the pair respond to each other live, in real time; in the experimental one, however, the interaction witnessed by the child is the video of the mother from the real-time episode, replayed – in which her behaviour is obviously out of synchronisation with the ongoing interaction of the baby.

The results show a similar pattern to those of the still-face experiment. Infant reactions during the live presentations largely parallel the 'normal' interaction in the other study (although there is much less smiling than in the face-to-face condition – itself an interesting result). However, the replay condition sees a 'considerable change', in which the baby turns away from the image of the mother, frowns, grimaces, and yawns. Also as in the still-face experiment, short glances take over from the 'normal' pattern of long (5-second) looks. But one difference between the two perturbed conditions is notable: infants in the replay episode – unlike those in the still-face one – do not initially appear to try to re-engage the mother in satisfactory interaction. Instead, the feeling that Murray & Trevarthen claim is conveyed by the infant's behaviour here is one of 'detached, unhappy puzzlement or confusion (*ibid*: 191)'. The infant occasionally looks at his mother in an expressionless, hesitant way, possibly smiling or making tonguing movements upon some 'chance coincidental interaction' with the mother's replayed behaviour. But then, when his own acts and those of his mother again diverge, he turns away with frowns, grimaces, or yawns.

Overall, Murray & Trevarthen suggest, their results indicate that infants of six to 12 weeks are able to detect aspects of the mother's behaviour (e.g. her gaze direction and facial expression); can respond to these in terms of 'coordinated structures of interpersonal and emotional value'; and can regulate their own expressions in appropriate fashions that can be interpreted by the mother. Their efficient handling of these interactions as early as two months old in turn suggests they already possess a mechanism for categorising in emotional and expressive terms the human stimuli to which they are exposed: 'To this extent, the forms and communicative values of human emotions are innately formulated even though their uses in regulation of contacts and relations undergo considerable development in infancy and childhood (*ibid*: 194).'

As I will show later in this chapter, evidence of such possible underpinning to social – and more specifically linguistic – interaction has also emerged in adult communication: in particular, from adult variants on the still-face procedure (see Chapter 4.3.3, below). And, as demonstrated by its widespread use among infants

over the last 30 years, the still-face procedure's overall value has been seen as considerable, both for the simplicity of its use, and for the wealth of behavioural detail that it elicits. However, as Adamson & Frick point out in their 2003 history and review of the paradigm, the results that it obtains can be questioned on several counts, such as their narrow focus - they supply data only on the still-face episode, not the subsequent recovery process – and the precise 'functional interpretations' proposed for the infants' actions.

Among examples of more recent interpretative schemas, they quote the Infant Regulatory Scoring System (IRRS) used by Weinberg & Tronick (1994: 1506), who analyse the co-occurrence of specific facial expressions with a list of specific behaviours (e.g. inhibition/freezing: 'the infant is glassy-eyed') to examine the existence and significance of coherent patterns of infant expressivity[2].
Murray & Trevarthen's 1985 studies represent an early attempt to fill the interpretatory gap, albeit an attempt apparently based less on formal data and more on impressions and previous (unattributed) descriptions of normal and perturbed infant behaviour. In spite of this, however, their findings have a value that goes beyond that of the detailed behavioural observations that they make. As summarised by Adamson & Frick (2003: 461), the main outcome of the still-face procedure is the demonstration of the dramatic behavioural change, or still-face 'signature', that it elicits: the 'increased gaze aversion and decreased smiling that [contrasts] markedly with an infant's normally attentive, gleeful engagement during face-to-face interactions'. This signature, which is clearly present in the data captured by Murray and a colleague, indicates a strong disruption in the interactive process caused – even at this early stage of development - by the non-responsiveness of a partner, and provides at least some evidence for possible underlying mental or affective states.

[2] Although the IRRS postdates Murray & Trevarthen's experiment, this was not the case with the coding system used by Weinberg & Tronick for infant facial expressions - Izard et al's AFFEX ('System for identifying affect expressions by holistic judgments'; 1983) - or Izard's MAX ('Maximally discriminative facial movement coding system'; 1979), both of which link specific facial expressions to discrete infant emotions. However, the validity of both schemes and of the theory underpinning them - Differential Emotions Theory – has since been challenged (cf Oster et al 1992, Matias & Cohn 1993, Camras et al 1993); according to Matias & Cohn (*ibid*: 529), for example, investigators should be 'extremely cautious' in drawing inferences about emotions on the basis of MAX-specified discrete negative facial expressions.

4.1.2 Still-face distress: a cognitive response or an affective one?

Murray & Trevarthen's 1985 experiments can also be criticised on the grounds of scale: only eight infants took part in the still-face study, and only four in the replay one. Nonetheless, the overall effects they noted have proved extremely robust, and indeed have been observed in babies across a wider age-range than those described above. According to Delgado et al (2002: 311), the still-face responses of infants as young as one month and as old as seven months 'consistently include decreased smiling, increased grimacing and distress, increased crying, and decreased gazing at the parent.' Indeed, still-face distress has also been noted in infants who are even younger: Nagy (2008: 1779) reports on a study of 90 neonates, three to 96 hours after birth, who decrease eye contact and otherwise react negatively if their interaction partner becomes unresponsive; they also continue to show signs of disturbance even after the interactant becomes responsive again. 'These results indicate that even newborn infants sensitively monitor the behavior of others and react as if they had innate expectations regarding rules of interpersonal interaction.'

Well-attested though the still-face phenomenon is, conclusions on the psychological mechanisms contributing to it have been more contentious. More than one theory is in the frame, with contenders including a nascent theory of mind: Striano, in her 2004 review of the literature, notes that the still-face paradigm does not provide much insight into infants' understanding of the underlying reason why interpersonal contact was broken and comments that the developmental origin of such understanding is of interest because one of the most significant and unique aspects of human cognition is 'the inclination to probe and consider others' minds (i.e. theory of mind…)' (Striano, 2004: 468). Amongst the questions raised over the development in early years of an understanding of internal mental states, she points in particular to the continuing debate over whether an understanding of others' motives or intentionality may already be present at birth or whether it emerges later in development. The development of infant social initiative-taking in interaction[3] offers one method of revealing how and when such understanding develops:

[3] An example that Striano cites is Cohn & Tronick's 1987 study of infant-mother interaction at the ages of 3, 6, and 9 months. Their results indicate a significant increase in the infants' social initiative-taking after the age of 6 months: up to that point, positive facial expressions such as smiling are

> It is possible that before the end of the first year, infants manifest an awareness of intersubjective intentions, or the attentional states or affects that provide cues to these intentions. Accordingly, if the human infant expresses a nascent understanding of intentions, such understanding should be observable in the early months once they more systematically engage and reciprocate in the context to dyadic interactions.
>
> (*ibid*: 469)

The hypothesis suggested here – a *cognitive* element of intention-recognition contributing to the initial stages of human interactivity – has considerable implications, not least for my own inquiry into the extent and nature of control exerted by an Addressee over a Speaker. On the face of it, it is one that Murray & Trevarthen's study appears to confirm: the behaviour of infants here alters , not when their mothers' interaction with them is halted for a visible, obvious reason (i.e. the entrance of the experimenter), but when such a reason is absent, with each change in their environment (including the Replay condition) giving rise to 'a distinctive motivational state coordinating and integrating discrete acts in profiles that convey a personal significance' (*op cit*: 191). Striano herself, however, expresses caution at this point. It is possible, she says, that the infants were using their mother's eyes as a signal for communication in the still-face condition and 'were simply distressed when contact could not be re-established. Such ability would imply the use of the eyes as a cue to communication but would not imply an appreciation of the underlying mental state or the underlying intent of the person' (*op cit:* 469). The suggestion of a link between gaze and interactivity is, of course, important here.

The picture is complicated by the presence of yet another hypothesis that has been proposed to explain attentive and affective behaviours in infant-adult interaction. According to this 'discrepancy model', which Murray & Trevarthen point out is *not* confirmed by their findings, negative reactions to the still-face experience demonstrate the development in the infant, not of a rudimentary theory of mind, but of an understanding of social context: of the relative likelihood of certain social

---

adopted by the infant *after* the mother has displayed them. By 9 months, however, infants become significantly more likely to smile before the mothers do.

situations. When an infant's expectations of a situation are violated, distress seems to result. In addition, as Delgado et al note, it has been observed that minor changes in the 'still-face' parent's behaviour can modify the infant's response: if the face is still but happy, or if the parent touches the child, the negative response is reduced. So are young infants capable of distinguishing such nuances in their social context, and responding accordingly – with less apparent disturbance where a still-face episode comes closest to the interaction conditions they are used to?

Delgado et al themselves address the problem by applying to the infant situation the principle that eye-contact indicates an adult's availability for interaction. 'Adult gaze direction', they point out, 'is an important social indicator that has not been previously studied in the context of the non-interactive still-face condition (*ibid*: 312)'. In a study of 43 6-month-olds, they compare the reactions of babies whose mothers adopt a standard still-face expression with those whose mothers adopt this expression in a slightly different context: to look above the child at a picture. The hypothesis here is that most infants do not experience the gaze-at still-face condition, which therefore represents a greater departure from their expectations than the gaze-above condition. If the discrepancy model is valid, the infants should react more negatively to the gaze-at condition than to the gaze-above one.

However, the results prove negative, with no difference appearing between the two conditions. The gaze-above group responded in the same way as those in the gaze-at one, confirming the general still-face effect by showing decreased gaze and smiling, and increased negative affect and crying. Among the reasons for this lack of sensitivity to gaze differentiation, Delgado et al acknowledge a possible problem in distinguishing between vertical and horizontal gaze shifts. However, they continue, 'due to the unfamiliar and potentially unpleasant nature of the still-face condition, it is [also] possible that during the still-face episode infants are responding affectively to the lack of interaction by the parent rather than cognitively to the distinct social contexts created by changes in eye orientation' (*ibid*: 316). Importantly, too, they suggest shifting the theoretical emphasis from the cognitive to the affective dimension, and add that further research is needed to provide a fuller understanding of the still-face effect and of infants' understanding of social contexts. The next section describes some of this research, with its outcomes.

4.1.3 The urge to share attention with others

Delgado et al's results appear to invalidate the discrepancy model of infant-adult interaction; how far do they seem to challenge the hypothesis that infants are aware of intentional differences in their adult interactants? Striano warns that, while the Delgado findings seem to indicate that 6-month-olds lack such interpersonal sensitivity, further test controls are still required: the age-range of participants needs to be extended, and the reason why infant-adult contact is broken should be more evident. If, however, the universal still-face response persists even under these conditions, it would suggest a 'possible primacy of interpersonal communication and social expectations in early human ontogeny' (*op cit*: 470), with social expectations or the drive for communication playing a more decisive role in explaining the infants' reactions than the motive or underlying mental state behind an interactant's behaviour.

Striano's 2004 investigation of this hypothesis extends the work of Murray & Trevarthen and Delgado et al in terms of both scale – 152 [120 + 32] babies were involved – and infant maturity: reactions are captured from 3-month-old, 6-month-old and 9-month-old children, the ages being selected on the grounds of their importance as key developmental transition points. As the intention here is to establish whether the infants show a different still-face response as a function of the experimenter's intention, the reason for the interruption – the cognitive dimension – is made more salient. Thus, in the first stage of her two-part study, the behaviour is recorded for 120 infants of all three ages when confronted with either:

a) the experimenter's still-face directed at the child, followed by one directed at a blank wall, the two separated by periods of normal interaction using infant-directed speech, or
b) the experimenter's still-face directed at the child, but with another person (the mother) standing beside the child and reading aloud, followed by a still-face directed at the mother – again with periods of normal interaction separating the two episodes.

According to the cognitive model of infant-interlocutor interaction, this clearly-differentiated pattern of interruption should produce differing still-face responses when the reason for the interruption is not evident, e.g. in the two look-away conditions. In particular, this differentiation is expected to be especially pronounced in the 'look-away-at-wall' condition, compared to the 'look-away-at-person' one, since 'wall' provides a less salient reason to look away than 'person' does. Striano finds, however, that such differences do not appear: 'Results showed that infants across all ages manifested a still-face effect… In relation to differences between infants' reaction as a function of the reason contact was broken, especially in the look-away conditions (look away at person or wall), there was scant evidence of any differentiation' (*ibid*: 474)[4]. In the second stage of Striano's study, undertaken to check possible differences between infants' reactions to strangers and to their mothers, 32 babies from the three age-ranges are involved. Here, only a still-face-away effect is studied: the mothers look away from their infants, either in response to a sound or for no apparent reason. Again, the infants show a reliable still-face effect, gazing more at their mothers during normal interaction than during look-away, and there are no significant differences between the still-face sound and no-sound episodes.

Overall, the results show that, by the age of three months, infants are already sensitive to others' attentional states, responding with a still-face effect when contact is interrupted for whatever reason. According to Striano, these findings point to what may be 'a unique and deep-rooted inclination in humans', the urge to share attention with others:

> What the current studies demonstrate is a deep-rooted drive to communicate with others and not be ignored…This inclination did not undergo much change over the first year and did not depend on the familiarity of the interactive partner. It suggests that human infants' drive to communicate and to have others attend to them starts well before the end of the first year with the emergence of systematic joint attention…The key is to assess whether such

[4] At the same time, interesting variations were captured between the behaviour of different age groups in different conditions, such as the 3-month-olds both gazing and smiling reliably more than the older infants in the 'normal' conditions, and also making more positive vocalisations than their seniors.

> inclinations reflect a human adaptation that affords unique aspects of culture such as language and theory of mind.[5]
>
> (*ibid*: 478)

These results suggest that, even at a preverbal stage of human development, connections exist between sensitivity to facial expression (especially that of the eyes) and communicative behaviour: connections that may have an important affective element and, that, either in the real world or that of virtual reality, encourage or inhibit interaction. The following sections of this chapter consider the existence and development of such connections among adult interactants: among people talking to each other.

4.2 Looking and talking: studies of eye contact among adult conversants

As described in Chapter 2, detailed psychological analysis of dialogue is a relatively recent area of study. The same is not true of research into the basic functions and mechanisms of adult interpersonal interaction, including eye contact, although focus on this particular area has been curiously episodic. For a long period, according to Wieser et al (2009), gaze has rarely been included in investigations of face processing: a surprising fact, since 'even in naïve folk psychology gaze direction is regarded as probably the most important social signal' (Wieser et al, 2009: 93)[6]. The

---

[5] Csibra (2007) takes the connection between communication and infant sensitivity to facial expression a stage further. Arguing that the human gaze – given extra salience by the distinctive iris/sclera contrast of the human eye – has evolved for use as an expression of indexical reference, he suggests that the tendency of infants to follow the gaze of an interaction partner demonstrates their ability to use gaze as evidence for a reference to something in the outside world. This, in turn, is an indication that they understand *what* is being referred to: that they are engaging in an early, pre-linguistic version of communication and are therefore enabled to 'learn via communication at an age when they would not be able to understand symbolic or iconic reference.' His later (2010) paper widens the scope of his argument by suggesting that infants' innate sensitivity to eye contact (and to high-pitched, slow-paced 'motherese', or child-directed speech) allows them from the earliest age to practise a fundamental communicatory process: the identification of, and response to, ostensive communication from another person – communication directed at *them*. The recognition of such communicative intentions can then, Csibra continues, help them to comprehend infant-directed communication by 'triggering inferential processes that identify and interpret manifestations of the informative intentions whose presence is implied by the ostensive signals' (*op cit*: 161).

[6] An illustration of the prevalence of such a view is provided by the following advice in a nursing textbook (Walker et al, 2007: 148): 'Eye contact with a friendly smile normally conveys interest and a willingness to engage. People frequently use avoidance of eye contact to signal that they have no wish to engage in social interaction. This may be why nurses become adept at walking through a

situation is now being remedied, with considerable impetus being provided by the development of virtual reality techniques (see Chapter 4.3, below). However, even before this lapse of interest in the subject, a substantial body of findings had already been gathered that identified a range of social functions to which eye contact makes an important contribution and, as will be shown (see Chapter 4.3.2), these findings continue to be highly influential.

As listed in Kleinke's major research review of 1986, the functions served by eye contact are: providing information; regulating interaction; expressing intimacy; exercising social control; and supporting tasks such as learning and bargaining. Within some categories, a variety of sub-functions has also been identified; thus, gaze is used to provide information about liking and attraction, attentiveness, competence, social skills and mental health, credibility and dominance (Kleinke, 1986: 81). Similarly, eye contact as an instrument of social control is used to persuade, deceive, ingratiate, and dominate (increased levels of gaze) and appease or submit (averted gaze). Prolonged gaze (i.e. staring) can be used to elicit avoidance/escape, as demonstrated by Ellsworth et al's 1972 studies of experimenters staring at road-users waiting at traffic lights, which produced significantly shorter crossing times in the subjects (Ellsworth et al, 1972: 302). However, here as elsewhere in the social control area, context and perceived motivation play important parts in determining the nature of a response: people, Kleinke notes, 'comply more to requests from gazing rather than nongazing experimenters in situations in which gaze is interpreted positively' (*ibid*: 83).

Within this body of work, the role played by eye contact in *verbal* performance seems something of a minority interest. Nonetheless, some investigation of this topic has taken place, and this section discusses studies that, incidentally or directly, focus on relationships between gaze and talking, and thus begin to provide formal indications of the control that an attentive Addressee can exert over an adult interactant. Chapter 4.2.1-3 describe research dating from the period leading up to the Kleinke review, with particular reference to the work of Adam Kendon on gaze direction during dialogue and its association with dialogic roles: Speaker or Addressee. Chapter 4.2.4 presents an important change of research focus, in which

---

ward without making eye contact. But while this prevents diversions, it can leave patients feeling ignored and frustrated.'

the impact made by the gazing Addressee on actual communicative competence is analysed, with evidence supplied by the verbal data itself.

4.2.1 Kendon's 'natural history of gaze-direction'

Within the overall area of gaze research, Kendon's studies of naturalistic dialogue are among the most important, with an influence that is still felt 40 years later. They are also among the earliest. As he comments in his 1967 account of gaze direction in social interaction, his main aim is to contribute to 'the almost non-existent literature on the natural history of gaze-direction as it occurs within the context of ongoing conversation between two people' (Kendon, 1967: 24).

In this study, 13 subjects take part in 30 minutes' unstructured conversation, during which they are simply asked to 'get to know one another', and the results – a wealth of detailed data – demonstrate that the relationships between gaze patterns and interlocutor roles are both distinctive and complex: participants' gaze patterns, for instance, are regularly shown to differ according to whether they are Speakers or Addressees. These differences are subject to quite large individual variations; nonetheless, an Addressee tends to look at a Speaker with fairly long gazes, broken by short gazes away, while a Speaker alternates more equally between gazing at and away from the interlocutor, with the gazes away being longer than those used during listening. Furthermore, the gaze patterns of a Speaker differ according to the hesitation patterns created during his utterances: distinguishing between fluent and hesitant speech, Kendon states that a Speaker tends to look at an Addressee much more when speaking fluently than when speaking hesitantly (50% of the time during fluent speech, as compared to only 20.3% of the time during hesitant speech (*ibid*: 39-40). Additionally, a Speaker is more likely to be looking at the Addressee at the moment when he ends a phrase (defined as a 'complete grammatical unit') than when he resumes speaking after such a phrase boundary pause.

The study draws a further distinction between a phrase boundary pause and a hesitation pause (breaks in fluency, perhaps involving interjections like 'um' and 'er'), and demonstrates that, in the former, the Speaker looks at the Addressee as he comes to the end of the phrase, continues to look during the pause, and then tends to

look away as speech starts again. Hesitations, however, are accompanied by a reduction in the extent to which the Speaker looks at the Addressee, and he tends to look back at the Addressee as fluent speech re-starts.

From these variations, Kendon draws an important conclusion:

> We may now see something of the part that gaze-direction plays for the interactants, in regulating each other's behaviour. In withdrawing his gaze, *p* is able to concentrate on the organization of the utterance, and at the same time, by looking away he signals his intention to continue to hold the floor, and *thereby forestall any attempt at action from his interlocutor* [my emphasis]. In looking up, which we have seen that he does briefly at phrase endings, and for a longer time at the ends of his utterances, he can at once check on how his interlocutor is responding to what he is saying, and signal to him that he is looking for some response from him. And for his interlocutor, these intermittent glances served as signals to him, as to when *p* wants a response from him.
>
> (*ibid*: 42)

Kendon goes on to divide the regulatory function thus played by eye-contact into two distinct types of Speaker activity. It allows him to monitor the dialogic situation (notably for Addressee attentiveness); it also appears to enable him to control its development by indicating to the Addressee how – and particularly when – the latter may respond. A salient finding in connection with the former activity is that the Speaker seems to look at the Addressee at 'points of uncertainty' in his discourse: thus, he looks at the Addressee more often at the end of a long utterance than at the beginning, presumably to see whether attention is still assured; he also looks at his interlocutor when he asks a question. The regulatory function, meanwhile, is governed by a different principle: changes in gaze direction, insofar as these are regular, may function for the Addressee as indications of Speaker intentions and expectations, especially in terms of the time available for talking. Thus, the sustained gaze that frequently ends a long utterance indicates that a Speaker is coming to the end of his utterance, and expects an Addressee response. The reverse also occurs: the Speaker can forestall a response from the Addressee by not looking at him.

In a section of his paper that has particularly interesting implications, Kendon considers correlations between gaze-accompanied speech and *speed* of speech

production, and finds a positive result. By computing the speech rates (i.e. the number of syllables per half-second) for 92 long utterances taken from eight of the study subjects, he demonstrates that in all cases the mean speech rate is higher during looking than it is during not looking. Expressed in terms of continuous speech episodes, this translates into 73% of episodes where speech rate was faster while looking rather than while not looking, with the reverse – where looking accompanies a slower speech rate – being correspondingly small.

Enhanced speed of speech – evidence of fluency – when Speaker is looking at Addressee; decreased fluency – i.e. hesitation – accompanying abstraction of Speaker gaze from Addressee; the search for Addressee attention at 'points of uncertainty' in speech production; deliberate abstraction of Speaker gaze as a way of preventing the Addressee becoming Speaker in his turn… what interpretations can be laid on findings such as these? Kendon himself sees them as evidence of a system of information-gathering and signalling between Speaker and Addressee, with the focus apparently more on the former than the latter:

> When *p* [the Speaker] looks [at the Addressee] at the end of his utterances, or at the ends of his phrases within an utterance, he is checking on his listener's responses, in particular he may be looking to see if *q* [the Addressee] is still attending to him. By looking at *q*… he also signals to him that he is giving him his attention, and thus if, in looking at *q*, *p* sees that *q* is looking at him, he sees that he is being 'received'. The mutual gaze, momentarily held, at least, would thus appear to be an integral part of the set of signals that people are on the look out for [in] an interaction as indicators that each is still taking account of the other.
>
> (*ibid*: 48)

However, there are other possibilities, which become apparent if Kendon's data is considered with reference, not to the Speaker but to his partner, and to the timing and impact of the latter's eye-contact with the former.

Kendon himself acknowledges that the overall attention paid in the study to the degree of eye-contact between Speaker and Addressee is general rather than precise: 'We have not …been able to explore what of the other person *p* sees, or is on the look out for, when he looks at him. The presumption has been that when *p* "looks at" *q*, he looks at him in such a way that, were *q* to "look at" him, their eyes would meet.

But this is only a presumption' (*ibid*: 53). However, as his study also shows, the chances that the gaze of *q* – in the role of 'looked-at' Addressee – may coincide with those of the 'looking' Speaker *p* are reasonably good, as the Addressee's own gaze-pattern is that of 'fairly long' gazes at his interlocutor, broken only by 'very brief' gazes away. It therefore seems arguable that, when a Speaker looks at an Addressee, the latter will at least be looking back. There also seems to be a reasonable chance that, subject to individual variations, this exchange of looks will result in actual eye-contact: mutual gaze.[7]

Given these assumptions, the descriptions outlined above of Speaker behaviour may *also* be descriptions of the influence that Addressee behaviour – and particularly of Addressee attentiveness, indicated by eye-contact – can have over Speaker functionality. Indeed, these descriptions are consistent with the possibility that Addressee attentiveness is not just a signal that encourages a Speaker in his utterances, but operates at a more fundamental level, by playing a causal role. Thus, it may be that a Speaker talks faster – and therefore more fluently – *because* of the Addressee's returned gaze: it could be this, not the Speaker's 'checking- up' gaze alone, that enhances the production rate. Thus, too, when a Speaker seeks the assurance of Addressee attentiveness at moments of raised anxiety during his utterances, he may be looking not merely for practical confirmation that his time is not being wasted, but for input at some deeper level that enables him to recover confidence and fluency. Equally – because a Speaker is of course also an Addressee-in-waiting – the Speaker who does not wish to relinquish the floor to his interlocutor is achieving this goal by 'blanking' the other: not merely indicating a wish, but attempting to deprive the other of the enhanced facility of speech that mutual gaze seems to empower.

Kendon does not raise these possibilities and, indeed, the emphasis throughout his discussion seems to be on conscious, deliberate actions and decisions made by his interactants, rather than on automatic responses to stimuli they provide: thus, the Speaker is 'able to concentrate' on his utterances, 'signal' his intentions, 'check' on

[7] A supplementary study of a portion of Kendon's experimental population seems to confirm this assumption, together with the individual variability involved. An examination of five conversations shows that somewhere between a quarter (26.9%) and over two-thirds (70%) of the time spent by individuals in looking at their partners involved mutual gaze, with scores of between 40% and 50% being the most commonly achieved.

interlocutor responses, can 'forestall' an Addressee response by not looking at him. This emphasis may, of course, be mainly an effect of the vocabulary used, and, even if the interactions Kendon observes do take place at a conscious level, his data provides suggestive evidence of the influence that an Addressee has over a Speaker, in addition to the Speaker's impact over the Addressee[8]. However, we have to look elsewhere for indications that this influence takes the form of non-ostensive stimuli rather than ostensive signals: that is, triggers that need not be consciously attended to in order to achieve their effect. This point, which was raised first in the previous chapter, is one to which I will frequently return.

4.2.2 During dialogue, does looking create liking?

Further light is thrown on patterns of eye-contact during speech in a later study by Kendon & Cook and here, alongside the issue of individual gaze habits, the question of affect explicitly enters the discussion. Gaze patterns, they comment (Kendon & Cook, 1969: 482), are a function 'both of the encounter and of the individual himself, but the relative importance of the two factors is uncertain. It is important to know how consistent subjects' gaze patterns are, and how much they are affected by the identity of the person they are talking to.' Earlier work on different aspects of gaze, they continue, seems to point to a single overall conclusion: that looking at a person is a 'signal of liking' and, therefore, the more Interlocutor A looks at Interlocutor B, the more B will respond to this signal by liking A more, because he thinks A likes him. In an experiment again involving 30 minutes' unstructured conversation, they investigate this affiliatory potential of gaze patterns, along with further aspects of gaze timing.

On gaze timing, findings include:

---

[8] It is worth noting, however, that Kendon is well aware that deeper, more automatic levels of interaction exist, as evidenced by his discussion of what he calls 'the Look': The perception of being looked at, he suggests, shows that 'one is being taken account of by another… [and] it seems reasonable to suppose that this will have quite marked arousing consequences (*ibid*: 59)'. He points to the Look's potentially threatening quality, similar to that described by Malinowski and others (see previous chapter): 'To be subjected to the continual gaze of another is a very unnerving experience, for to be the object of another's attention is to be vulnerable to him' (*ibid*: 48). And, acknowledging that external contextual factors play a part in determining the result of gaze-driven arousal, he describes subtle variations in the quality of the gaze itself – 'whether the eyes are "staring", or "narrowed", how the eyebrows, mouth and eyelids are disposed' (*ibid*: 59) – as well as differences in the type of encounter in which it occurs.

1. Length of gaze is a 'consistent aspect' of a subject's behaviour, although this varies according to whether he is speaking or listening
2. Large correlations exist between overall amount of gaze and utterance, and length of gaze and utterance, showing that subjects who 'look or speak a lot do so in long rather than frequent gazes and utterances'
3. 'Some tendency' exists for subjects to match gaze lengths
4. The more the Speaker (S1) says, the more often the Addressee (S2) looks while listening, and the shorter S2's looks are
5. The longer S1's utterances, and the greater percentage of time that he speaks, the shorter and more frequent his gaze while speaking

(*ibid*: 490-1)

On the relationship between subjects' amount of looking-time and their subsequent evaluation of each other, Kendon & Cook find the hypothesis that 'looking creates liking' partially confirmed. While there is no correlation between the percentage of time that subjects looked and their evaluation of each other, a correlation is shown to exist between length of gaze and evaluation, and a negative correlation between evaluation and frequency of gaze. 'It appears that long and infrequent gazes are much preferred to short frequent ones; sheer amount of looking is not sufficient in itself, for it must be correctly distributed. This fits in with popular stereotypes about "steady gazes" on the one hand and "shifty-eyed" people on the other' (*ibid*: 492).

Finally, they propose a choice of possible explanations for some of the above findings, of which the most interesting in the present context is the interactants' tendency to match their gaze lengths. One explanation for this, they suggest, is a simple causal link between the two with imitation as the mechanism that drives it. A second is that some other factor is at work, and this may be related to the success of the interaction as a whole: thus, 'subjects who looked in short frequent gazes, especially while listening, were not liked. This suggests that the pattern of long utterances by one subject and short frequent gazes by both is symptomatic of an interaction that is not running smoothly' (*ibid*: 493). They find the second of these two more likely, and dismiss the first as a 'not very informative hypothesis'. (However, as I hope to show in Chapter 6, Kendon & Cook's doubts over connections between imitation and interaction are perhaps misplaced.)

Overall, they suggest that gaze and utterance patterns are linked in complex ways, but stress that further work needs to be done, 'preferably using confederates whose gaze and action [i.e. utterance] patterns are "programmed"… [thus making it] possible to determine the effect of one subject's gaze and action patterns on those of the other subject' (*ibid*: 493). Again, I will return to this point later (see Chapter 4.3).

4.2.3 Under the gaze of the Addressee

A third explanation offered by Kendon & Cook for the gaze-matching phenomenon is based on the theory of affiliative equilibrium, according to which 'the increase in one subject's gaze length will *cause* the increase in the other subject's gaze length' (*ibid*: 492-3). This theory is advanced in one of the earliest studies of gaze and verbalisation, Argyle & Dean's 1965 discussion of the hypothesised balance between the inclinations – demonstrated by eye-contact and other aspects of intimacy, such as physical proximity – to approach or avoid an interactant. Hypothetical 'approach' functions served by eye-contact include seeking information on communicative success (via feedback on the reaction of the other), and indicating that the information channel is open[9]; affiliative motivation – the establishment and recognition of a social relationship – is a third. However, eye contact may also create anxiety, especially if it is more than the subject feels appropriate. Argyle & Dean propose that this conflict between the approach and avoidance drives is resolved when an individual engaging in social contact reaches an 'equilibrium level' of eye contact, with greater physical proximity (indicating intimacy) being balanced by diminished eye contact and vice versa.

In an investigation of this hypothesis, they observe subjects taking part in a three-minute conversation, seated at different distances – 2', 6' and 10' – from fellow 'conversationalists' (actually confederates who gaze at their interlocutor throughout

[9] Thus, Argyle & Dean explain, an interactant who is in eye-contact with another 'know[s] that the other is attending primarily to him, and that further interaction can proceed' (Argyle & Dean, 1965: 291). Using an example from Goffman (Goffman, 1963: 94), they also point out that eye-contact places a person under some obligation to interact: 'when a chairman or waiter allows his eye to be caught he places himself under the power of the eye-catcher' (*ibid*: 292).

the interaction[10]). And, while the hypothesis itself is of only remote relevance here, the same is not true of some of Argyle & Dean's findings; indeed, they arguably offer early research evidence that – as suggested above – the influence of an Addressee over a Speaker is exercised below the level of conscious awareness.

The first of these intriguing results is an exception to the findings that otherwise confirm Argyle & Dean's equilibrium hypothesis. While eye contact between interactants decreases at closer distances (especially between 2' and 6'), it is never reduced to zero; instead, in conditions where the interlocutors are directly facing each other, there is almost as much eye contact at 2' as at 6' – even though the subjects are 'very uncomfortable' in the former condition. (Attempts to lean backwards are prevented by the position of their chairs; however, they seek to escape from this constraint by looking down, shading their eyes with their hand, narrowing their eyes, scratching their heads, and blowing their noses.)

Two possible explanations are advanced for this finding: either that eye-contact (EC) and proximity do not form part of the same equation, or that there are 'such strong positive forces behind EC that it is difficult to reduce it to zero' (*ibid*: 303). These forces, Argyle & Dean continue, are 'the need for *some* feedback, to ensure that the channel is still open, and to avoid sheer rudeness in view of the conventional social pressures to engage in some EC'. This second interpretation, they suggest, is supported by the fact that eye-contact levels dropped substantially when subjects were placed at right-angles to their interlocutors: a position in which eye contact can be 'more voluntary' and the continual gaze of the confederate less apparent. I would suggest, however, that the participants' dilemma when in the face-to-face seating arrangement arises from a lower-level imperative than mere social convention. When face-to-face, they are caught in a similar triangular relationship (eye-contact + anxiety + speech) to that experienced in the phatic communion situation, with similar results. Inescapably faced with a remorselessly attentive addressee, they maintain eye-contact – and, furthermore, they *keep talking*. As Argyle & Dean note, the conversations recorded were 'perfectly normal'.

[10] The topic for their conversation is suggested by a picture card, about which they have to make up a joint story.

The second finding of interest emerges from the post-experiment interviews with the subjects. These show, Argyle & Dean continue, that 'only one or two persons [out of 80, 24 of whom were eventually placed in the right-angled seating arrangement] realised that they were being gazed at, or that they were talking to a confederate; and their results were no different from those of other subjects' (*ibid*: 298). So here is a picture of Speakers maintaining a flow of 'perfectly normal' conversation, even under difficult conditions, in response to a 'signal' consisting of a constant confederate-Addressee gaze – a signal whose constancy the Speaker is unaware of. Given this unawareness, I would suggest that what the Addressee is providing here is not an ostensive signal, but a non-ostensive stimulus: a spur to action that elicits, not a conscious decision to react to consciously-processed evidence of Addressee attention, but a response to a visual cue, which need not be consciously registered in order to achieve its effect.

In Argyle & Dean's study, the subjects' verbal output is only a by-product of the investigation (sadly, their paper does not include records of the conversations themselves). However, work on gaze patterns by Kleinke et al (1975) shows a slight shift of emphasis, with the investigation of gaze behaviour during dialogue being extended to cover the amount of dialogue that takes place. And this study, rather than providing evidence of a causal link between sustained eye-contact and the compulsion to talk, seems to demonstrate the reverse effect: 'eclipse' by a non-attentive hearer. Here, 54 subjects are interviewed using a pre-selected list of questions for 24 minutes each. Manipulations built into the experiment include the level of interviewer gaze during the interaction: constant, intermittent (with the length of gaze periods controlled by the experimenter), or none. In the no-gaze condition, the interviewer looks either to one side of the subject or down at the question list, but never directly at the interviewee.

Although the impact of interviewer – i.e. Addressee – gaze on the Speaker subject is not the main focus of the investigation, findings produced on this score are rich in implications. They show that:

1. Interviewers in the constant-gaze condition are rated as the most attentive interactants by the subjects, with those in the non-gaze condition being the least attentive

2. Subjects make 'significantly briefer statements' (Kleinke et al: 1975: 120) in the no-gaze condition than in the others
3. Interviewers talk most in the no-gaze condition, although their actual utterance lengths do not differ between the three conditions: 'interviewers talked more in the no-gaze condition only because subjects gave briefer answers and more questions had to be asked [to fill up the time slot]' (*ibid:* 120).

Furthermore, in a final stage of the study designed to show post-interview subject attitudes, subjects with non-gazing interviewers sat farther from them during debriefing, compared to those in the other two groups.

Taken together, the first and second of these findings seem consistent with the hypothesis that perceived Addressee non-attentiveness can significantly reduce Speaker productivity. The third finding arguably takes the hypothesis a stage further, hinting at something like a spiral of interaction decline: the 'non-attentive' interviewer receives brief replies from the subject, and this in turn forces the interviewer to talk more, which will again – due to the apparent lack of interviewer attention – elicit only brief responses. Repeated under non-experimental circumstances, this is a dialogue that will soon grind to a halt.

4.2.4 Gaze windows and the 'listener's meaning'

While the ground-breaking explorations of Kendon and others seemed at the time to open the way to decades of further detailed work on the role of eye-contact in communication, the reverse is true. As Bavelas et al (2002) note in their summary of the literature, Kleinke's 1986 paper represents the last major review of the field, by which point the main thrust of research was turning away again from the specific scrutiny of gaze in dialogue to re-focus on external variables such as interpersonal attitudes or personality differences. However, the continuing minority interest in microanalysis of gaze-utterance relationships was maintained by Bavelas and her own research group, with results of considerable importance within the context of my own research. Here, the interest lies not just in the relationship between eye-contact and communication but, specifically, in the impact of the Addressee's gaze

behaviour on the Speaker's linguistic performance. As she shows, this impact can be considerable.

Expanding their blunt assertion about the 'tenuous foothold' occupied by listeners in most communication theories, Bavelas et al describe the Addressee as commonly seen as either a 'speaker-in-waiting', who is present during the other's speech but not active (reducing conversation to a series of alternating monologues) or as simply non-existent or irrelevant (Bavelas et al, 2000: 941)[11]. They contrast this with the theoretical concepts of grounding collaboration proposed by Clark, Schober & Clark, and others (see Chapter 2), and emphasise the dynamic, moment-by-moment nature of the Clarkian 'collective activity' that takes place when two people use language (Clark, 1992: xvi). It follows, they continue, that the listener's moment-by-moment responses to the speaker 'arguably play an important role in producing the dialogue' (Bavelas et al, 2002: 568), and they suggest that this role can also be seen in enabling the production of narrative. Seen in this light, narrative – which on the face of it is monologic in character – itself becomes a type of dialogue. An important series of experiments, published in 2000-2002, tests this general hypothesis.

The experiments' specific goal is to analyse the types of listener response made during the narration of a 'close call' (i.e. 'near thing') incident. Bavelas et al distinguish here between 'generic' listener responses (standard back-channels such as nodding, 'uh-huh', and 'yeah') which are not specifically connected to what the narrator is saying, and 'specific' ones which, in contrast, are closely connected with what the narrator is saying at a given moment and are not necessarily transferable to other narratives. Examples – such as gasping in horror, imitating the speaker's gesture[12], and adding an appropriate phrase to the story – are restricted neither to audible responses nor to visual ones but, rather, embrace both modes: frequently at the same time. It is by treating these 'visible acts of meaning' (Bavelas & Chovil: 2000: 163) as integrated wholes that, the Bavelas group suggests, the extent of an

[11] They ascribe part of the responsibility for this exclusion of the listener to the 'deeply embedded' influence of the Shannon-Weaver communication model of 1949, in which the channel between interactants is strictly one-way, from sender to receiver.

[12] In the earliest study described below, the latter were summarised by the (naïve) analysts as 'co-telling' acts in which the listener is 'more of an actor than an observer of the story…[He] selects out information from the narrative plot and *acts upon* it. That is, the listener acts like someone in the story (or like the narrator while telling the story)' (Bavelas et al, 2000: 946)

Addressee's role in enabling a successful narrative can be assessed. In their view (2002: 569), these specific responses go beyond simple indications of understanding and contribute to the development of the narrative: at brief but frequent intervals, the listener becomes a co-narrator. Their work builds on that of Kendon and others, with a view to understanding both the timing of the co-narrator's contribution and its impact, the relevance of which is made clear in one of the group's specific predictions for their investigation: if dialogue (including storytelling) is always collaborative, they suggest, then distracting the listener from what the speaker is saying should affect the quality of the storytelling. 'That is, the narrator needs a listener to tell a good story: a good listener is a collaborator, a partner in storytelling' (2000: 945)

The studies concerned are based on two sets of experiments, which analyse listener responses (both generic and specific) to stories told under different conditions. In the first set, participants in 46 dyads take turns to tell each other a 'near-escape' tale, with the listener directed to respond in one of four ways: to 'just listen', to listen with a view to summarising the story afterwards, to listen with a view to retelling the story afterwards, and – the distraction condition – mentally counting the number of days till Christmas. The second set consists only of a single experimental group of dyads (narrators working under a distraction condition; see Chapter 4.2.5, below) and a control group, whose task is simply to listen to the story closely.

One of the resulting studies – in fact, a detailed account of behaviour within the control group in Experiment 2 – focuses on one specific variable: speaker gaze. This, Bavelas et al suggest, is the factor likely to have the 'strongest and most consistent relationship to a listener response' (2002: 569) and, to test this, Addressee responses (both audible and visible) are recorded, with the timing of the gaze interactions between the two being measured and analysed. The findings confirm and extend those of Kendon, and they also seem to confirm my own suggestion that, when a Speaker looks at an Addressee, it is reasonable to suppose that the Addressee will be looking back: that mutual gaze – a 'gaze window' – will occur.

Bavelas et al show that – against the overall background of sustained gaze from an Addressee, briefer and more frequent glances from a Speaker – a regular pattern of mutual gaze takes place at key points in the story, when a Speaker appears to seek a

response by looking at the Addressee. Only when a response is elicited does the Speaker quickly look away and continue with his story.

An excerpt from their data (2002: 571*)* shows three examples of a gaze window in action. Underscores indicate periods of mutual gaze, with listener's responses appearing directly under the words they accompany (a fourth response, accompanying the word *castle,* does not occur in a window):

> 'First of all I have to tell you that my sister calls me surefoot Charlotte cause I
>
> always trip. I'm not, I'm not clumsy but I just, if there's a place to trip I'll find it.
>
> *Mm* *nod*
> *+ nod* *+ smile*
> *+smile*
>
> So we're exploring in the castle and there's this [slight pause] tall ladder going
> *nod*
>
> up to a window, and I was going to climb up to it. '
> *nod*

Overall, Bavelas et al conclude, such timings of the Addressee response support the hypothesis that a collaborative process is at work, achieved via joint action: 'Speaker gaze creates the opportunity for a listener response, and the response then terminates that gaze. Neither of the individuals alone controls when and where a listener response occurs' (2002: 572). This is clearly a more complex pattern of interaction than the Speaker 'signalling' proposed by Kendon, and the Bavelas group themselves reject the suggestion that this is simply a stimulus-response system in which the Speaker evokes an Addressee response. Rather, they propose, it is the Addressee's response that seems to terminate the speaker's gaze and therefore ends the gaze window:

> The listener tended to respond when the speaker looked at him or her, and the speaker tended to look away soon after the listener responded. Together, speakers and listeners created and used the gaze window to coordinate their actions. They demonstrated an efficient and precise use of gaze, not only to regulate turn exchanges as already proposed in the literature, but also to seek and provide listener feedback without signalling a turn exchange. (*ibid*: 577).

However, it can be argued that these findings also support the hypothesis I put forward in the previous chapter: that a Speaker's performance is in some way *dependent* on evidence of Addressee attention. Once this evidence is received, the Speaker is equipped – licensed – to meet the challenge of shaping a key part of his production. Indeed, such a suggestion is by no means inconsistent with the Bavelas group's own conception, developed from that of Schober, Clark, and others, of face-to-face verbal interaction as a process grounded in continuous, dynamic collaboration. 'We propose,' Bavelas et al state, 'that face-to-face dialogue is shaped by social as well as syntactic and semantic processes. That is, dialogue is more than the individuals' production and comprehension of language; there are essential on-line collaborative processes as well' (2000: 941). And, as demonstrated by the body of work that starts with Kendon and includes the grounding theorists, a crucial component of these processes is mutual gaze: eye-contact. Furthermore, as the Bavelas work shows, eye-contact works as a means whereby an Addressee elicits speech from a Speaker, not just at the initiation of a speech episode, but throughout it, on a moment-by-moment basis: the excerpt from 'Surefoot' Charlotte's story, quoted above, takes about 20 seconds to say aloud, but still contains no less than three occasions for Charlotte to seek license to continue speaking, and for her interlocutor to grant it to her.

4.2.5 A badly-bodged story: inducing dysfunction in the Conceptualizer

If the Bavelas study described above provides evidence for a theory of the Addressee as speech elicitor, the results of the experiments which gave rise to the gaze-window work provides dramatic support for the reverse: a theory of the Addressee as speech destroyer.

In Experiment 1 – in which the 'distraction' group have to mentally count the number of days till Christmas – Bavelas et al note that there is a marked contrast between the Addressee reactions of this group and those in the other sets, where Addressee responses of both types occur at the rate of approximately every 3.5 seconds. Amongst the 'counters', however, the rate of generic responses (e.g. listener responses, such as a nod or a 'mm', that are not specifically connected to the

narrative's content) drops to under half that of the attentive Addressee, while the rate of specific responses (those closely geared to the narrative's content, such as a gasp of horror or the completion of a speaker's phrase) falls to less than one-tenth of the attending groups. Furthermore, when rated (by new assessors) on a five-point scale, from 'very poor' storytellers to 'excellent, for a nonprofessional' the narrators in the 'counting' group tell their stories significantly less well than those in the other groups. 'Distracting the listeners,' Bavelas et al conclude, 'affected the overall quality of their narrator's story telling, indicating a reciprocal effect of listener on narrator. No matter how good a story plot is, a good listener is crucial to telling it well' (*ibid*: 947).

Here we find experimental confirmation of the television actor's maxim that, to an actor, an attentive audience is essential (see Chapter 3, n.6). And, in Bavelas et al's second experiment, this point is made even more explicitly. Here, the quality of the close-call narratives is assessed in more detail, with their endings, in particular, measured against four criteria: pace; relevance of any post-climax narrative; absence vs presence of attempts to justify the choice of story; and 'choppiness' (e.g. pauses or disfluencies during or between sentences, producing noticeable gaps). Additionally, Addressees in the experimental[13] group are set a distraction task designed to obviate a problem that might have influenced results in the earlier test, where the required calculations possibly disengaged subjects almost completely from the social interaction of the storytelling. In Experiment 2, distraction-group listeners are asked to count the number of words the narrator utters beginning with *t*, pressing a button each time they hear one: a task that requires them to attend closely to the narrators' words and – crucially – to look at them 'constantly' (2000: 947) , though without attending to the content of the narration itself. Also crucially, the narrators this time are truly naïve: they are *not* told what the listeners have been asked to do.

As the results demonstrate, the stories told under these more demanding conditions – in which specific responses are 'virtually eliminated' (2000: 949) – suffer decisively. They are told less well overall, and the dramatic conclusions are treated particularly poorly. The endings are over-abrupt or 'choppy'; the narrators circle around and tell the ending more than once; and explanations of the close call are often given. As an

[13] Twelve dyads in all, with another 12 dyads in the control group.

example of a badly-bodged ending, featuring already–explained information, dysfluency, and a pointless attempt to justify the perfectly obvious, Bavelas et al quote a tale from a 'particularly skilful' narrator who, with a colleague, had had a narrow escape when working in a logging camp (a double asterisk follows the story's climax):

> So this tree's falling, falling, falling. And he was ahead of me, and I was behind him, and *just* the end of the tree clipped my foot. And it felt like, like a *whip* hitting my foot.** And so ah after I, I mean, I saw it fall and we both go *diving* into the thing cause we knew – I mean, I don't know how exciting that is but afterwards, ah, I mean, we chuckled about it at lunch. Cause it's always funny if you don't get landed on, sure it was a hoot, but (stylised laugh). Um, I just thought that was, ah, that was funny that, ah. Like *usually*, the easy way to go out is go to either side, and that way it'll fall and you're on either side. But since we had no escape room, we knew it was comin' at us, so we had to run for our lives basically, which puts a little excitement into the job too, cause it's fun, rappelling down trees and stuff and, and what-not. So… that's all!
>
> (2000*:* 949)

The narrative faults that build up in the post-climax sentences are startling evidence of some profound dysfunction in the narrator's Conceptualizer. As analysed by Bavelas et al, they include pointlessly 'talking on and on' when the story was over; adding irrelevant information ('rappelling down trees'); a change in delivery from smooth to choppy; and justifying the obviously 'close-call' nature of the account while 'at the same time seeming almost to apologize for or retract the story'. As an output from a system whose functions, in Levelt's words (*op cit*: 9) are 'conceiving of an intention, selecting the relevant information…, ordering this information, keeping track of what was said before,' it is shockingly poor, and this lapse is all the more startling when compared with this Speaker's normal production standards. The staring Addressee has indeed become a 'co-narrator', but a malevolent one.

In conclusion, Bavelas et al point to the continuing reciprocity of the dialogic process that their collaborative model of storytelling implies: 'Even in highly asymmetrical dialogues, speaker and listener roles are not fixed and separate. Rather, their relationship is reciprocal and collaborative, in that the narrator elicits responses from the listener, and *the listener's responses affect the narrator* [my emphasis]… The

essential contribution of listeners must be included to understand language use in face-to-face dialogue' (2000: 951).

4.3 Programmed interactions: virtual discourse situations and their outcomes

Kendon & Cook's 1969 investigation of gaze patterns in social interaction was carried out under conditions that now seem extremely burdensome. The interactants were watched by a team of three observers who physically recorded the onset and ending of each gaze and speech episode with button presses, the timing of which was recorded on punched tape. Kendon & Cook were happy with this procedure – '[it gives] a complete record of the interaction and the gaze patterns of each subject' (*op cit*: 484) – but, as stated above (Chapter 4.2.2), they had doubts about the design of the interactants' own part in the experiment. As with Kendon's earlier investigation, this was simply to 'make each other's acquaintance' during a 30-minute conversation, and the data recorded during this unstructured scenario was in some cases so complex as to be 'rather baffling'. Overall, the authors concluded, further work needed to be done, ideally using confederates with 'programmed' gaze and action patterns: 'In this way, it will be possible to determine the effect of one subject's gaze and action patterns on those of the other subject, and on the way the other subject sees him' (*ibid*: 493-4).

Forty years later, this research ideal has been realised, with results that, while targeted at a specialised area of engineering, are of considerable relevance both to the field of social psychology in general, and to my own investigation. As described in Chapter 2, developments in computer science – and, in particular, in human-computer interface design – are now making substantial contributions to knowledge about human-human interaction, and doing so within experimental constraints that are increasingly tightly controlled. The experimental context here is a computer-generated virtual environment (VE), in which human subjects are invited to take part in conversations with interactants who are either computer-generated or real-world interlocutors whose reactions are edited online by the computerised environment. The theoretical context within the field of human-computer interaction is that of

'presence' and 'co-presence' research, the distinction between which is explained by Garau et al (2005: 105). The first 'aims to understand what leads to people's sense of "being there" in the virtual environment (VE) despite knowing rationally that it is not "real"'[14], while the second 'aims to understand how to enhance the sense of being with other people in the VE and is of interest for all those applications that involve some form of social interaction, from collaborating with remote human users… to practising public speaking with a virtual audience.' (The same source also explains the difference between two widely-used labels for 'human' inhabitants of a virtual room, 'avatar' and 'agent': when a virtual human represents a real human user (e.g. the partner in a computer-mediated remote conversation), the image is referred to as an avatar, while an agent is 'driven by an artificial intelligence or by simple pre-scripted behaviors… The challenge with human-human communication is to drive avatar behaviors that enrich, rather than hinder, communication between remote participants ' (*ibid*: 105).)

Within such a context, research outcomes demonstrating the validity of the whole VE experience as a simulacrum of human-human communication are as important as those describing the behaviour of the human participants in detail. Nonetheless, these results, derived from work in a variety of VE paradigms, show in a tightly controlled manner how Speakers react when placed in dialogic conditions like those described in the present chapter and the previous one: when confronted, for example, with an audience that goes to sleep (cf Chapter 3.3); or whose gaze patterns do not fit the expected rhythm of a conversation (cf Kendon & Cook's comments on gaze distribution (Chapter 4.2.2)); or even – in a possible parallel to the infant behaviour observed by Murray & Trevarthen and others (Chapter 4.1) – when faced by an Addressee whose expression seems less eagerly receptive than expected. The following sub-sections of this chapter outline some of this work which, as will be seen, offers further experimental support for the extended model of the Speaker that I am trying to construct.

[14] The finding that humans behave to computers *as if* they were people is now well-established, as demonstrated in Reeves & Nass's 1996 review of human interaction with new media (phenomena described here include people being polite to computers).

4.3.1 In the VE suite: the public speaker's nightmare

One finding on Speaker-Addressee interaction that has emerged from the virtual environment provides noteworthy support for Bavelas et al's theory of the listener as co-narrator, while at the same time confirming the fears that all public speakers experience introspectively. In a virtual reality study of humans interacting with different types of computer-generated audience reaction, Pertaub et al (2001) demonstrate that giving a short prepared speech to an indifferent (virtual) audience does indeed have a negative effect on speakers' communicative abilities – even though they are perfectly aware, at a conscious level, that the audience is an imaginary one.

Given the general value of virtual environment paradigms to research into naturalistic situations (including naturally-occurring speech), it is interesting to consider the details of Pertaub et al's methodology, as well as its application. Although the overall goal of their study is to assess how far virtual reality presentations can be used in habituation therapy for social phobia, the response variables used in the assessment relate strongly to the subjects' communicative abilities and, in some cases, specifically to their verbalising abilities: for instance, statements requiring a post-experiment yes/no response include 'I was in constant fear of forgetting my speech' and 'My thoughts became confused and jumbled when I spoke before the audience' (Pertaub et al, 2001: 9). The second of these has a clear relevance to my own inquiry: participants are here being invited, in Leveltian terms, to reflect on the workings of their Conceptualizer.

One of the two independent variables in the experiment is the type of computer-generated audience behaviour which greets the speaker[15]. Three virtual audiences are involved, each consisting of eight male avatars. One audience consists of a 'friendly' group, whose facial and body animations express support for the speaker and engagement with the topic; eye contact here is maintained about 90 per cent of

[15] The second variable is the degree of immersion experienced in the virtual reality scenarios: half the subjects give their speech (a five-minute presentation, prepared in advance) when wearing a virtual reality headset that places them in a stereo version of the seminar room, and half present their talk to a desktop monitor showing the room and audience. Interestingly, the impact of the headset-generated experience appears to have been greater for the female subjects in the group. A possible cause may have been less experience with computer game-playing amongst females, but a check on this showed no significant difference between the males and females taking part in the study.

the time. The second virtual audience, in contrast, is a speaker's nightmare: the avatars in this negative scenario 'fell asleep, slouched in their chairs, slumped forward on the seminar table, orientated themselves away from the speaker, leaned backwards, put their feet on the table, *avoided establishing eye contact* [my emphasis], and one even got up and walked out of the seminar room during the talk' (*ibid*: 6). The third audience is neutral in terms of emotional expression and – unlike the other two – is completely static throughout.

An additional feature of the animated scenarios is speech: scripted 'back-channel' responses for the audience members, ranging from 'I see' and 'That's interesting' to 'That's absolute nonsense'. Both these audio inserts and the avatar animations in general are contingent to some extent to the ongoing interaction with the human participant, as they are controlled by an unseen human observer. 'We attempted,' Pertaub et al explain, 'to ensure some form of co-ordination between the responses of the audience and the content of the speech being given. An operator seated at a remote terminal unseen by the subjects used the distributive capabilities of the virtual environment to trigger the next reaction in the sequence at an appropriate moment… The flexible timing of the animated responses and audio clips was deliberately incorporated into the scenario to foster a sense of interactivity and avoid making subjects feel that the audience was responding at wholly inappropriate points in the talk' (*ibid*: 6-7).

Before the experiment, the 40 subjects involved in the study complete a questionnaire that indicates how they generally feel about speaking in public. This includes negative statements such as 'My thoughts become confused and jumbled…', together with positive ones such as 'I have no fear of facing the audience'. Afterwards, they complete a modified version in which a tense change ('My thoughts *became* confused and jumbled…') relates the situation to the one they have just experienced. Correlations between these before/after scores under the three audience conditions provide the most important results of the study.

Pertaub et al find that the results from the negative group differ dramatically from those from the positive and neutral sets. In both the latter cases, there is a positive correlation between the groups' pre- and post-experiment questionnaires, showing that their response to their avatar audiences can be predicted from their *normal* level

of anxiety about public speaking, indicated by their responses to the pre-experiment questionnaire. However, this is not the case with the subjects who face the inattentive audience. As demonstrated by the lack of correlation between their two questionnaire scores, their response to the hostile situation tended to be one of 'high anxiety' (*ibid*: 13), irrespective of their normal level of concern over speaking in public. That is, the audience's lack of attentiveness has wrecked their confidence in themselves as Speakers, however high this had been before the talk. Furthermore, and crucially, this damage has taken place despite the subjects' full awareness that their audience is a virtual one only – an effect that might be compared to a case such as the Muller-Lyer illusion, where the subject's intuitive response that the lines are of different lengths is not affected by the knowledge that the lines are in fact of the same length. It is also an indication of the role played by affect, as opposed to cognition, in speech production, and it appears to extend the suggestion made earlier (Chapter 3.2) that the Leveltian self-monitor is under the dual control of conscious and automatic mechanisms. Notwithstanding Levelt's assertion that speaking is 'usually an intentional [and therefore consciously planned] activity' (*op cit:* 20), this raises the question of whether the highly controlled planning that drives the Leveltian model of the whole Conceptualiser might *also* be subject to a dual-control process, with automatic, pre-conscious reactions to the external situation taking a share in utterance creation. As will become clear in the next two chapters, the automaticity of interaction processes usually seen as consciously controlled is a major theme of my investigation.

Meanwhile, further light is thrown on the nature of such reactions in later work by the same group (Slater et al, 2006), which demonstrates that phobic public speakers feel anxiety when speaking to a virtual audience, but *not* when addressing an empty (virtual) room. In this further test of 'presence' – of whether people respond naturally to virtual representations, even though they know these are representations only – the aim is to present the subjects with less extreme virtual realities than those used in the positive/negative study, thus reducing the likelihood of eliciting a response simply because the situation is so extreme. The virtual audience created for this study is therefore a neutral one who, though they make eye contact and indicate varying degrees of attention, do not exhibit either the enthusiasm or the negativity of the earlier experiment. The subjects are speakers who, in response to a preliminary test

on fear of public speaking, have shown themselves either highly confident or phobic. Reactions from representatives of both groups addressing the neutral audience are compared to those elicited by delivering a talk to a virtual environment consisting simply of an empty room.

The results show that, in front of the virtual audience, both groups respond according to type: the confident speakers are unaffected by the presence of the audience, whereas the mere fact of the audience's existence created an anxiety response in the phobic group. Pertaub et al comment that this is what would be expected in similar circumstances in real life, and thus counts as a 'presence' response (*ibid*: 8). However, the reaction to the 'empty room' situation is different: while confident speakers remain confident, the phobic subjects display significantly less anxiety when faced with an 'empty room' than their counterparts do when faced with a virtual audience, even though it is a neutral one. At some level of awareness, their Conceptualizer is not fooled: it knows the difference between solitude and the presence – even an imaginary one – of an Addressee, and reacts accordingly.

4.3.2 Looking and talking with avatars

The relevance to speech decision-making of the work just described lies principally in the way it supports and extends Bavelas's model of the 'listener as co-narrator': one who, by the nature of their reactions – or their absence – plays an active part in constructing a Speaker's output. As is obvious, the participants' productions are here all monologues, delivered from the (virtual) distance of a speaker's podium. However, VE research also shows examples of virtual dialogue which again support the extended model of the Speaker that I am proposing.

Amongst the data thus made available, the evidence of the specific power of eye-contact is particularly useful, both in my own context (as explained in the previous chapter) and within the VE field itself. Here, it is the richness of its effects that makes it valuable. As explained by Garau et al (2001: 1), in a reference back to the studies of Argyle and Kendon, it serves at least five distinct communicative functions: regulating conversation flow, providing feedback, communicating emotional information, communicating the nature of interpersonal relationships and

avoiding distraction by restricting visual input, and the importance of gaze in virtual communicative environments is one of the main subjects of the Garau group's work.

In their 2001 experiment, subject pairs conduct a 10-minute conversation over a video link under one of four conditions. In one, interlocutors see and hear each other normally; in another, they only hear each other. In the third, the image of the interlocutor is replaced by an avatar whose head and eye movements are not related to the conversation, except by accident. In the fourth, the head and eye movements of an avatar 'interlocutor' are determined by the conversational role – Speaker or Addressee – of the subject that it represents, using the earlier findings of Kendon, Argyle and others to determine differentiations between the avatar's gaze lengths when speaking and listening.

The response variable, assessed by post-experiment questionnaire, is 'quality of communication', classified under four broad headings: the extent to which the subjects have felt the experience was like a 'real face-to-face conversation'; the extent to which they felt involved in the conversation; the extent of experienced co-presence (i.e. of interacting with another person rather than with a computer interface); and the extent to which subjects positively evaluate their partner and enjoyed the conversation.

The results demonstrate that, while the video environment produces the most satisfactory reaction to the conversation, the conversations involving the 'inferred gaze' avatar – i.e. the one whose gaze behaviour was related to the conversational role it was taking – produce the second highest score on three out of the four criteria (the exception being co-presence). Indeed, the inferred gaze score does not significantly differ from the video one in terms of involvement and 'face-to-face' experience: a finding that both confirms the validity of using virtual environments in research on human social reactions and again underlines the importance of gaze in any dialogue, virtual or real-world. The inferred-gaze avatar also 'consistently and significantly' outperforms the random-gaze avatar and, in the overall analysis, significantly outperforms the audio-only condition. According to Garau et al, this suggests that in contrast with a randomly animated avatar, one with meaningful animations can indeed contribute to perceived communication quality' (*ibid*: 7). Meanwhile, the 'meaningless' head and eye movements of the random avatar

actually perform worse, if anything, than the audio stream in terms of developing a realistic co-presence: as one subject comments, 'It just kind of sat there doing something rather than helping.' (*ibid*: 7)[16].

Further work by Garau's group[17] has confirmed the positive impact of inferred – as opposed to random – virtual gaze on human subjects, but with an added dimension of some interest. In Garau et al (2003), a study of 48 subjects is described which involves two independent variables: inferred vs random avatar gaze and high vs low realism of the avatar's appearance. (The 'low realism' representation is a Disney-like figure of indeterminate gender; 'high realism' avatars are clearly distinguishable male and female humanoids.) Using a procedure and response variables similar to those described in the previous experiment, the group demonstrates that the match, not just between Speaker and 'Addressee' behaviour, but between 'Addressee' behaviour and *appearance* makes a significant difference to how far the conversation is perceived as successful by its human participant. For the lower-realism avatar, for example, the (more realistic) inferred-gaze behaviour reduces face-to-face effectiveness, whereas for the higher-realism avatar, effectiveness is increased by the same, more realistic, inferred-gaze behaviour.

> For the lower-realism avatar, the inferred-gaze model has a consistently negative effect on each response variable… The opposite is true of the higher-realism avatar… Consistency between the visual appearance of the avatar and the type of behavior that it exhibits seems to be necessary: low fidelity appearance demands low fidelity behavior, and correspondingly higher fidelity appearance demands a more realistic behavior model (with respect to eye gaze)'.
>
> (Garau et al, 2003: 6)

A secondary, but also intriguing, finding concerns the impact of the audio input of the experience on the human partners. Although, with the higher-realism avatars, the inferred-gaze interaction model outperforms the random one in creating a 'face-to-face' experience for participants, and also induces a sense of co-presence and a

---

[16] Arguably, this distinction between random and conversation-related gaze – and its differing impact on the conversation itself – gives support, from an adult experimental population of 100, to a phenomenon merely glimpsed in Murrey & Trevarthen's study of four young infants: the negative reaction (detachment, confusion, puzzlement) of an interactant when a partner's gaze pattern appears inappropriate to the interactant's own behaviour.

[17] The Virtual Environment and Computer Graphics group, University College London.

positive partner evaluation, it does not create a greater sense of 'involvement' (the other response variable). Here, neither the type of avatar nor the type of gaze makes any difference to the participants' sense of absorption in the conversation, or their ability to keep track of it. Instead, the 'overwhelming majority' of respondents claim that their attention is focused on the avatar's *voice* rather than its appearance. Garau et al suggest that this result might be partly ascribed to the relatively minimal behaviours designed for the avatars, and stress the significant impact of the avatars' visual appearance on the other response variables. This impact even extends to the participants' physical behaviour: 'Without exception, all participants stood facing their partner's avatar throughout the entire conversation. They took care to maintain a suitable interpersonal distance and felt *compelled* [my emphasis] to display polite attention' (ibid: 7): a manifestation, in a laboratory-contrived environment, of the compulsion to obey the behavioural and communicative commands created by the presence of an 'aware', attentive Other: the reaction described in Chapter 3.

4.3.3 'Damped' reactions: creating Addressee indifference to order

As demonstrated by the research described in this section, the degree of control that can be exercised over the interlocutor actions being studied has increased dramatically since Kendon and Cook described their ideal, 'programmed', experimental paradigm. Indeed, the use of novel VE techniques is now opening the way to achieving in full this ideal of systematically testing dyadic interactions between, not just a human and a humanoid, but two human conversants, talking – as they think – naturally to each other. This, as Robinson & el Kaliouby comment in their introduction to a Royal Society collection of papers on emotional expression in man and machines, has so far presented a considerable challenge because of the range of parameters involved (Robinson & el Kaliouby, 2009: 3442). The work they are introducing is a study by Boker, Cohn et al, which translates the head movements and facial expressions of a human interactant (actually a confederate) into those of an avatar replica, making it possible to study, on the one hand, the reactions of the naïve dialogue partner to these programmed modifications, and on the other, the reactions of the real-world confederate to the reactions thus artificially stimulated in the naïve participant.

The purpose of this experiment is to investigate what happens when the expressiveness of an Addressee's reactions to a Speaker is systematically attenuated, or 'damped' – an outcome of considerable relevance to my own research – and the methodology used is also of interest. The 27 naïve participants and the six confederates hold two eight-minute conversations with each other over a video-conference link. The confederates see an ordinary live video of their partners, while participants are informed that the video image they will see of their interlocutor has been edited to cut out forehead, ears, and other surrounding features. However, this is not the only image manipulation that has taken place. What the participants actually see is an avatar of the confederate, created on the fly with motion tracking software, which relays the confederate's head movements and facial expressions – and, at one minute intervals, reduces their expressivity. Thus, a smile can become slightly less wide, a nod of the head less vigorous.

Viewed individually, these modifications appear extremely subtle and, indeed, go unnoticed (at a conscious level) by the naïve participants: according to Boker, Cohn et al (2009: 3488), none of the participants mentioned that they thought they were speaking with a computer generated face, or noticed the experimental manipulations. The study's findings show, however, that these unremarked manipulations have a regular and complex effect. When faced with an attenuated nod or turn of the head from the avatar, or a damped facial expression, naïve participants *increase* their own head movements. But this apparently compensatory reaction then sets in train an *imitative* one from the confederate who, while aware of the purpose and method of the experiment, does not know when his image is being manipulated: witnessing these increased head movements of the participant, the confederate responds by automatically increasing *his own*.

In discussing these findings, Boker, Cohn et al suggest that the participants' increased expressivity may be an attempt to elicit a response from their partners in line with what was expected, rather than the attenuated version: a possibility that parallels the initially-increased infant activity noted in the Murray & Trevarthen 'still-face' study (Chapter 4.1.1), with its established origins in negative affect. They also put forward the hypothesis that the head movement responses, in particular, indicate the existence of a 'shared equilibrium' in dyadic interaction; thus, when one

conversant's perception is disturbed, both conversational partners respond in a way that compensates for the disturbance. 'It is as if there were an equilibrium energy in the conversation and when we removed energy by attenuation, and thus changed the value of the equilibrium, the conversational partners supplied more energy in response and thus returned the equilibrium towards its former value' (Boker, Cohn et al, 2009: 3492).

Stressing the importance of possible connections between one interlocutor's head movements and the other's facial expressions, they describe circumstances that also fit the problems experienced by the failed story-tellers observed by Bavelas et al (*op cit*: 2000; cf Chapter 4.2.5):

> The attenuation of facial expression created an effect that appeared to the research team as being that of someone who was mildly depressed…[It] can also be related to cognitive states or social context. For instance, if one's attention is internally focused the attenuation of facial expression may result. Interlocutors might interpret damped facial expression of their conversational partner as reflecting a lack of attention to the conversation.
>
> (*ibid*: 3493)

The increased head movement of the subjects in the avatar study 'may have been' efforts to elicit more responsive behaviour in their partners; it would be interesting to know whether the videos of Bavelas et al's narrators show similar elicitatory responses. (It is, of course, possible that the choppy, over-extended endings of the story-tellers in the Bavelas distraction conditions represent a *verbal* attempt to elicit some kind of appropriate reaction from the Addressees.) Also in this connection, Boker, Cohn et al recall investigation of infant behaviour when confronted with simulated maternal depression. Here, as in the Murray & Trevarthen study, the subjects tried to elicit a change in their mother's responsiveness, in this case by smiling at her, turning away and back, and smiling again: 'when they fail to elicit a change in their mothers' behaviour, they become withdrawn and distressed' (*ibid*, 3493, referring to Cohn & Tronick 1983). They comment: 'Had we attenuated facial expression and head motion for more than a minute at a time, naïve participants might have become less active following their failed efforts to elicit a change in the confederate's behaviour. This hypothesis remains to be tested' (*ibid*: 3493).

The results of such tests would be extremely interesting, as would an analysis of the content and delivery of the speech that accompany these 'damped' interactions. Findings here could perhaps bring together theories that have emerged from developmental studies, on the one hand (cf Murray & Trevarthen, Striano) and sociolinguistics, on the other (cf Bavelas et al), and thus contribute to a unified account of the relationships between Speakers and Addressees.

4.4. Summary and conclusions

This chapter reviews evidence, drawn from a variety of research areas, of the impact made during face-to-face verbal (or pre-verbal) interaction by facial expressions and – in particular – by eye contact. It opens with an account of the robust effect produced in infants as young as six weeks old by an interactant who 'blanks' them, either by looking away or by staring at them without expression or response (the still-face effect). The overall picture that emerges is one of an initial attempt by the infant to regain interlocutor attention, followed by behaviour that may be interpreted as evidence of distress, together with withdrawal from the interactive situation: responses that are clearly affective rather purely cognitive.

The second main section of the chapter discusses a range of studies of eye contact between adult interactants, starting with Kendon's ground-breaking and very detailed work on gaze-direction. Kendon's findings show that there are clear gaze-direction patterns on the part of both Speakers and Addressees, and that there is a positive correlation between gaze-accompanied speech and its speed: when the Speaker is looking at the Addressee, his mean speech rate is higher. Another aspect of Speaker-Addressee gaze behaviour is described by Argyle & Dean, who put conversationalists in uncomfortably close physical positions in which eye contact is (artificially) maintained; despite their discomfort, however, the subjects continue to talk and the conversations are 'perfectly normal'.

Records of the speech content of these early studies are not available but, with the more recent work of the Bavelas group, the relationship between content and eye-contact moves into the foreground. Bavelas et al's main premise is that an

Addressee's responses – hitherto overlooked in most investigations of communication – play an important role in dialogue production. For this reason, narrative can be regarded as a limited case of dialogue, even though only a single Speaker is involved. The group's work outlined above describes narratives told under a variety of Addressee conditions, and its findings include evidence that Speakers and Listeners collaborate to create 'gaze windows' – episodes of mutual gaze – that help them coordinate their actions. Also included is evidence of how listener indifference can dramatically affect a Speaker's performance: stories told to staring but abstracted Addressees emerge severely damaged, with endings repeated, irrelevant information offered, and delivery broken up by pauses and interjections.

The third main section of the chapter considers evidence from a relatively novel research field that offers much future promise for studies of social interaction: the development of computer-generated virtual environments. Among the findings here is confirmation of Bavelas et al's theory of the Addressee as co-narrator, with a demonstration (see Pertaub et al, Chapter 4.3.1) that Speakers' communicative abilities are impaired when giving a speech to an inattentive virtual audience – even though they are aware at a conscious level that the audience is an imaginary one. Research on virtual dialogue, which builds on the work of Kendon and others, also demonstrates the importance to the Speaker of appropriate gaze behaviour from the Addressee: for example, participants exposed to a variety of Addressee responses (some of which are computer-generated) prefer an 'Addressee' whose gaze patterns conform with Kendon's findings, as opposed to one programmed to react randomly. Meanwhile, another study of computer-mediated dialogue gives detailed evidence of how Speakers react when exposed to artificially-diminished attention on the part of a (human) interlocutor: they appear to increase their efforts to engage and retain their Addressee's interest. The chapter closes with the suggestion that further work with virtual reality environments could offer valuable further insights into Speaker-Addressee interactions, especially if combined with analysis of accompanying (human) speech content.

Overall, these findings appear to support the hypothesis outlined in Chapter 3: that speech production can be strikingly influenced by its hearer's attentiveness or the lack of it. In particular, this influence seems to affect the message level where, in

Levelt's framework, a Speaker is seen as forming an intention, selecting and ordering the relevant information, and keeping track of what was said before. It is these aspects of speech production that suffer from the Addressee's abstracted stare in Bavelas et al's experiments, and from the indifference of the virtual audience in the work of Pertaub et al.

The evidence from the virtual reality studies, in particular, also gives support to my suggestion that the mechanisms involved at the conceptual level of speech production are of two types – conscious/controlled and automatic. The conscious knowledge that a nightmare audience is a mere fiction does nothing to ease the speech-maker's discomfort, and the reverse also holds: phobic speakers display anxiety even before a neutral virtual audience but, when giving a speech to an empty (virtual) room, lose their fears. Additional support for dual-process theories of speech production is provided by the existence of coordinated, moment-by-moment speech patterns established by Kendon and others; given the speed and persistence with which these occur, they seem less likely to be the product of intentional (controlled) decision-making and action than of the low-level, automatic system of Speaker-Addressee interactions that underpins my main hypothesis.

All this raises a further question. If Addressee non-responsiveness damages an interlocutor's ability to produce speech, what must a Speaker do to prevent this? The solution seems obvious: he must do all he can to engage and retain his Addressee's attention – as, of course, must the Addressee when it is his turn to speak. For both, making their contributions relevant to the other is crucial if they wish to achieve their communicative goals by successfully continuing their dialogue. It can therefore be argued that, unless a Speaker achieves and continually maintains relevance to an Addressee, the Addressee's interest will be lost, and this will in turn impair the Speaker's production abilities and hence the communication process itself.

The implications of this suggestion – including the means by which such a feedback mechanism can be supported – are explored in the following chapters

.

Chapter 5

Automaticity in social interaction

Introduction

In this chapter, I propose an answer to my two research questions and explore some of the issues connected with it. Since the discussions that have led me here have been wide-ranging, I start by reviewing the ground covered so far: the existing research background (mainly drawn from psycholinguistics) discussed in Chapter 2; the hypothesis of low-level Addressee influence proposed in Chapter 3, backed by examples of naturally-occurring phatic communication; and Chapter 4's discussion of evidence from a variety of sources that supports this hypothesis.

I then propose the central claim of my thesis (Chapter 5.2): that Speaker sensitivity to Addressee needs is not merely a desirable accompaniment to verbal interaction, but a prerequisite to its success. My suggestion – which combines both my research questions in a single premise – is that, unless a Speaker continually maintains Addressee attentiveness by achieving relevance, the loss of the Addressee's interest will damage the Speaker's performance, and the dialogue itself will suffer. In the following sections, I discuss ways in which the Speaker can maintain this feedback loop successfully and consider how far the mechanisms involved operate automatically rather than intentionally.

I start by discussing a powerful tool for investigating the automaticity issue, dual-process theory, and review recent findings on the two types of cognitive system involved, the relationship between them, and their connections with memory processes, on the one hand, and the massive modularity hypothesis, on the other. I also draw attention to an apparent gap in the literature: detailed applications of the theory to language use itself.

As a step towards applying the theory to this seemingly under-researched area, I identify aspects of social interaction that support Addressee-sensitive behaviour and have been shown to originate from neural areas associated with automatic processing. Amongst these, one neural region in particular has been implicated in the distress caused by social exclusion (Chapter 5.4.2) and I suggest that this may also be involved in the sharp decrease in Speaker skill that accompanies Addressee 'blanking'. I also survey evidence on the neural foundations of responsiveness to facial movements.

Two other areas of research may provide further insights into the automaticity of processes supporting social interaction: mirror neuron theory and Theory of Mind. After looking briefly at some findings from this research, I end by considering possible interactions at a neural level between automaticity and the language production process itself.

Principal sources are: dual-process theory (Chapter 5.3), Frankish & Evans, Smith & DeCoster, Mercier & Sperber; cognitive neuroscience (Chapter 5.4), Lieberman, Eisenberger et al, Baron-Cohen, Schilbach et al, Adolphs; mirror neurons (Chapter 5.5), Gallese, Shamay-Tsoory et al, Rizzolatti & Arbib, Iacoboni et al, Brass et al, Hickok, Cisek & Kalaska, Catmur et al, Neal & Chartrand; Theory of Mind (Chapter 5.6), Baron-Cohen, Frith & Frith, Rilling et al; language and automaticity (Chapter 5.7), Ullman, Adolphs.

5.1 From models of the Speaker to models of dialogue: the argument so far

The two questions with which this inquiry began – about a Speaker's choice of content and its relevance – were simply put, but have emerged as far-reaching in scope. While a possible answer to both of them can now be attempted, the background to these suggestions is, as the previous chapters have shown, a large and complex one, and this therefore seems an appropriate moment to review the complete argument so far.

Its starting-point is Levelt's influential model of the Speaker as a solo performer, processing speech decisions from the moment of their inception to their emergence as an audible linguistic product, ready for consumption by an Addressee. Although Levelt acknowledges the importance of the Speaker's function as an *interlocutor,* a partner in a speech activity ('The speaker's skills of language cannot but be tuned to the requirements of conversation', *op cit*: 29), the Leveltian Addressee remains in modelling terms something of an abstraction and, indeed, as somehow superfluous to the whole production process, since the Leveltian Speaker is also his *own* Addressee, monitoring his utterances at every stage from conceptualization to physical verbalisation. Meanwhile, the Speaker's knowledge of his actual, real-world Addressees – who they are, where they are etc – is merely one of the constituents of the declarative knowledge store that contributes to his speech decisions: a relationship that seems to fit with Bavelas et al's crisp comment about the 'tenuous role' occupied in most theories by listeners (Bavelas et al, 2000: 941).

As the relevance of a Speaker's output to an Addressee was the subject of my other research question, my first priority was to explore theoretical approaches to speech production that pay greater attention to the role of the external Addressee than the one that Levelt gives it, with special reference to the 'common ground' model developed by H.H. Clark and others, and to Pickering & Garrod's theory of automatic dialogic interaction. Both of these, in emphasising the interlocutor's importance in any theory of the Speaker, also considerably expand the theory's scope. Indeed, they double it, for both stress the idea that the Speaker-and-Addressee is a single unit, working *together* to further the goals they bring to the interaction. As Clark comments, it is only when we study two people talking to each other that many essential aspects of these contributions will emerge.

It is Pickering & Garrod's model of dialogic interaction in particular that opens the way in my narrative to addressing the two aspects of Speaker competence that I have selected as focus: message generation and relevance assessment. A central component of this model, which features a tightly-interlocking system of Speaker-Addressee alignment at multiple levels, is the mechanism that enables such alignment to take place. For Pickering & Garrod (2004:172), this is a 'primitive and resource-free' system, based on priming and – importantly – operating automatically,

which enables the lexical, syntactic and articulatory choices of Speaker A to be effortlessly absorbed and reproduced in turn by Speaker B. The introduction of automaticity to a model of the Speaker is not in itself new: Levelt also stresses its importance, though principally in connection with the formulatory and articulatory sub-systems that convert a message into its final form, working top-down: 'There is no feedback from processors down the line (except for some Formulator-internal feedback)' (*op cit*: 16)[1]. As envisaged by Pickering & Garrod, however, the contribution of automaticity to dialogue is far more complex, with the priming process operating in each interlocutor in *both* a top-down and a bottom-up way, with additional interchanges at each sub-system level. As a result, the interlocutors' output and the conceptual representations underlying it are claimed to converge, ensuring mutual comprehension without the need for conscious audience modelling. By virtue of their alternating roles as Speaker and Addressee, both parties in a dialogue thus in effect guide each other to think and talk about the same thing in the same way.

Does this process go further, and guide the interlocutors' actual choice of content? Pickering & Garrod draw the line at this point, acknowledging that their model deals with the structural aspects of speech production and interpretation rather than directly with content determination: 'We propose that alignment channels only produce a direct link between the structures that the interlocutors use in language processing… [The process] does not determine exactly what the speaker produces (as this depends on his long-term knowledge) or how the addressee interprets what he hears "beyond" the level of the situation model' (*ibid*: 213). Nevertheless, their development of the automaticity issue – and particularly its focus on the ongoing interaction between Speaker and Addressee – represents a major advance in theories of speech production. It not only sheds more light on the striking fluency and flexibility of naturally-occurring speech than Levelt's 'solo performer' model; it also – to a greater extent, Pickering & Garrod claim, than the theory of groundedness – addresses the issue of how relevance to an Addressee can be achieved when talking at the natural speech rate of two to three words per second.

---

[1] While he accepts that automatic, *en bloc* retrieval of complete message strings is also possible, he sees processing at the conceptual level as mainly under a Speaker's central, conscious, control: a view to which my own account proposes some modifications.

The explanation in both cases focuses on the possibility that these characteristics may derive, not from conscious decision-making by the Speaker, but from Speaker-Addressee interactions that operate automatically at a more basic cognitive level, and determine to an important extent the speech that is actually produced.

This proposal is significantly extended by work in the apparently unrelated field of computerised natural language systems. Here, as I describe in Chapter 2.5, the suggestion has been made that low-level, automatic mechanisms can process not only verbal but also non-verbal signals from a dialogue partner, and that the latter also play a role in achieving Speaker-Addressee alignment; moreover, it has been suggested that affective factors may make an important contribution to successful dialogue behaviour, alongside cognitive ones. The potential extension of dialogue theory to include non-verbal behaviour – behaviour that a listener can display while still *remaining* a listener – opens up a new area of Speaker-Addressee interaction for exploration, and one that, while focusing primarily on the Addressee, paradoxically may throw some light on the automaticity or otherwise of some of his partner's speech output. In particular, it may explain an aspect of naturally-occurring dialogue that has received little attention to date, but which has an obvious significance: what is it that brings a Speaker to open a dialogue in the first place (or, indeed, abandon one)? In other words, what makes verbal interaction succeed or fail?

5.2 The successful Speaker and the Addressee-Speaker feedback loop

It is at this point in my argument that I propose the central hypothesis of this thesis: that a low-level, automatically-operating system of interaction exists between two interlocutors which enables, not the Speaker, but the *Addressee* to influence a dialogue, thus ensuring that the Speaker's output remains continually processable at the Addressee end. I also suggest that this hypothesis, if confirmed, could provide some answers to the two questions with which I began: Why do Speakers say what they do? And how do they usually achieve optimal relevance, given the time and other constraints involved?

As described in Chapter 3, I initially explore this proposal by considering informally-gathered examples of natural, spontaneous dialogue. Most of these are cases of phatic communion, or small talk: a somewhat under-researched category of communication chosen here on grounds of its social prevalence, its inherently dialogic nature and its close connection with the affective factors that have been shown to have an important impact on successful communication. Even at this informal level, the results throw some interesting light on the issue of speech automaticity, in that they both illustrate and – in one area – extend the limited set of automatic processes that Levelt ascribes to the message level in his model. They certainly demonstrate the existence of 'whole messages' (*op cit*: 21), accumulated through experience and easily retrievable under time pressures from long-term memory. (I will consider this point in detail in Chapter 6.) But they also seem to indicate that the monitoring process whereby a Speaker examines, controls and even abandons a message before it is formulated may be under automatic control to a greater extent that the Leveltian model suggests.

This raises the question of what stimulus – internal or external – triggers this controlling process and, here too, my initial inquiry appears to yield some results. Phatic communion is, by its very nature, the verbal accompaniment of behaviour 'at the psychologically crucial margins of interaction' (Laver, *op cit*: 217), and any non-verbal behaviour which seems to lead to dialogue being initiated or rebuffed at these delicate moments is arguably a useful indicator of the underlying elicitatory or inhibiting mechanisms involved. My informally-gathered examples offer a preliminary indication that such consequences – speech elicitation or speech eclipse – do indeed seem to be linked to one particular aspect of non-verbal behaviour which, though realised in a variety of ways, can be summed up as Addressee attentiveness. Thus, one potential partner in a phatic exchange, whose behaviour shows awareness of the presence and of the equal awareness of the other, acts *through that very awareness* as an automatic stimulus to interaction: ultimately, to verbal interaction. Equally, a potential partner who, either deliberately or unintentionally, fails to show attentiveness to the other seems to have the power to instantly shut down a Speaker's wish to create a message, let alone formulate and articulate one. Either way, at these extremes of Addressee/Speaker relationships, it

seems to be the influence of the Addressee – exerted through low-level, non-verbal means – that is decisive.

How far have these suggestions been confirmed by formal research? As shown in Chapter 4, there is some evidence that the giving (or withholding) of eye-contact by the Addressee can have a positive (or negative) influence on the would-be Speaker, and that this influence makes itself felt at a very early age.

In the context of adult interaction, Kendon (see Chapter 4.2.1) contributes a particularly important finding on the relationship between Speaker production rhythms and eye contact with his interlocutor: namely, that gaze episodes seem to take place at 'points of uncertainty' during speech, such as at the end of phrases within utterances. He also finds that speech production is faster when the Speaker is looking at his interlocutor and therefore in a position to establish mutual gaze. In the context of research on child development, the distress caused by non-responsiveness from a communicative partner is demonstrated by the robust 'still-face effect', in which a previously responsive adult suddenly stares expressionlessly at an infant of three months or even less (an effect noted by one researcher in infants as young as 96 hours). An arguably similar result has been observed by Bavelas et al in adults, amongst whom a 'still-face effect' produced by manipulating subjects' response to a story-teller can decisively wreck the latter's narrative powers (see Section 4.2.5). Meanwhile, story-tellers in the control group, whose audience responds in a spontaneous manner, show a regular pattern of seeking eye-contact with an Addressee at key points in their story. Only when mutual gaze is established does the story-teller continue: a finding that appears to confirm Kendon's 'points of uncertainty' premise by linking this uncertainty directly with processing at a conceptual level.

Further support for the influence on a Speaker of an attentive or indifferent Addressee comes from recent work in the field of computer-generated virtual environments. Here, giving a speech to a virtual audience of grossly inattentive listeners has been shown to have a strongly affective impact on subjects, producing a state of high anxiety – however confident they were beforehand, and however aware they are at a conscious level that their audience is only a computer-generated image. Other findings from the VE field seem to indicate that video-link interactions

between real-world interlocutors, while reciprocal, may not be symmetrical: when the expressions of Interlocutor A are digitally attenuated, or 'damped', Interlocutor B becomes *more* expressive, whereupon Interlocutor A (who is unaware of the attenuation) becomes more expressive in turn – arguably, a result of a self-governing mechanism whereby the responsiveness of dialogue partners can be kept more or less balanced.

From its starting point with Levelt's 1989 model of speech production to the subtle dialogic interactions captured in recent VE research, my narrative has drawn a picture of Speaker activity much of which appears to take place below the level of consciousness. Nor is this preconscious activity focused on the production process itself. One of its most important aspects, I would suggest, is concerned with the reception and processing of the stimuli by which an alert, attentive Addressee appears to license the Speaker's efforts, or by which an Addressee's expression of abstraction or indifference inhibits speech production to the point of eclipse. And the importance of these stimuli lies, not only in the positive outcomes of such licensing, but also in the second-by-second warnings that they may give of the onset of Addressee non-responsiveness.

As I have shown, sensitivity to the latter is vitally important to a Speaker who wants to maintain his relationship with his Addressee. Furthermore, if a dialogue has been successfully initiated by this point, the Addressee – who is also by turns a Speaker – is subject to the same pressures and priorities. If they want a dialogue to continue, this sensitivity must be maintained at all costs: a point that brings me to the central claim of this thesis. Put simply, the task of the ongoing Speakers – both of them – is now to forestall in their Addressees the 'blanking' that decisively damages speech production at multiple levels. In order to achieve the goals they bring to the dialogue, they must obviously say things. But, *in order to keep saying them*, they must retain their interlocutor's interest. In other words, achieving relevance to an Addressee (and maintaining it on a second-by-second basis) is not just a desirable end-result of dialogic speech production, but an absolute prerequisite; without it, speech production will falter and communication itself will fail.

I want to suggest, therefore, that the two aspects of speech production referred to in my research questions – message content and relevance achievement – are not

merely linked by their general importance to interpersonal communication. They are instead connected in far more specific terms, by a tight – and mainly automatic – feedback loop that enables communication to function as a single unified system in which the contribution of each party, as both Speaker and Addressee, goes to increment what has gone before, and become the foundation for what will come next.

The question that now arises is: how is such a loop sustained? What resources are available to interlocutors for ensuring its maintenance? In one sense, answers are not hard to find: as indicated by the Addressee-Speaker interactions described in earlier chapters, participants in a dialogue have an enormous array of possibilities to call on in their continuing efforts to engage, assess, and retain interlocutor attention. This range is multi-modal: visual (e.g. eye contact) and auditory inputs predominate, but others – such as touch – are not excluded, and direct input from the physical environment can be complemented by conceptual information drawn from memory. It is also clear that different types of process are involved: resources available to the successful Speaker include both the deliberate, effortful audience modelling which Pickering & Garrod see as required to establish 'common ground', and preconscious processes such as the linguistic priming mechanisms that Pickering & Garrod themselves describe. Because different types of process are involved, the time taken to produce Addressee-oriented utterances also varies widely, with the quick-fire dialogue of Examples 2 or 6 at one end of the spectrum (see Chapter 3.2 and 3.4), and painstaking trawls through long-term memory and planning at the other. In contrast, however, the need to check if Addressee orientation has been successful makes itself felt extremely frequently, as demonstrated by studies of interlocutor eye-contact: for example, the narrator of the 'surefoot Charlotte' story (see Chapter 4.2.4) checks her listener's reaction every few seconds, and does not proceed until a satisfactorily attentive reaction is secured.

However, these answers lead to other, perhaps more fundamental, questions. The evidence of such second-by-second assessments of audience reaction – along with the fluency and overall speed demonstrated during most stretches of spontaneous conversation – appears to support my suggestion that the feedback loop crucial to Speaker success operates mainly, if not entirely, on a low-level, automatic basis. But, again, the question arises: how does this work? How does automaticity

contribute to everyday conversational behaviour? One valuable line of approach has already been mentioned, although in passing. This is now the point at which to consider fully the relationship between the successful Speaker and dual-process theory, which postulates, as Evans puts it (*op cit*: 454), 'two minds in one brain'.

5.3 'Two minds in one brain': the dual-process theory of cognition

Dual-process accounts of reasoning, learning, social cognition, and other aspects of cognition have emerged in their modern form over the last 40 years. Although varying considerably in their details, dual-process theories agree on one core proposition: that there may be two distinct processing mechanisms for carrying out a given task, which use different procedures and may give different (and perhaps conflicting) results.

According to Frankish & Evans in their introduction to an important collection of research in the field,

> Typically, one of the processes is characterized as fast, effortless, automatic, nonconscious, inflexible, heavily contextualized, and undemanding of working memory, and the other as slow, effortful, controlled, conscious, flexible, decontextualized, and demanding of working memory. Dual-process theories of learning and memory have also been developed, typically positing a nonconscious *implicit* [authors' emphasis] system, which is slow learning but fast access, and a conscious *explicit* one, which is fast learning but slow access.
>
> Frankish & Evans (2009: 1)

It is the idea that the Speaker may have access to two such distinctively different processes that may offer a key to understanding the mechanisms supporting the direct, low-level system of Speaker-Addressee interaction central to my own hypothesis. In the following sub-sections, I review some of the work carried out in the dual-process area, and indicate potential areas of relevance to my own inquiry.

5.3.1 Dual-process theory: terms and definitions

There seems to be general agreement that the various sectors of the dual-process field have evolved largely independently, with little connection between them: a lack of connection, Frankish & Evans comment, that is a 'reflection of modern psychology' (*ibid*: 11). One result is the wide variety of terms and definitions current in the field at large, and I will start by explaining some of these.

Dual-process theory is also known as dual-system theory, with fast, automatic, effortless, high-throughput processes being collectively referred to as 'System 1' processes, and slow, controlled, effortful, low-throughput processes as belonging to 'System 2'. The 'System 1/System 2' label is replaced in some accounts by the label 'Type 1/Type 2' to describe the same process groupings. I will here refer to the whole field as 'dual-process theory', but will also continue to use the 'System 1/System 2' label for the process classifications that it advances.

There is also disagreement on the actual number of systems involved: Evans, for example, adds a set of 'System/Type 3' processes: a meta-system that controls which of the other two should be called into play at any given moment (Evans, 2009: 48). Overall, however, the picture that emerges is that of a rapidly-functioning, intuitive, associatively-oriented 'mind', operating below the level of consciousness and alongside a volitional, slower-moving, and logically-oriented processor which – unlike its fellow – is to some extent open to conscious scrutiny and deliberate use. The following sub-sections describe this theoretical core in more detail, and then outline some of the differing developments to which it has given rise.

5.3.2 Dual-process cognition: features of the core theory

The two cognitive systems that the theory proposes are each characterised by a range of properties, some of which are outlined in Table 5.1. Again, there are several ways of classifying these. The grouping I have chosen here simply lists them in terms of their most obvious attributes: general processing style in Group 1; required level of consciousness in Group 2; typical inputs in Group 3; and evolutionary status in Group 4.

SYSTEM 1	SYSTEM 2
Group 1 ('Core')	**Group 1 ('Core')**
Fast	Slow
Automatic	Controlled, volitional, intentional
High processing capacity	Limited capacity
Low effort	High effort
Parallel	Sequential
Independent of working memory	Demanding of working memory
Group 2	**Group 2**
Preconscious, non-conscious	Conscious
Sub-personal	Personal
Group 3	**Group 3**
Associative	Rule-based
Contextualised	Abstract
Intuitive	Reflective
Pragmatic, belief-based reasoning	Logical reasoning
Implicit knowledge	Explicit knowledge
'Learn slowly, retrieve fast'	'Learn fast, retrieve slowly'
Group 4	**Group 4**
Evolutionarily old	Evolutionary recent
Shared with animals	Distinctively human

Table 5.1: Properties ascribed to the two systems proposed in dual-process theory

Although opinion is divided to some extent on the properties listed in Groups 3 and 4, agreement on those in Group 1 appears to be more or less general. The System-1

mind – often described as a set of sub-systems – is a high-speed processor, crunching huge amounts of data from varying inputs, automatically and with little effort, and with only the final output being potentially available to conscious awareness; by contrast, System 2 is seen as working slowly and laboriously, handling items of data one after the other rather than in a massively parallel way and, crucially, being under the control of its human owner. A Group 1 property that is particularly powerful in both systems is its relationship with working memory: working memory acts as the 'gatekeeper' to System 2 operations, offering them up to conscious control and slowing them down. System 1, meanwhile, functions independently of working memory and this accounts for its speed and automaticity.

Despite the variations in terminology, there is also a general consensus on the first property in Group 2: the extent to which we know what is taking place in our own minds. The activities of System 1 are opaque: we are not aware of them happening, although we can be aware of their outcomes. System 2, however, is seen as open to conscious scrutiny. We analyse our intentions and confront problems; interrogate our memory for names, dates, words, procedures; use the results to construct 'what if' solutions; balance one solution against another.

Although they operate so differently, both systems are seen as working together to promote effective cognition: both can be employed in the same context, and to further a single goal. Thus, when I look at the anagram UTSOVNCI with the aim of spelling an aristocratic title (*Times 2*, 18.10.10: 21), my System-1 mind instantly de-codes it into VISCOUNT, although I have no idea how this has happened. Meanwhile, the briefer – and apparently easier – NRABO takes me plenty of System-2 trial and error. Frankish (2009: 92) invents a similar example: 'Take long division. Imagine someone with a natural talent for arithmetic. We ask them, what is 21,582 divided by 11, and they immediately respond with the answer, '1962'. We ask them how they worked it out, and they say they don't know – the answer just came to them'. As he points out, however, few people can do long division in this way; instead, they 'get a pencil and paper, write down the numbers, then perform a sequence of simpler divisions and subtractions, dividing 21 by 11, writing the integer part of the answer above the '21' and the remainder below, and so on, in the usual style.' Finally they read out the figure that has accumulated above the line: overall,

the result of a System-2 approach, but with intermediate steps such as a simple subtraction problem being accomplished by the automatic, rapidly-calculating System 1. 'Ultimately,' Frankish concludes, 'the process breaks down into actions that are the product of subpersonal [i.e. System-1] rather than personal reasoning.'

As acknowledged by Levelt (see Chapter 2.1.1) – and also suggested by Pickering & Garrod's dialogue alignment model – similarly automatic processes contribute to speech production. But does System-1 automaticity extend right up the Levelt model to the message level: are message-level decisions – essential to establishing and maintaining relevance to an Addressee – made in the same fast, effortless etc way, independent of working memory? The examples that I give in Chapter 3 seem to show that this might be the case, and Frankish's version of dual-process cognition seems to support this possibility. . Moreover, he continues (ibid: 97), System 2 will therefore be 'heavily dependent' on System 1 in several ways:

> First, [it] will be dependent on System 1 for its *inputs* [author's emphasis]. Conscious, personal-level reason can begin only after a great deal of preconscious processing has been completed – processing that determines which problems become the focus of attention and what information is consciously recalled for use in solving them… Second – and more controversially – System 2 will be *causally* [author's emphasis] dependent on System 1. The intentional actions involved in personal reasoning will themselves be generated by subpersonal cognitive processes. These will include the processes involved in deciding to initiate personal reasoning, choosing reasoning strategies, directing attention, selecting, generating and manipulating inner speech and other sensory imagery, together with the mechanisms of language comprehension, self-monitoring, self-regulation, and many other processes, depending on the nature of the task.
>
> (*ibid*: 97)

If message-level decisions in speech production are included in this account of the System-1/System 2 relationship, it should be arguable that automaticity plays a greater part in dialogic exchange than even Pickering & Garrod envisage.

5.3.3 Parallel or sequential processes? A brief review

The question of how the two systems are related to each other – to which Frankish's is only one of the many solutions proposed – is one of the biggest issues facing researchers in the field and possibly represents the deepest divide in dual-process theory itself. It raises a number more specific questions: in what order, for instance, do the two systems operate – simultaneously or sequentially? If simultaneously, how do they collaborate? And what happens if they disagree? Strong arguments have been given both for a parallel (i.e. simultaneous) mode of system interaction and for a sequential one. Although these issues do not play a major part in my own theory, I will briefly mention some of the salient points

A parallel relationship would involve a continuous interplay between the two systems, with both competing for attention and acceptance. Sloman (1996: 15) describes experiments in the psychology of reasoning that seem to illustrate this. Here, associative (System 1) and rule-based (System 2) processes yield incompatible results, with the System 1 results persisting however much participants try to ignore them. As Sloman puts it, 'Both systems seem to try, at least some of the time, to generate a response. The rule-based system can suppress the response of the associative system in the sense that it can overrule it. The associative system however always has its opinion heard and, because of its speed and efficiency, often precedes and thus neutralizes the rule-based response.' A similar pattern, whereby System 1 makes a sustained and continuous impact on the simultaneously-operating System 2, is also proposed by Smith & DeCoster (2000: 112).

In sequential models of dual-system processing, these questions of overlap and running competition do not apply. As described by Evans (2009: 45-6), the assumption here is that a fast, automatic, System-1 process 'precedes and shapes' subsequent conscious, controlled, effortful System-2 reasoning. On this approach, competition between the two systems is resolved by the sequencing pattern used. Thus, System 1's output – a default mental model produced by heuristic[2] processing

[2] Differences over the meaning of this term represent another example of confusing terminology current in dual-process work, and perhaps reflect the lack of connection between the fields from which the theory has emerged. According to Evans, the field of social cognition draws a contrast between heuristics – speedy, low-effort cognitive functioning – and the more effortful type required for considered rule application: a distinction, Evans (2009: 36) comments, 'between a quick and dirty

– is presented for consideration to the analytic processes of System 2. Where this second-stage processing is shallow and careless, Evans explains, the final output is effectively a heuristic response. However, depending on 'motivation, cognitive ability, instructional set, *time available*, [my own emphasis] and so on,' analytic processing may lead to revision and replacement of the initial default model as a result of more effortful reasoning.

Frankish's conception of a System 2 that is heavily dependent on System 1 for input and motivating factors seems to take the sequential model as far as it can reasonably go while still preserving some connection between the two systems. However, it is only one of the major variations that have evolved within dual-process theory's developing structure. Other important ones include the role played by memory in the activities of the two systems and, by extension, in their relationship with learning. This dual model of memory is discussed in the next sub-section and, as I hope to show in the following chapter, a further extension links dual-process theory – in particular, System1 – with speech production that is automatically geared to ensuring Addressee relevance.

5.3.4 'Separate but interacting': duality in memory systems

A major contribution here is made by Smith and DeCoster who propose that System-1 and System-2 processes are linked to separate memory systems that use two 'fundamentally different' principles of operation. The first of these – a slow-learning memory system – supports the processing mode associated with System 1, and the other – a rapid-learning memory system – feeds into the operations of System 2.

---

heuristic form of processing and a slow and careful systematic form [that] could, *but need not* [author's emphasis], reflect an architectural distinction.' As he and other theorists point out, however, System 2 can also make use of heuristics, drawing on a quick and careless thinking style as well as on slow, careful, consideration. A further complication lies in the fact that, in some contexts, 'heuristics' refer to rule-based structures, or rules of thumb, that have to be retrieved from memory and evaluated, presumably consciously. An example given by Frankish (*ibid*: 94) of this deliberate application of a rule is the recognition heuristic, 'If you have to say which of two items best satisfies some criterion, then choose the one you recognise.' Within relevance theory, the relevance-theoretic comprehension procedure is seen as an automatic heuristic which operates at a sub-personal (non-conscious) level. Mercier and Sperber have argued that, within their massive modularity framework, higher-order reasoning is also carried out by intuitive argumentative mechanisms which yield inferences about premise-conclusion relations (see Chapter 5.3.5). So their differences from Evans et al. are mainly about the working of System 2. On the parallel vs sequential issue, Sperber favours a 'pandemonium' model in which most automatic procedures operate in parallel and compete for processing resources.

Smith & DeCoster (2000: 109) start from the assumption that there is a functional incompatibility between the demands made on humans by existence. On the one hand, they need to record information slowly and incrementally, thus accumulating a large sample of experiences in memory so that expectations and long-term stable knowledge can be based on average, typical environmental properties. This requires a 'slow-learning memory system which could be termed "schematic" because it matches the typical properties assumed for schema in social and cognitive theories.' At the same time, however, humans also need to learn new information rapidly, so that a novel experience can be remembered after only one encounter. This calls for a 'fast-binding' system that can store episodic records of details of specific experiences.

As evidence for this hypothesis, Smith & DeCoster point to disassociations observed in amnesia patients between different types of memory-related capability: specifically, between the capacity to learn and remember general regularities in the environment (e.g. character traits), as distinct from consciously recollecting particular events. They continue:

> A model featuring two separate but interacting memory systems can account for these sorts of evidence. The slow-learning system is a collection of overlapping systems that are involved in sensory, perceptual, and motor output processes. The operation of these systems generally does not depend on conscious awareness or attention. These systems are responsible for translating input representations (e.g. visual patterns of letters) to output representations (e.g. word meanings). They also function as memory systems. Learning takes place as the system processes each stimulus and involves small, incremental alterations of representations in ways that facilitate repetition of the same processing… These representations are then used preconsciously to process and interpret new information by categorizing, filling in unobserved details, and the like.
>
> (*ibid*: 110)

Meanwhile, the fast-learning memory system, which depends on the hippocampus and related brain structures, is responsible for rapidly constructing episodic memories – new representations, resulting from single experiences – that bring together different aspects of an experience or an object in its context.

Smith & DeCoster propose that these differences in the two memory systems directly affect the operations of dual-process theory's System 1 and System 2. The slow-learning system drives a preconscious, fast-operating 'associative processing mode' that functions as a pattern-completion mechanism: here, a stimulus calls to mind the wealth of information and affective associations that have accumulated in the past. The fast-learning system helps support a 'rule-based processing mode' which draws on symbolically and culturally transmitted knowledge and rests on human linguistic abilities. Importantly, Smith & DeCoster point out that these abilities, in turn, draw on *both* underlying memory systems.

As Table 5.2 shows, there are clear overlaps between Smith & DeCoster's proposal and core aspects of dual-process theory, notably in connection with the issues of processing speed and automaticity. These are particularly highlighted in connection with the rule-based mode, and for the same reason cited in other versions of the theory: the processing bottleneck that System 2 necessarily involves. Here, it is as much the implementation of the 'rules' as their retrieval from memory that slows things up: the process of using the retrieved explicit representations to guide processing is, Smith & DeCoster explain, 'necessarily sequential and relatively slow (in contrast to the fast, parallel constraint-satisfaction process that can be used with associative knowledge representations). The reason is that only one rule can be explicitly used to guide processing at a time. Rule-based processing is thus more effortful and time-consuming than associative processing' (*ibid*: 112).

ASSOCIATIVE PROCESSING	RULE-BASED PROCESSING
Draws on associations…	Draws on symbolically-represented rules…
… which are structured by similarity and contiguity…	…which are structured by language and logic…
… and learned over many experiences	… and can be learned in just one or a few experiences
Occurs automatically	Occurs optionally when capacity and motivation are present
Occurs preconsciously, with awareness of the *result* of processing	Occurs often with conscious awareness of processing steps

(after Smith & DeCoster, *ibid*: 111)

Table 5.2: Theoretical properties of associative and rule-based processing modes

There are two important corollaries to this equation of the associative processing mode with fast retrieval speed and the rule-based mode with the reverse. The first can be simply expressed as 'learn slow, retrieve fast' vs 'learn fast, retrieve slow(er)'. To this, Smith & DeCoster (*ibid*: 115-6) add an interactive dimension, whereby over time, the results of slow, effortful reasoning can become established in the associative system, and vice versa:

> Suppose someone repeatedly uses a step-by-step rule-based process to make an inference or solve a problem – perhaps just counting on one's fingers to get the answer to 2 + 3. Repeated trials create the conditions for associative learning, so eventually the same answer can be retrieved by pattern-completion from the associative system, rendering the step-by-step procedure superfluous… With

> enough practice, therefore, the answer to such a problem just pops into consciousness.

Equally, information can move the other way: people can, perhaps, reflect on their own past accumulated experiences and condense them into a symbolically-represented rule. For example, frequent and successful fishing trips with a friend may lead someone to realise that the friend always knows where the fish are biting: a conclusion that in turn leads to further considerations, such as *how* the friend knows. As Smith & DeCoster comment, 'Turning this knowledge from a mere association built up from repeated experiences into a symbolic representation has several benefits – the knowledge can be used flexibly, applied in other contexts, or it can be communicated to other people.'

What happens when there is a conflict between the outputs of the associative and rule-based systems of this memory-based model? Citing instances where this results from a clash between argument strength (e.g. weak) and appeal of source (e.g. attractive or expert), Smith & DeCoster suggest that the crucial factors that govern the overall response are motivation and cognitive capacity. Given that the rule-based system demands effort and attentional resources, they say, there has to be some motivation – such as a desire for accuracy or argument validity – in order to use it. If this is lacking, the response will generally be determined by the relatively effortless associative system. They also make the further crucial point (*ibid*: 117) that 'cognitive capacity' refers to *temporal* as well as attentional resources:

> Rule-based processing generally takes longer than associative processing… and, because it requires attention, it is more subject to disruption by distraction, interference, and so forth. Thus, responses that are made quickly or when the perceiver is busy or distracted likely will be controlled by the associative system. However, given adequate time and freedom from distraction, rule-based responses (because of their greater subjective validity) may override associative responses.

The second key point that Smith & DeCoster make about the link between the associative system and fast retrieval speed has to do with the huge variety of outputs it can produce. Its pattern-completion mode of operation, operating automatically and preconsciously, has such a wealth of accumulated representations to draw on that even a limited stimulus can activate a wide variety of these, and they may go well

beyond the purely informational. Thus, the sight of a mug activates the intuitive conceptual information that it is used to hold coffee, while the sight of a friend may bring an affective response of warmth and affection.

Another feature of the associative system is that it can use currently available cues to retrieve representations that were stored on past occasions when similar cues were present. 'Through associative processing, information that has repeatedly been linked to an object in the past is automatically brought to mind whenever we perceive or think about the object again. This information can fill in unobserved details or can even change the way people perceive existing features of an object' (*ibid*: 111). In Chapter 6, I consider some of the implications for Addressee-oriented speech of such automatically-retrieved links, whether to objects, situations, or individuals.

5.3.5 Dual-process theory and the massive modularity hypothesis

Dual-process theory has interesting links to the massive modularity hypothesis (Sperber 2005, Carruthers 2006), which is being explored in current work in relevance theory. In the psychology of reasoning, for instance, Mercier & Sperber draw a distinction between 'intuitive inference' and 'reflective inference' (which they call 'reasoning proper'), and discuss their relations to System-1 and System-2 reasoning: 'Some analogies seem obvious. Both system 2 and reflective inference are characterized by control, effortfulness, explicitness and, (at least virtual) domain-generality. They contrast in all these respects with system 1 and with intuitive inference' (Mercier & Sperber, 2009: 156). Again, they suggest that a 'clear asymmetry' exists between one type of inferences – System 1/intuitive, found in 'all animals endowed with rich enough cognitive systems' – and a second type – System 2/reflective, that 'may well be absent in non-human animals and that, even in humans, are used much more sparingly than the first type'.

However, Mercier & Sperber take the two-system distinction beyond a simple list of contrasting characteristics to considering the distinctive goals served by each type of inferencing process – and, especially, by the second. The main function of reflective inferencing, they argue, is *not* to improve individual cognition (as many philosophers have thought), but to contribute to interpersonal communication by enabling speakers

to produce persuasive arguments and hearers to evaluate those arguments and thus defend themselves from being accidentally or deliberately misinformed.

This approach has been developed within the theoretical context of the 'massive modularity' hypothesis, which treats the human mind as composed of many specialized, autonomous, processing modules, each with its distinct evolutionary history, and each with its own input conditions, specific procedures, and characteristic outputs. At first sight, this multi-process model seems to stand in marked contrast to the dual-system approach. A massively modular mind would perform a wide range of automatic inferences, justified by regularities existing in different cognitive domains – a pluralist rather than dualist view of cognition. There would therefore not be a single type of System-1 inferencing, but many, carried out by different domain-specific modules that would yield valid results only for their particular domain. These would include three distinct types of metarepresentational module of particular interest to research on communication: a 'mindreading' module specialised for processing representations of mental states, a 'comprehension' module specialised for processing representations of speakers' meanings, and an 'argumentative' module specialised for representing logical or evidential relations among propositions. The claim that all three types of metarepresentational module perform intuitive/System-1 inferences brings aspects of mindreading, communication and argumentation into the scope of automatic vs controlled processes.

This integration of aspects of a pluralist architecture into two broadly distinguishable types of inference process allows Mercier & Sperber to map the massively modular model onto the constructs of dual process theory. All the inferential modules, they suggest, perform their function of modifying or adding to an individual's beliefs at a System-1 level: 'The modification of the stock of beliefs …that results from spontaneous inference occurs without the individual's attending to what justifies this modification' (*ibid*: 153). However, the output of the argumentative mechanisms (which yields conclusions of the type 'P is a good enough reason for believing that Q') is an intuition about premise-conclusion relations which can be used to justify actually drawing the conclusion Q. According to Mercier & Sperber, the argumentative mechanism works at a preconscious, System-1 level, taking

> … as input a claim and, possibly, information relevant to its evaluation, and [producing] as output reasons to accept or reject that claim. The workings of this module are just as opaque as those of any other module, and its immediate outputs are just as intuitively compelling. We accept as self-evident that a given pair of accepted assumptions of the form *P-or-Q* and *not-Q* justifies accepting the conclusion *P*, but this compelling intuition would be hard to justify….The argumentation module provides us with reasons to accept conclusions, even though we may be unable to articulate why we accept these reasons as reasons.
>
> (*ibid*: 154-5)

Mercier & Sperber complete their mapping of inferencing processes onto the dual system by distinguishing between the argumentation module's intuitively-accepted output (i.e. P is a good enough reason to accept conclusion *Q*) and the process of actually accepting the conclusion Q. Such conclusions, accepted with an awareness of the mentally-represented reasons for accepting or rejecting them, are described by Mercier & Sperber as 'reflective'. As they put it, 'Conclusions accepted for a reason are not intuitive but are, we will say, "reflective"… and the mental act of accepting a reflective conclusion through an examination of the reasons one has to do so is an act of reflection… Since reflective inferences involve the representation of reasons, they well deserve the name of reasoning proper' (*ibid*: 155-6). Thus described, reflective inferences are System 2 processes, and share some of the properties ascribed to such processes in my discussions above.

Mercier & Sperber apply this intuitive/reflective distinction to specifically communicative ends by suggesting (*ibid*: 159) that the primary function of reasoning is a social one, linked to the 'massive human reliance on communicated information.' The evolutionary pressures that, amongst animals, have ensured 'honest signalling' – e.g. of worthiness as a potential mate – translate amongst humans into a strong selective pressure for filtering mechanisms that ensure reliability in communicated information, either by assessing the reliability of communicators themselves, or by assessing the reliability of communicated contents. The function of intuitive argumentation mechanisms might therefore be, they argue, to regulate the flow of information among interlocutors 'through persuasiveness on the side of the communicator and epistemic vigilance on the side of the audience' (*ibid*: 166).

5.3.6 Are language processes System-1 operations?

In a later paper, Mercier & Sperber present their own version of the Addressee-Speaker feedback loop: 'For communication to be stable, it has to benefit both senders and receivers; otherwise they would stop sending or stop receiving, putting an end to communication itself' (Mercier & Sperber, 2011: 60). Although their focus is on the influence of the Speaker rather than the Addressee, their proposal nonetheless describes one of the possible mechanisms whereby the two systems envisaged in dual-process theory may help to ensure sustained – and therefore successful – communication. However, there is a further reason why their discussion of communication in a dual-process context is of interest here. Their hypothesis appears to be one of the few recent dual-process accounts that focuses on language use specifically and in detail: a gap in the narrative that can perhaps be explained by the fact that linguists themselves have not been heavily involved in the development of dual-process accounts. This is not to say that the subject is entirely ignored in current dual-process models, but it figures more as an incidental than a dominant theme.

For instance, Carruthers (2009: 118-20) proposes a model of System-2 operation in which beliefs and goals emerge into consciousness as a result of cyclical operations of System 1, among which he includes the mental rehearsals of bodily and speech actions. In the latter case, a wide variety of System-1 subsystems are involved, including perceptual systems, bodily states, belief/desire generating systems and 'practical reason systems' as well as language production and comprehension. The output is 'inner speech', a set of (mainly auditory) representations of the form the rehearsed speech item would take which, in a process reminiscent of that proposed by Levelt, the language comprehension sub-system can then process. Language also plays a key role in supporting the tripartite model of the mind proposed by Stanovich, in which the architecture of human rationality (i.e. System 2) is restructured into two levels, 'algorithmic' (calling on fluid intelligence) and 'reflective' (calling on goals/ epistemic values). A crucial link between the two – cognitive decoupling, or the preparation of representations for use in 'what-if', hypothetical reasoning – is performed at the algorithmic level, and draws heavily on

language for its functioning. 'Decoupling…' Stanovich comments (2009: 63), 'is a cognitively demanding operation. Any mindware that can aid this computationally expensive process is thus immensely useful, and language appears to be one such mental tool. Language provides the discrete representational medium that greatly enables hypotheticality to flourish as a culturally acquired mode of thought. For example, hypothetical thought involves representing assumptions, and linguistic forms such as conditionals provide a medium for such representations.'

Although Stanovich is focussing here on the use of language as a medium for System-2 thinking, he had earlier included it, along with other domain-specific processes such as perception, in the range of automatic systems that make up System 1 (Stanovich, 2004). Here he is not alone. There seems to be agreement on the role of language as a mediator/enabler for cognitive processing at the conscious (System-2) level: as Frankish suggests, 'the conscious mind [is] a language-dependent virtual machine' (Frankish & Evans, *op cit*: 22). But it is also clear that some language processes themselves are System-1 operations, working according to the standard System-1 pattern of processing: fast, effortless, automatic, undemanding of working memory, with only the final output accessible to conscious awareness and manipulation. This is the clear implication of Frankish's explanation of the distinction between personal reasoning (i.e. System-2 mental processing) and the processes underpinning it:

> Although personal reasoning itself is conscious, the beliefs and desires that motivate it typically will not be. Actions can be consciously performed even if we do not consciously reflect on the reasons for performing them… Most of our behaviour is generated without the involvement of personal reasoning. Think about the actions involved in such everyday activities as driving a car, *holding a conversation* [my emphasis], or playing sports. These are intelligent actions, which are responsive to our beliefs and desires (think of how beliefs about the rules of the game shape the actions of a football player), and a great deal of complex mental processing must be involved in generating them. Yet, typically, they are performed spontaneously with no prior conscious thought or mental effort. Indeed, giving conscious thought to such activities is a good way to disrupt their fluidity.
>
> (*op cit*: 93)

Something like automaticity of language output also seems to be implied in Smith & DeCoster's discussion of time constraints on cognitive processing (see Chapter 5.3.4): 'Responses that are made quickly or when the perceiver is busy or distracted likely will be controlled by the associative system…' Evans, too, allocates to Type/System 1 the functioning of the 'perceptual, *language* [my emphasis] and memory systems that provide content for working memory form' (*op cit*: 48), although he acknowledges that this can present problems where language itself is concerned[3].

Effortlessness; speed; spontaneity … These are features of naturally-occurring speech that we have met before. They characterise the speech on 'something close to autopilot', illustrated in Chapter 3 by the examples of phatic communion. There are, for instance, few signs of careful reflection in the garden-lovers' dialogue (Chapter 3.2, Example 3) and none in my exchange with the weary passer-by (Chapter 3.4, example 7). And this effortlessness is not entirely the result of the automaticity ascribed by Levelt to the lower-level processes in his model; there is a message-level element in it as well. Responding to the elicitatory eye-contact made by their (future) Addressees, the interactants in these examples *decide* to talk; they therefore have to *decide* what to say. And these decisions are made against a background of many other demands on their cognitive resources, notable among which is the simple need to keep progressing through a public space.

Taken together, these circumstances seem to indicate that the lower speech production levels are not the only candidates for System-1 status in the dual-process framework; the message level is also a candidate, and not just because it can draw on stock phrases easily retrievable from long-term memory[4]. Both the intention to speak and the selection of appropriate content appear to be at least partially supported by System-1 processes as well: high-speed, automatic processes that can handle large amounts of data in parallel and, via a 'learn slow, retrieve fast' memory system, can respond to even a limited stimulus by recalling a mass of associated

[3] Dual-system theorists, he comments, generally address the issue of modularity by siting modules within System 1 due to their manner of functioning: rapid, unconscious, automatic. 'But this now seems problematic to me. Language appears to be modular in the strong sense but is (a) uniquely human [therefore a System-2 candidate] and (b) is an essential prerequisite for type 2 processing, facilitating higher order and explicit representations of knowledge' (*ibid*: 40).

[4] Although these certainly play a part in supporting 'effortless' verbal interaction; see Chapter 6 for a discussion of speech routines and routinisation.

material. And if, for casual speech at least, the Conceptualizer operates according to System-1 principles, do we here have the essential link in the Addressee-Speaker feedback loop which enables a Speaker to retain Addressee interest by producing relevant utterances – all the while complying with the heavy time pressures of normal dialogue?

At the moment, direct evidence for such a link from dual-process work seems thin on the ground. However, as I hope to show in Chapter 6, one can approach a dual-process account of speech production from other perspectives and, in the meantime, research in the field of neuroscience offers valuable insights into the crucial mechanisms supporting dialogue: those underpinning social interaction itself. These developments are outlined in the following section.

5.4 Dual process, personal interaction, and the evidence from neuroscience

A central assumption of my thesis is that the interaction between Speaker and Addressee, while indisputably the outcome of mental and emotional activities, cannot be usefully investigated without taking into account the mechanisms that support these activities at the most fundamental level: that of the brain. As Posner & Raichle commented almost 20 years ago, 'The cognitive studies of the last 30 years have furnished a basis for decomposing complex [cognitive] behaviour. These studies have provided hypothetical but well-reasoned sequences of mental operations. Although we cannot be sure the sequences are real, we can test their reality in the process of locating the corresponding brain areas' (Posner & Raichle, 1994: 16). The reasoning behind this suggestion has not changed, while the technological means of implementing it are increasing dramatically.

Admittedly, these advances have been accompanied by frequent warnings as to the precision and reliability of the results obtained. For example, Nishitani et al (2005: 60) point out that, 'activation of any area in a brain imaging study does not mean that the neural substrate of the mentioned functions is seated (only) there; rather, it indicates that the activated area is involved in, or *may be* [my emphasis] an important

node in, a widely distributed neuronal network.' Nor does the problem of imprecision apply only to issues of brain topography. Cabeza & Nyberg comment (2000: 35): 'In the context of regional activations observed across cognitive domains, it must… be noted that activation of one and the same region in two distinct domains need not imply that the region has the same functional role in both cases.' Instead, it has been argued that the functional role of a brain region depends, at least in part, on its neural context: that is, the 'pattern of interactions among brain regions.' From this perspective, a particular brain region should not be seen as dedicated to given specific functions, but may play a role in a variety of cognitive and other operations.

However, the wealth of findings that is building up in the field of cognitive neuroscience provides, at the very least, invaluable insight into potential relationships between the brain and the cognitive and behavioural processes that it supports. Moreover, the stronger the evidence of neural activity connected to one particular aspect of social interaction or another, the more confidently we can start to identify and characterise that aspect itself. In considering the contribution made by the dual process framework to speech production I will therefore now look briefly at the neural systems identified by research in social cognitive neuroscience[5] that could support the Addressee-Speaker interactivity essential to successful communication.

Specific areas covered in the remainder of this chapter include neural activities directly connected to the feedback loop described above (i.e. eye contact between interactants and its absence), and the possible contribution of mirror neuron theory to social interactivity. I also return to a subject that I have not considered in any detail so far: theory of mind (ToM), or the ability to attribute mental states (e.g. beliefs, desires and intentions) to self and others in order to explain and/or predict behaviour.

.

5.4.1 Carving cognitive activity 'at the joints': the contribution of neuroimaging

As noted above, the use of neuroimaging techniques in the field of social cognitive neuroscience have made a major contribution to cognitive theory. One example,

[5] Itself a rapidly-expanding field in its own right, with Internet hits totalling less than a hundred in early 2001, and 215,000 in late 2010

which gives a particularly clear illustration of how a general area of cognitive activity can be 'carved at the joints' to show subtle processing distinctions, is provided by Lieberman et al (2004) in a study of self-referencing processes. Lieberman, whose later work includes the intriguing suggestion that the processes associated with System 1 and System 2 are supported by identifiably distinct neural areas, points out (2007: 267) that, by their very nature, some social psychological processes appear exclusively either automatic or controlled. Self-reflection – deliberately, effortfully thinking about oneself – seems to be an example of the latter. However, Lieberman et al show a distinction, reflected in the neural regions activated, between self-representation in abstract terms and self-representation as an outcome of personal experience: e.g. in the form of self-schemas such as 'I am a footballer', with all the history and connotations that this implies.

Lieberman et al's study asks football players and actors to make self-descriptiveness judgments about words relating to either their high-experience domain (e.g. acting, with words including 'comedian' and 'creative') or their low-experience one (e.g. football playing, with words including 'team player' and 'stamina'), and their responses to these verbal representations show suggestive differences in both speed and neural activation. Actors respond more quickly to acting words than to athletic ones (1,168ms vs 1,308ms), with a similar response to athletic words shown by the athletes, while retrieval of non-schematic self-knowledge – as evidenced by their responses to words that did not describe them - is relatively slow. This contrast seems to reflect the distinction between the two types of memory associated with System 1/System 2, and therefore to categorize the retrieval of knowledge related/not related to subjects' self-schemas as being associated with the activation of automatic and voluntary processes respectively.

Additionally, Lieberman et al find (2004: 428) that, 'when participants judged the self-descriptiveness of trait words, different neural structures were recruited depending on how much experience the participants had in the relevant activity domain.' Thus, retrieval of non-schematic self-knowledge is shown to be associated with activity in the dorsomedial prefrontal cortex and medial temporal lobe, whereas automatically accessible schematic self-knowledge is associated with activity in the

ventromedial prefrontal cortex, amygdala, basal ganglia, lateral temporal cortex, and medial parietal cortex.

5.4.2 The automaticity of social rejection

As indicated by the above example of neural differentiation between types of recall from memory**,** neuroscientific investigation of cognitive processes can offer promising insights into the mechanisms, automatic or otherwise, underlying social cognition. Another example – and one that is highly relevant here – is provided by neuroimaging work on social exclusion.

Introducing a study on the subject, Eisenberger et al (2003: 291) point out that the maintenance of social bonds is a crucial – because highly adaptive – requirement for mammalian species, the young of whom must stay close to their caregivers in order to survive. For this reason, they suggest, the neural processes involved may have 'piggybacked' onto the physical pain system and, in an fMRI investigation of this hypothesis, they investigate causal links between social pain and a neural region strongly implicated in signalling the affective distress arising from physical pain, the dorsal anterior cingulate cortex (dACC).

In the first experimental element of this study, subjects in the scanner watch a virtual ball-catching game that they believe is being played by two other human 'participants'; but, in a situation analogous to footballers sitting on the reservists' bench - they are prevented from taking part themselves because of alleged technical problems. In the second, the subjects do take part but soon find themselves pushed out of the game: after seven throws, their partners refuse to pass them the ball, and continue this exclusion until the game ends. (In fact, the partners are as virtual as the game itself, being the product of a preset computer program.) In a post-scan questionnaire, subjects then describe how they felt about being excluded.

The fMRI results for the second episode show activation of the dACC, and this correlates with the social distress reported by the subjects afterwards. Additionally, ACC activity is also produced during the – perhaps less painful – episode in which subjects are prevented from joining in by external circumstances. These findings support Eisenberger et al's hypothesis:

> A pattern of activations very similar to those found in studies of physical pain emerged during social exclusion… Activity in the dorsal ACC, previously linked to the experience of pain distress, was associated with increased distress after social exclusion…. The neural correlates of social pain were also activated by the mere visual appearance of exclusion in the absence of actual exclusion.
>
> (*ibid*: 291-2)

Interestingly, dACC activity is negatively correlated with activation in a neural area previously associated with the regulation of physical pain distress: the right ventral prefrontal cortex [6]. Activity here is in turn negatively correlated with self-reported social distress. However, this only applies to the study in which the subjects 'knew' they were being personally excluded. 'Explicit awareness of exclusion', Eisenberger et al continue, 'may be needed before individuals can make appropriate attributions and regulate the associated distress' – an activity requiring conscious decision-making.

Overall, the neural activity reported here and in other studies on the subject strongly suggest that affective distress over losing social connections is both as 'painful' as physical pain and – importantly – is an automatic, System-1 response to a stimulus: fast, effortless, independent of general intelligence and cognitive load. In addition, as Eisenberger et al point out, this system is sensitive to signals that can range from the merely circumstantial – as in the first experimental episode – to evidence of highly dysfunctional social interaction: the (apparently) deliberate and sustained exclusion of the subject by conspecifics. It seems reasonable to suggest that conversational 'blanking', of the type I have described in Chapter 3 and elsewhere, has a definite place within that range, and that it is the affective and social distress automatically induced by an occluded gaze (whether deliberate or accidental) or the fixed stare of non-responsiveness that, as demonstrated by Bavelas (see Chapter 4.2.5) interferes so drastically with speech production.

If this is the case, additions can perhaps be made to Levelt's model of speech production. As noted in Chapter 2.1.1, one aim of my own inquiry has been, if possible, to identify some of the processes involved in the Leveltian message level

[6] Labelled 'right VLPFC [ventrolateral prefrontal cortex]' by one of Eisenberger's co-authors, Lieberman (2007: 274),

and, in particular, in its intra-Conceptualizer loop. In connection with this, I later raised the question: why do the speed and impact of the 'blanking' experience not appear to tally with Levelt's characterisation of message construction and monitoring as *controlled* processing, with self-corrections 'hardly ever' being made without some degree of awareness (see Chapter 3.2). If the distress inspired by social exclusion is, as its neural correlates may suggest, an automatic process, then here we have evidence that automaticity plays a greater part in the workings of the Conceptualiser than Levelt proposes.

5.4.3 Extracting meaning from faces: neural foundations of sensitivity to facial expressions

If the social pain caused by an averted or unresponsive gaze can be shown to have neural foundations, what about the reverse: the obviously compelling nature of successfully-accomplished eye-contact? And, given the general responsiveness to facial expression that this demonstrates, what are the neural mechanisms involved?

Here, as in the case of the 'piggyback' that social separation has performed onto the mammalian pain system, an adaptive explanation has been suggested. Surveying work in the field such as that of Perrett & Mistlin (1990), Baron-Cohen draws attention to their finding that a specific cell grouping exists in the monkey superior temporal sulcus that specifically responds to the gaze direction of another animal. These cells are, therefore, responsive to the 'state of attention of the other individual', and have the primary function of detecting whether the other individual is 'looking at me'. As he points out, the evolutionary benefits of such hard-wiring are obvious: 'It is clearly highly adaptive to become aware than another organism has you within its sights' (Baron-Cohen, 1995:90).

Work described by Puce & Perrett (2003/4; 12) shows that similar sensitivity exists in the human posterior superior temporal sulcus, together with an ability to make fine distinctions between types of facial movements. For example, seeing a mouth opening produces a stronger level of response than seeing one closing; equally, a different level of response is produced when observing eyes averting their gaze from the observer, compared to eyes focusing their gaze on the observer. 'Augmented neural responses to eye aversion movements,' they conclude, 'may be a powerful

signal that the observer is no longer the focus of another's attention.' They also suggest that human brains are very sensitive to distinctions created by facial expressions accompanying verbal or non-verbal communication in differing affective contexts, and the importance of these gesture-affect blends is also noted by Baron-Cohen:

> It would seem that whenever [the Eye-Direction Detector[7]] detects a pair of eyes that are in mutual contact with its own, this triggers physiological arousal with pleasurable consequences. There is clear evidence of physiological arousal produced by mutual eye contact. For example, galvanic skin responses increase with mutual eye contact… and brain-stem activity has been reported in response to eye stimuli in monkeys…These measures of arousal might, of course, be indicators of positive or negative emotion. However, in the case of human infants the evidence suggests positive emotion, since eye contact reliably triggers smiling.
>
> (*op cit*: 42)

Further evidence of connections between eye contact and pleasurable emotion is provided by Schilbach et al, whose fMRI study of joint attention initiatives between subjects and (virtual) partners shows important contrasts between the neural activity involved in inviting a partner to gaze at an object and the activity involved in responding to such an invitation. The latter – looking at an object gazed at by the 'partner'[8] – activates the anterior portion of the medial prefrontal cortex, whereas the former – using eye gaze to direct the partner's attention at the object – activates the anterior ventral striatum: a neural area connected with reward processing. Findings from an additional behavioural study are consistent with this result: according to answers to post-experiment questionnaires, subjects find initiating joint attention 'significantly more pleasant' than responding to others' initiative. Schilbach et al conclude (2010: 2713) that joint attention engages mechanisms that may contribute to an intrinsic motivation to engage in the interpersonal coordination of perspectives, and suggest that 'this could be closely related to the phenomenon's impact on human cognitive development by

---

[7] Or EDD: one component of the mind-reading model in the neuronormal, whose impairment, Baron-Cohen suggests, may help to account for some aspects of autism. Other components are the Intentionality Detector (ID), the Shared-Attention Mechanism (SAM) that handles triadic representations and, finally, the Theory-of-Mind Mechanism (ToMM), a 'system for inferring the full range of mental states from behavior' (*ibid:* 51).

[8] Who is viewed by the participant on a screen; participants are informed that these images are controlled by real persons outside the scanner, whereas they are actually computer animations.

contributing to the uniquely human motivation to engage in shared, social realities.'

A further example of neural responsiveness to eye-contact and affective reaction that he describes traces a connection between two separate brain areas associated with automatic activity. This is between the human superior temporal sulcus (STS) – one suggested site of EDD cells – and the amygdala, the area of the limbic system which generates the states of mind required for brute survival: flight, fight, and 'excessively friendly behaviour', or appeasement (Carter, 2000:142). Baron-Cohen cites evidence that the amygdala itself contains both face-sensitive and eye-direction-sensitive cells, in addition to those responding to facial expressions of emotion: a finding that, with other evidence, he uses to support the claim that the EDD function is located in two different nodes within a circuit that connects the amygdala and the STS region.

In a major review of the neural bases of social cognition, Adolphs (1999: 469) also emphasises the part played by the amygdala in enabling human interactivity to take place. Studies in humans and other primates, he writes, have pointed to several structures that play a key role in guiding social behaviors: among others, the amygdala, the right somatosensory-related cortex, and the ventromedial frontal cortices. 'These structures appear to mediate between perceptual representations of socially relevant stimuli, such as the sight of conspecifics, and retrieval of knowledge (or elicitation of behaviors) that such stimuli can trigger.' In a typical, emotionally salient real-life situation, he comments, all three will operate in parallel:

> The amygdala will provide a quick and automatic bias with respect to those aspects of the response that pertain to evaluating the potentially threatening nature of the situation, or with respect to allocating processing resources to those stimuli that are potentially important but ambiguous; ventromedial frontal cortex will associate elements of the situation with elements of previously encountered situations, and trigger a re-enactment of the corresponding emotional state; and right somatosensory-related cortices will be called upon to the extent that a detailed, comprehensive representation of the body state associated with emotional or social behavior needs to be made available.
>
> (*ibid*: 477)

In a later paper (2001: 235-6), Adolphs acknowledges the intricacy of these relationships – 'The sequence of events leading from perception of a socially relevant stimulus to the elicitation of a social behavior is complex and involves multiple interacting structures' – and suggests three possible patterns of interaction. In the first, the structures involved in social cognition may directly modulate cognition**.** In the second, they may modulate emotional state, which then modulates cognition indirectly, while in the third they may 'directly modulate perceptual processing via feedback' – the initial input to which may be completely outside the scope of conscious awareness, as evidenced by the finding that the subliminally presented facial expressions can cause amygdala activation.

A model of facial perception that resolves some of these intricacies has been proposed by Haxby et al, who stress the neural distinctions between processing invariant aspects of faces (i.e. those that determine identity) as opposed to changeable ones (i.e. eye gaze, expression, lip movement). The first type of representation, they suggest, is handled in particular by the fusiform gyrus, while the second is handled by the pSTS. These two regions, together with the neural area responsible for early perception of facial features, form the core of their model, which is then extended to explain the complex functioning of facial perception overall. This is accomplished, Haxby et al suggest (2000: 228) via the participation of other neural systems:

> Face perception provides information that is used to access knowledge about another person, to infer his or her mood, level of interest and intentions; to direct one's own attention to objects and events that others are looking at; and to facilitate verbal communication. The results of functional brain imaging suggest which brain regions are recruited to process some of these kinds of information. These brain regions are part of neural systems that perform other cognitive functions... However, they become part of the face perception system when they act in concert with [visual brain areas] to extract meaning from faces…

Thus extended, Haxby et al's distributed model of face perception places a strong emphasis on the superior temporal sulcus, as it is responsible for liaising with the different brain regions that handle spatial information and that are needed to process gaze direction, speech perception, comprehension (see Note 14, below), and facial

expressions of emotion, in which the amygdala plays a part. In this way, data from one cognitive system can inform the perceptions of a second: in the case of face perception, for instance, 'information about the emotional tone of an expression appears to facilitate the accurate recognition of expression.' And these inter-system relationships do not stop there: the regions called upon by the core system can, in their turn, also participate in other functions by interacting with other systems. 'For example,' Haxby et al continue (*ibid*: 231), 'intraparietal regions that act in concert with the superior temporal sulcus to mediate shifts of spatial attention in response to received gaze are also involved in directing spatial attention in response to other visual cues and, perhaps, to auditory, somatosensory, and endogenous cues, as well.'

Although the review I have given here of neuroscientific research into social interaction has been brief, it was designed to draw attention both to the extent of work in this field and to the possible extent to which brain regions associated with automatic (i.e. System 1) function are involved in such interaction. In the following two sections, however, I move away from the dual-process framework to look briefly at two other areas of research that may offer major insights into the mechanisms supporting automatic interactivity: mirror neuron theory and theory of mind (ToM).

5.5 Alternative mechanisms for interaction: mirror neurons

Although dual-process theory and its variants are of obvious relevance to the issues of interactivity and automaticity that I am exploring here, there are others that also appear highly promising. One that has received a great deal of attention is mirror neuron activity, and any account of the automaticity of social interaction would be incomplete without some discussion of the topic. First identified in macaque monkeys, mirror neuron cells are activated both when an individual makes an object-directed action (e.g. grasping a peanut) and when it observes another individual make a similar action. From the mid-1990s onwards, evidence has been found suggesting that comparable mirror cells exist in humans: certain areas of the human brain (one of them Broca's) are activated during an imitation task, thus confirming the hypothesis of a 'direct matching' mechanism that maps a visual or kinematic description of the action observed onto an internal motor representation of the same

action (Iacoboni et al, 1999:2526). The brain regions specifically involved are the lateral prefrontal cortex (opercular – i.e. Broca's – region) and the right lateral parietal cortex.

As Singer et al (2003/4: xv) comment, the obvious relevance of the discovery of this automatic matching system to social interaction has aroused great interest: 'In particular, such neurons provide a neural mechanism that may be a critical component of imitation and our ability to represent the goals and intentions of others.' Nishitani et al (*op cit*: 62) expand the argument: 'Communication, both verbal and nonverbal, requires that the interacting individuals "stay tuned." Because the conspecifics certainly are very similar in their main characteristics, it is then also mandatory that each subject's action and perception rely on closely linked neuronal circuitries – one individual's output is the other (similar) individual's input.' So how far has mirror neuron theory been shown to support these links?

Areas of inquiry that have been approached from this perspective include empathy, intersubjectivity, the evolution of language, mentalizing, and conditions in which the capacity for empathising and mentalizing appear to be impaired (e.g. autism). In the view of Gallese (2003/4: 174), for instance, 'preliminary evidence' suggests that the neural structures active during the experience of sensations and emotions are also active when the same sensations and emotions are to be detected in others. Findings cited in his review include an example of pain-related neurons responding both to pain in a subject's hand and to the sight of pinpricks being applied to the hands of another. The opposite effect has also been observed: a brain-damaged patient with an impaired ability to experience disgust also had difficulty in detecting disgust expressed by other people. 'These results,' Gallese comments, 'seem to suggest that once the capacity to *experience* and *express* a given emotion is lost, the same emotion cannot be easily *represented* and *detected* in others [author's emphases].'

Again within the context of empathy, Shamay-Tsoory et al (2009:624) have suggested there are behavioural and neural distinctions between two types of empathic response: a basic system of emotional contagion supported by a brain region (inferior frontal gyrus cortex/IFG) linked to MN activity, and a more advanced cognitive one involving insights into another person's perspective

(supported by the ventromedial prefrontal cortex). Using experimental methods that include Baron-Cohen et al's emotion-recognition test ('Reading the Mind in the Eyes'; 2001), they report a dissociation between the two empathic systems, with IFG-damaged patients showing extremely impaired emotional empathy and emotion recognition. While they acknowledge that doubt has been cast on the role played by mirror neurons in social cognition, they conclude: 'These results present an empirical evidence that the MNS is essential for emotional empathy.'

Additionally, some of the work on connections between autism and the mirror neuron system appears to show links between mirror neuron activity and the recognition of emotional states in others. For example, Dapretto et al (2005) have found significant differences between mirror neuron activity in normally-developing and in high-functioning autistic children who are asked to either imitate or simply observe facial expressions showing anger, fear, happiness, neutrality, or sadness. Their findings suggest that a normally-functioning mirror neuron system may be the mechanism supporting this part of the reciprocation process in the neuronormal:

> In line with previous findings in normal adults, the fact that typically developing children showed increased MNS activity even when simply observing an emotional expression … indicates that this mirroring mechanism may underlie the remarkable ability to read others' emotional states from a mere glance at their faces. The lack of MNS activity during both the imitation and the observation of emotional expressions in our sample of children with ASD [autism spectrum disorders] provides strong support for the hypothesis that early dysfunction in the mirror neuron system may be at the core of the social deficits observed in autism[9].
>
> (Dapretto et al, 2005: 3)

In connection with language use itself, the involvement of Broca's area in the 'direct matching' process is particularly intriguing and, indeed, attracted attention very early. In the view of Rizzolatti – one of Iacoboni's co-authors in 1999 – it is the observation/execution matching system that has provided an evolutionary bridge

[9] A description (Sacks, 1994, quoted by Baron-Cohen, 1996: 140) of the highly achieving autistic individual, Temple Grandin, gives some insight into how this lack may feel when experienced subjectively: her youth, Sacks writes, was marked by a feeling that 'something was going on between the other kids, something swift, subtle, constantly changing – an exchange of meanings, a negotiation, a swiftness of understanding so remarkable that sometimes she wondered if they were all telepathic.' Although the work of Dapretto et al and others suggests a connection between mirror neuron activity and such effortless social adjustment, the impact of a disturbed mirror system on imitation itself has been challenged (cf Leighton et al, 2008); see also Catmur et al, 2007 (below).

from action to communication: 'Our proposal is that the development of the human lateral speech circuit is a consequence of the fact that the precursor of Broca's area was endowed, before speech appearance, with a mechanism for recognizing actions made by others. This mechanism was the neural prerequisite for the development of inter-individual communication and finally of speech' (Rizzolatti & Arbib, 1998: 190). And, in their 1999 paper, Iacoboni et al suggest four distinct explanations for Broca involvement:

> There are several reasons to expect that, if a direct mapping for manual imitation does exist, it should involve Broca's area (area 44). First, area 44 is one of the relatively few cortical areas where distal movements (the type of movements imitated in this experiment) are represented in humans… Second, area 44 is considered the human homolog of monkey area F5… in which an action observation-execution matching system exists. Third, Broca's area is the motor area for speech, and learning by imitation plays a crucial role in language acquisition. Fourth…, language perception should be based on a direct matching between linguistic material and the motor actions responsible for their production. Broca's area is the most likely place where this matching mechanism might occur.
>
> (*op cit*: 4)

The role played by Broca's area in social interaction will be examined in more detail below (see Chapter 5.7).

More recent developments in the mirror neuron field have been seen as supporting the hypothesis that mirror neurons are involved, not just in action recognition and imitation, but also in understanding the intentions of others, as expressed in their actions. Tests of brain activity under different circumstances (e.g. viewing actions within, or separated from, a specific context, such as drinking tea) appear to indicate that, in addition to the direct-match mirror neurons that fire during the execution and observation of the same motor act, there are also neurons that are visually triggered by one motor act (e.g. grasping a cup), but that discharge during the execution 'not of the same motor act, but of another act, functionally related to the observed act (e.g. bringing to the mouth)' (Iacoboni et al, 2005: 533). Indeed, this study continues, the existence of such neurons had been reported from the very early work on monkeys, but their role at that time remained unclear. The new findings, on the other hand, appear to attribute a function to these other, 'logically related', mirror neurons, and also imply that they may be part of a chain of neurons coding the intentions of other people's actions - the motor acts that are most likely to follow the act observed: 'To ascribe an intention is to infer a

forthcoming new goal, and this is an operation that the motor system does automatically.' (*ibid*)

However, it has also been argued that such an interpretation is pushing the explanatory powers of the MN system too far. Jacob, for instance, points out that Iacoboni et al's results are consistent with an alternative sequence of events, in which MN activity *follows on*, rather than leads to, an observer's representation of the goals behind another's actions. In its place, he endorses an 'inverse model' of internal action whereby, while MNs in an agent's brain compute the motor commands required for achieving a prior intention, MNs in an observer's brain will compute a representation of the same motor commands, based on a representation – gained via perceptual cues – of this prior intention. The MNs' task is thus to 'predictively compute the best motor command suitable to satisfy the agent's intention' (Jacob, 2008: 190).

Others have also expressed doubts: for example, Brass et al (2007: 2120) suggest that the mirror network seems to be involved in action understanding only in situations in which no active inferential processing is required to identify the goal of the behaviour observed. And Hickok, while he acknowledges that the MN system's hypothetical role in action understanding is an 'interesting and prima facie reasonable idea', concludes his detailed analysis of the issue by pointing out that, despite the proposal's widespread acceptance, it 'has never been adequately tested in monkeys, and in humans there is strong empirical evidence, in the form of physiological and neuropsychological (double) dissociations, against the claim' (Hickok, 2009: 1240)

Nor is this the only type of challenge that has been offered to the MN theory constructed over the years by Iacoboni and his colleagues. Another takes a different approach, by extending both the physical area possibly involved in 'mirroring', and the functionality involved. In an account of work on action rehearsal in monkeys, Cisek & Kalaska (2004) point out that mental rehearsal of actions activates motor circuits both when an action is directly observed and when information about it is given to subjects in advance. But mirror neurons – cells sited in the ventral premotor cortex – do not appear to support this mechanism: they only respond to natural

actions directly viewed, and do not appear to be involved in the prediction of future actions or events based on arbitrary information

Cisek & Kalaska's study, however, demonstrates that another group of cells, sited in the dorsal premotor cortex (PMd), may be able to contribute to such mental rehearsal of actions. Unlike mirror neurons, they say, activity here does predict the directionality of impending movement before it starts; furthermore, and importantly, the cells respond to abstract visual cues that become associated with the action via training rather than in the natural course of events. After training in their experimental task, the monkeys show PMd activation both during performance and observation of actions, even before the action to be observed – a cursor movement on a computer screen – has begun.

Although, as Cisek & Kalaska acknowledge (2004: 996), this PMd activity differs in several key ways from MN activity in the ventral premotor cortex, it is also possible that they are 'functionally related': Both findings show that observation of external events can engage nominally 'motor' circuits in order to generate motor representations of the actions associated with those acts. They add that such covert simulation of motor acts may contribute to the assessment and understanding of observed events but warn that this mental rehearsal process is 'still poorly understood'.

Cisek & Kalaska's work has recently been extended by Catmur et al, who have demonstrated that, as in monkeys, both the human ventral and dorsal premotor cortices are involved in mirroring. Additionally, both sites respond similarly to counter-mirror training (see below): an important effect that reinforces the theory that, rather than being totally innate, the properties of mirror neurons can be acquired through associative learning (ASL). According to this theory, Catmur et al suggest, any motor areas with appropriate connections to sensory areas have the potential to show mirror effects, given sufficient mirror experience - a major extension to initial models of MN activity. They continue (2011: 2360): 'Sources of mirror experience, in which observation and performance of the same action occur in a contingent manner, include observing one's own actions, being imitated… and engaging in synchronous actions with others... ASL therefore provides an explanation for the

presence of multiple brain areas responding to both observation and performance of actions in brain imaging studies.'

The key relationship between MN functioning and training/learning has been the focus of earlier work by Catmur et al (2007), who use incompatible sensorimotor training – in which human participants perform movements with the index finger while observing little-finger movements, and vice versa – to investigate ways in which mirror neurons acquire their mirror properties. After training, participants in whom this counter-system has been induced show a reversal of the normal mirror effect: a muscle-specific mirror effect related to the little finger is detected when the participant observes index-finger movements, and vice versa. These results, Catmur et al say, provide 'strong support' for the theory that the mirror properties of the mirror system are not completely innate, nor – once acquired – do they remain fixed; instead, they appear to develop through the continuing sensorimotor experience of performing and observing actions. Furthermore, the stimuli involved need not be innately specified; they can be arbitrary ones (e.g. the sound of ripping paper) resulting from repeated sensorimotor pairings. Thus, they continue (2007: 1529), the mirror properties of the mirror system are 'genuine but not intrinsic – they depend on the experienced contingency, rather than the objective similarity, between stimuli and responses.'

Overall, Catmur et al point out, these findings indicate that, to some extent, the human mirror system is both a *product* of social interaction (an important source of sensorimotor experience) and also a social interaction *process*. Through involvement in language acquisition and Theory of Mind, it both contributes to human capacity for complex social interaction and depends on the social availability of appropriate sensorimotor experience: a finding that is directly relevant to the relationship between MN functionality and autism, as explored – for example – by Dapretto et al. However, the conclusions reached by Catmur et al are in opposition to Dapretto et al's claim that that early MN dysfunction may be 'at the core 'of the social deficits observed in autism. Catmur et al, in contrast, claim that problems experienced by autistic individuals in orienting to social stimuli may be a cause, rather than a result, of reduced MN activity: 'Impairments in attention to social stimuli are likely to limit the opportunity to learn sensorimotor links between observed and executed actions,

and therefore recent reports of reduced mirror-system activation in autism spectrum disorders… may be explained by the sensorimotor-learning hypothesis' (*ibid:* 1529).

An interesting contrast to accounts that link motor/embodied cognition theories to MN-driven recognition of emotion is given by Neal & Chartrand, who point out (2011:1) that, while understanding of the processes of facial mimicry and facial feedback to the brain is well-established, 'it remains unclear whether people can use facial feedback to make more accurate judgments about other people's expressions'

Using methods that reduce or amplify muscular facial feedback (respectively, Botox and a resistance-creating gel), they address the problem by exposing participants to Baron-Cohen et al's 'Reading the Mind in the Eyes Test' (RMET) used by Shamay-Tsoori et al, who show that impaired performance here is linked to damage to the IFG, and thus to the mirror neuron system. Impaired RMET performance has also been linked with autism (a condition where mimicry is reduced). None of the participants in Neal & Chartrand's study falls into either IFG-damaged or autistic categories; here, however, an impairment effect is produced by the novel manipulation of facial mobility, with emotion recognition being significantly less accurate, compared to a control group, amongst participants whose own facial responses have been paralysed with Botox. (In a separate experiment, participants' judgment of others' facial emotion is shown to improve significantly if their own facial responsivity is increased, via application of the restricting gel.) 'Thus,' Neal & Chartrand comment, 'reducing facial feedback appears to have broad functional effects on emotional processing, encompassing both emotional reactivity and emotion perception' (*ibid*: 5). This seems to indicate the existence of automatic mechanisms for emotional connectivity other than those involving the MN system: evidence, perhaps, that empathy does not activate mirror neurons.

Questions about the number and extent of 'mirroring' sites; questions about their potential to handle different types of information (naturally-occurring or arbitrary, directly-observed or abstract); questions about the scope of the insights they may or may not offer… These are major criticisms, among which is a fundamental query asked by Brass & Heyes (2005: 489): what, indeed, are mirror neurons for? 'Imitation,' they comment, 'might well be one of the things that mirror neurons do;

under some conditions, in some species, mirror neurons could be involved in the generation of imitative behaviour. However, mirror neurons could *do* imitation without being *for* imitation (authors' emphases).' While the whole mirror neuron field offers great promise for future research work, the nature and scope of this are clearly in need of continuing re-assessment.

5.6 Alternative mechanisms for interaction: Theory of Mind

Another theoretical area that has obvious relevance to a discussion of social interaction – whether automatic or volitional – is theory of mind/mentalizing, or mindreading. By any standards, this occupies an important position in accounts of how successful interlocutors bridge the communication gap. If a Speaker's representations of what his interlocutors want, perceive, and think are reasonably accurate – if, in other words, his mental model of them is well-founded and up-to-date – he is not, in Baron-Cohen's expressive term, 'mindblind' (Baron-Cohen, 1995: 2), but a mindreader. As such, he equipped to identify the thoughts, desires and intentions of other people, and thus to understand or predict their behaviour. More particularly, within the context of interpersonal communication, he is equipped to interpret or tailor the utterances exchanged in a way that retains interlocutor relevance and thus helps ensure that the interaction continues. (The difficulties experienced by even a skilled and committed Speaker when interacting with an autistic – i.e. 'mindblind' – conversationalist are illustrated in Chapter 3, example 1.)

Crucially in the context of my own inquiry, mentalizing appears to show many classic System-1 characteristics: fast, automatic, smooth-flowing, capable of handling large amounts of data simultaneously (or in parallel). As Baron-Cohen comments (*ibid*: 3), 'We mindread all the time, effortlessly, automatically, and mostly unconsciously. That is, we are often not even aware we are doing it – until we stop to examine the words and concepts that we are using.' But close examination of this apparent automaticity reveals several contradictions and paradoxes. As the neuroscientific evidence shows, ToM's undoubted contribution to Speaker success is more complicated – and more volitional – than at first appears.

Research on the nature and mechanisms of mentalizing, which dates back to the work of Premack & Woodruff in the 1970s, approaches interaction from a different direction than mirror neuron investigation. However, it also acknowledges the suggestion from the MN literature that mirror neurons and mentalizing may be connected, even though the nature of the connection remains obscure. Frith & Frith, in their 2003/4 review of the development and neurophysiology of mentalizing, speculate that the human brain may contain a 'mechanism that enables an understanding of the meaning of actions, a differentiation of the goals of actions and the means to reach them' (*op cit*: 53), and that mirror neurons might be involved in such a mechanism. They go on to ask whether this 'potentially innate' predisposition to understand actions, together with a predisposition to detect agency and a preference for attending to conspecifics (e.g. human faces), might contribute to the development of mentalizing, but point out that, by themselves, these components are not sufficient for such a task, since they are shared with a great many other species which possess no mentalizing ability. The implication here appears to be that ToM itself – for all its System 1 characteristics of speed and automaticity – is relatively recent in evolutionary terms, and displays distinctively human characteristics generally ascribed to System 2. This apparent contradiction at the core of the mentalising process raises the possibility that, in dual process terms, it may be a hybrid system that combines the capabilities of both Systems 1 and 2: in fact, a dual system in its own right.

On the basis of the evidence they review, Frith & Frith locate the mentalizing function in three brain areas: the temporal poles (TP), the posterior superior temporal sulcus (pSTS), and the medial prefrontal cortex (MPFC). Of these, the temporal poles are a region of particular interest in the context of language use, as the left temporal pole is 'frequently activated in studies of language and semantics' (Frith & Frith, 2003/4: 57). Mentalising tasks that involve left TP activation include understanding stories and cartoons, inferring knowledge, and empathising with social transgressions. But the picture is made more complex by the fact that the same or adjacent regions are also activated by other processes, including the identification of familiar faces and voices, the comparison of sentences to unrelated word strings, the comparison of narratives of greater or lesser coherence, and autobiographical memory. Frith & Frith tentatively conclude that this region is concerned with

'generating, on the basis of past experience, a wider semantic and emotional context for the material currently being processed' (*ibid*: 58), and add that it 'may well' also be concerned with the retrieval of scripts, as these are one component of the wider semantic context.

According to Satpute & Lieberman's application of dual-process theory to neural systems (2006), one of the systems hypothesized as supporting automatic cognitive processes is the lateral temporal cortex (LTC), which includes the temporal poles. . The implication is, therefore, that the mentalising activities identified above in the Frith & Frith review may be carried out automatically. The second of the three mentalizing areas that this identifies, the posterior STS, is also a zone claimed by Satpute & Lieberman to be associated with System-1 processes and here too, by implication, the mentalizing activities with which it is associated are also carried out automatically. According to Frith & Frith, these are similar to those that activate the temporal poles: inferring knowledge, understanding stories, cartoons, and social transgressions. Other processes that activate the same or adjacent brain regions include seeing speaking mouths (cf Puce & Perrett, Chapter 5.4.3), attending to action presented as moving points of light, attending to eye gaze (a point of particular relevance in the current context) and attending to biological motion in general. Frith & Frith comment:

> 'These observations suggest that this region is activated when observing the behaviour of living things and also when retrieving information about the behaviour of living things… Knowledge about complex behaviour and, in particular, the ability to predict the next move in a sequence of behaviour is extremely valuable in any social interaction and could underlie some of the precursors of mentalizing, like gaze following and joint attention. Indeed, it is known that activity in the STS increases when volunteers are asked to attend to gaze direction… The mentalizing system goes one step further and uses the observed patterns of behaviour to perceive the mental states that underlie this behaviour'.
>
> (*ibid*: 60-1)

However, the third of the three mentalizing regions – the MPFC, defined in the Frith & Frith review as the most anterior part of the paracingulate cortex – is different. Although it has direct connections to the other two regions, interesting distinctions have emerged between its functionality and that of the others. While, for example,

the Friths' review shows that the MPFC is activated by tasks that activate the other two regions (e.g. inferring knowledge, understanding stories etc), there appears to be a difference between activations elicited by stories demanding an understanding of physical causality and those demanding the ability to mentalize by handling 'false beliefs'. All three brain areas, according to Frith & Frith, are activated by the mentalizing stories, but the physical ones activated *only* the temporal poles and the STS: 'The MPFC seemed to be particularly linked to mentalizing since it was the only area that was not also activated by the physical stories' (*ibid*: 55). Again, they point out, only the MPFC out of the three mentalizing brain areas has been shown to be activated during participation in interactive games.

An extended version of this result, captured by Rilling et al, involves a neighbouring region, the dorsomedial prefrontal cortex (DMPFC). In an fMRI study of subjects playing interactive games[10] that involve estimating partners' mental states, each subject plays both against a computer-generated algorithm and against real-life human 'partners' outside the scanner (in reality, also the output of a computer program), with the focus of the experiment being to explore the specifically social function of the three 'classic ToM areas' (Rilling et al, 2004: 1700): the anterior paracingulate cortex (i.e. MPFC), the posterior STS and the temporal pole. For both games, the results showed activity in both the MPFC and the pSTS, a finding consistent with the apparently hybrid nature of ToM. Very interestingly, the results from these areas also show a clear distinction between the subjects' mentalizing activities when dealing with – as they thought – a human player, and a computer-generated one. Both regions, Rilling et al continue,

> … responded to decisions from both human and computer partners, but showed stronger responses to human partners in both games. The stronger response to human partners is consistent with the behavioural data showing that participants distinguished between human and computer partners, rejecting unfair offers from human partners more frequently in the [Ultimatum Game} and cooperating more often with human partners in the [Prisoner's Dilemma]. The fact that computer partners are able to activate this network, albeit to a lesser extent than human partners, suggests that either this neural system can also be activated by reasoning about the unobservable states of nonhuman

[10] The Ultimatum Game and the Prisoner's Dilemma.

> systems, or that participants imbue their computer partners with human attributes.
>
> (*ibid*: 1700-1)

In the light of the robust research findings on the anthropomorphisation of computers (see, for instance, Chapter 4, n. 14: Reeves & Nass, 1996), the second of these two suggestions seems well-founded: a possibility that is given further substance by the fact that Rilling et al's data 'show that people are more likely to engage theory of mind brain areas with computer partners when the latter are perceived to be responsive to their human partner's choices' (*ibid*: 1702). Is this finding also showing a version of the Addressee-Speaker feedback loop, in which an electronically attentive Addressee is – by its responsiveness – enabling its human interlocutor to access the processes that will in turn allow him to model the Addressee's 'mind' more effectively… and thus keep the interchange going?

Lieberman's own hypothesized mapping of dual-process theory onto neural regions also brings out the hybrid nature of the neural processes that support ToM, citing observed activity in both the pSTS and temporal poles (associated with automatic processing) and in the DMPFC (associated with controlled processing), and pointing to distinctions in the aspects of ToM activity with which the different regions are connected . One of these is the fundamental division between relatively automatic processes, such as sensitivity to biological motion (associated with the pSTS), and 'explicit propositional thought about the content of another's mind [which] would seem to fall squarely within the domain of controlled processes' (Lieberman 2007: 263-4). It is interesting to note, Lieberman continues, that 'the only medial activation (DMPFC) associated with theory of mind processes is associated with the internally-focused process of considering the contents of another person's mind. Sensitivity to biological motion and personal familiarity in lateral regions (pSTS and temporal poles, respectively) are both externally-focused processes that do not require consideration of a target's internal states.' (*ibid*: 264)

This same distinction can be perhaps be seen at work in the application of ToM to neural activity specifically associated, not just with language use, but with pragmatics – again an area where attention has to focus on the contents of the speaker's (or author's) mind. This is demonstrated by Ferstl & von Cramon's

identification of the same MPFC region's involvement in both mentalising tasks and language processing ones, as compared to control conditions. An example of the first type of task is identifying the feelings of the person described in the sentence pair, 'Mary's exam was about to begin. Her palms were sweaty.' An example of the second is deciding if there is a logical connection between the sentence pair, 'The lights have been on since last night. The car doesn't start.' (Ferstl & von Cramon, 2002; quoted by Frith & Frith, 2003/4: 65-6). Citing Sperber & Wilson's theory of relevance (1995), the Friths point out that, in this theory's terms, then 'pragmatics, the understanding of utterances, depends upon mentalizing whether or not this is required by the task instruction [as in Task 1]. This would apply also to the type (ii) sentences used by Ferstl & von Cramon… where logical connections had to be found. For instance, the example above may evoke the idea that "someone (stupidly or maliciously) left the lights on"' (*ibid*: 66).

Further evidence for neural connections between mentalizing and communication – which again points to the duality of the ToM system – comes from an fMRI study by Kampe, Frith & Frith (2003). Kampe et al take the view that a prerequisite to successful communication is the recognition of the intention to communicate: an intention conveyed by signals such as calling a person's name or making eye contact that indicate to the potential addressee that it is he/she who is the communicative target. 'The question we wished to address was whether the neural circuit involved in mentalizing is also engaged in the initial stage of communication, when the intention to communicate is signalled. Typically, a subject has to recognize that such a signal is directed at himself. If recognizing the communicative intention of another toward oneself triggers the mentalizing mechanism, then perception of a variety of signals, normally associated with the intention to communicate, should activate the neural circuit implicated in mentalizing' (Kampe et al, 2003: 2).

The study compares reactions to images of faces that look straight at, or away from, participants, and also the reactions to hearing their own name or someone else's called out. The prediction is that both direct eye contact and hearing one's own name would activate the 'classic three' neural correlates of ToM: the paracingulate cortex (of which the most anterior part is the medial prefrontal region described in the Friths' review), the STS, and the temporal poles. In the event, two out of the

three are activated, in both visual and audio modalities: the right paracingulate cortex and the left temporal pole. Additional activations in other brain regions are not observed.

Although subjects here are required to react, not to signals of communicative intent, but to stimuli that downplayed urgency[11], their reactions to the former seem to follow the dual pattern outlined by Lieberman: automatic, effortless TP activity for the processing of an external stimulus (the TPs have been associated with the processing of both familiar faces *and* voices), and the controlled, effortful MPFC activity required for considering the contents of another's mind. As Kampe et al explain, someone who hears his name called, whether or not he is required to react, still needs to mentalize along the following lines: 'I am Chris. I heard the word "Chris". Is this "me Chris" that is meant? Or any other Chris? Does the person who just called "Chris" want to address me?' (*ibid*: 9). It is arguably the MPFC that is specifically equipped to support this internal debate – one that, though conducted by a C-system process, still takes place fast enough to enable Kampe et al to add that 'in normal people[12]…mentalizing appears to be a rapid automatic process that does not require conscious effort.'

Overall, Frith & Frith close their 2003/4 review of research in the mentalizing field by concluding that the MPFC region associated with mentalizing tasks is activated whenever people are representing – i.e. modelling – the thoughts, feelings and beliefs both of themselves and others. They point out, however, that this is far from being the whole story. 'Clearly, other components of the mentalizing system need to supply the content of these thoughts, feelings and beliefs and their relation to people's actions. This knowledge is supplied partly from our knowledge of the world based on past experience applied to the current situation and partly from our observations and predictions about people's current behaviour (STS). Both types of knowledge help to understand the content of mental states and their relation to actions, and may be accessible via temporal poles and the STS' (*op cit*: 67). My own account turns to this crucial question of content in the next chapter. In the

[11] Their actual task was to detect faces with eyes closed in the visual condition, surnames rather than first names in the auditory one. Both of these stimuli appeared only infrequently.

[12] In contrast, they note, high-functioning autistic patients have commented on the surprise they have felt when they realised, around the age of 11, that being called by name indicated that somebody wanted to communicate with them.

meantime, it may be worth asking whether we are looking at not one 'Theory of Mind' system but two, each of which is equipped to deal with the distinctive inputs that are *together* necessary to model the minds of others efficiently (via fast-acting associative processes) and effectively (via fast-learning controlled ones)[13].

5.7 Language and automaticity: the challenge posed by Broca's area

As discussed in Chapter 5.4.2-3 above, the neural correlates of the non-verbal aspects of communication seem to demonstrate a strong System-1 presence. Again, both Lieberman's account of verbally-expressed self-reflection and the Friths' review of the neurophysiology of ToM show intriguing connections between aspects of language processing and neural regions with hypothetical connections to System-1 activity: the role played by the lateral temporal cortex (LTC), for example, in implicit semantic memory; that of the pSTS in watching the mouth of a speaker and thus aiding comprehension[14]; and that of the temporal poles in connection with wider aspects of semantic or pragmatic processing, such as script creation and retrieval. It is now time to narrow the focus and consider how far neural regions claimed to support System-1 processes have been shown to support speech production itself.

A substantial literature exists on the neural correlates of speech production, one example being Indefrey and Levelt's meta-analysis (2004) of 82 neuroimaging studies of regions involved in the core processes of word production. This identifies 15 regions, four in the right hemisphere and 11 in the left. In addition to Broca's and Wernicke's areas, they include a region activated during listening (part of the

[13] And which may function differently at different stages of development. A recent fMRI study (Sebastian et al, 2011) of ToM duality – here, cognitive vs affective understanding – has shown more activation in the ventromedial PFC in adolescent participants during 'affective ToM' conditions than is elicited in adults. (Interestingly, the VMPFC is a region listed in the X-system category of Lieberman's framework.)

[14] Cf Haxby et al (2000: 230): 'Lip-reading plays a large role in speech comprehension, even in people with normal hearing. Lip-reading improves hearing accuracy and lip movements that are inconsistent with auditory speech can cause hearing errors… Perception of non-speech mouth movements is associated with activity in the [STS]. Lip-reading, in the absence of sound, additionally elicits activity in auditory areas in the superior temporal gyrus that are also activated by hearing spoken words. This indicates that the representations of speech-related lip movement involves the coordinated activity of visual regions in the [STS], which are associated with the visual analysis of lip movement, and auditory speech regions in the superior temporal gyrus, which are associated with the analysis of phonemic content.'

precentral gyrus) and one that is involved in both word and face recognition (part of the fusiform gyrus). Importantly, given the focus of this thesis on the conceptualizing aspect of speech production, they identify one region – and one region only – as being involved in the conceptually-driven process of lexical selection: part of the left middle temporal gyrus, a region boundaried by the STS[15]. The Brodmann reference number for this whole area is BA 21, and here, perhaps, there may an indication of a possible connection with the neural version of dual-process theory, as this region is part of the lateral temporal cortex, hypothesised in Lieberman's framework as being associated with automatic processing.

However, evidence that specifically connects the language-associated brain areas with dual-process theory seems hard to find, with a major challenge coming from the area once regarded as exclusively specialising in speech production: Broca's (left-hemisphere BA 44/45). The challenge starts with the fact that the capabilities of Broca's are now far from simple to characterise: as summarised in Nishitani et al's review (*op cit*: 66), these have now gone 'far beyond its classical language functions, [and contribute] to action planning, action observation, action understanding, and imitation.' Speech production and comprehension, they continue, might be considered a 'highly developed' form of action execution/observation matching, with Broca's emerging overall as an orchestrator of time-sensitive perceptual and motor functions underlying verbal and non-verbal communication. They conclude, however, that a variety of questions still remain to be addressed in unravelling the 'multitude of brain functions' (*ibid*: 67) to which Broca's contributes, including the temporal activation sequences and connectivity patterns involved.

There is also the question of how functions are distributed within Broca's area, with mappings between tasks and the area's internal regions being the subject of much debate[16]. For instance, there is an apparent overlap between BA 44/45, both of

[15] However, they point out that research on process timing suggests that this region may be more implicated in lexical selection than in conceptual processing as such. 'It seems plausible that tasks like picture naming and word generation [two of the word production tasks studied], which probably activate quite different concepts, should only converge and enter a common pathway from the point of lexical selection onwards' (Indefrey & Levelt, 2000: 123).

[16] The results of a comparative study by Heim et al of semantic, phonological and syntactic fluency (2008; 1362) give an illustration: 'Phonological fluency activated BA 44 more strongly than semantic or syntactic fluency…. Semantic fluency did not elicit higher activation than the phonological fluency tasks in any part of Broca's region. No differences were observed between syntactic and semantic

which lie within the prefrontal cortex, and the area suggested by Lieberman as one of the neural regions that support controlled social cognition. (Indeed, there is an obvious connection here, as the PFC is an area concerned with executive functioning and control.) And it has to be acknowledged that connections between such an area and Broca's are problematic for any model that seeks to explore the the automaticity of the mechanisms supporting relevant speech production.

This problem can be addressed, however, by considering the involvement of other areas besides Broca's in language processing. For example, Lieberman's list of regions associated with automatic processes includes the subcortical structural system of the basal ganglia, and it has been suggested (Ullman, 2006) that this is connected to a variety of cortical areas, *including Broca's*. Furthermore, the connection with Broca's is seen as being a double one, in which each linking pathway serves a distinct function. Explaining his hypothesis, Ullman suggests (*ibid*: 481) that one of these parallel and 'largely functionally segregated' channels may involve the anterior part of Broca's, with the other serving the area's posterior regions. Each of these, he continues,

> …is hypothesized to subserve functions that cut across language and non-language domains. The "anterior" channel may be characterized as subserving the retrieval of lexical and semantic information stored in declarative memory. This memory system has been implicated in the learning and use of both conceptual-semantic and lexical knowledge… The "posterior" channel may be profitably viewed as subserving aspects of procedural memory. This system underlies the acquisition and real-time expression of motor and cognitive skills, especially those involved in sequential knowledge… [and has] also been implicated in grammar, in particular in aspects of the acquisition of grammar and in rule-governed grammatical composition.

Another approach to the potential role of automatic neural processes in linguistic processing is taken by Adolphs, continuing his focus on the amygdala as a key player in the social cognition field. Initially, he discounts any link between the amygdala and entities that are verbally represented rather than directly perceived. In a study of neurological patients with severe damage to the amygdala, Adolphs et al (1998: 470) demonstrate a contrast between reactions to 'unapproachable and

---

fluence. Thus, the activation of BA 45 observed during verbal fluency tasks seems to be not restricted to semantic processing as suggested by the literature…'

untrustworthy looking' faces, on the one hand, and verbal descriptions of these faces, on the other. While the subjects responded to the visual stimuli in an atypical way (judging them to look more approachable and more trustworthy than did neuronormal control subjects), they reacted normally to the verbal descriptions – 'perhaps because the stimuli provided sufficient explicit information such that normal task performance could result from reasoning strategies that did not necessarily require the amygdala' (Adolphs, 1999: 473).

However, as he points out, there is evidence of a possible connection between the amygdala and lexical representation, especially where the amygdala's prominent role in processing threatening stimuli is called into play. For example, in a PET study by Isenberg et al, neuronormal subjects are asked to complete a modified version of the Stroop test, involving the colour naming of words that have either threatening or neutral connotations. The resulting PET scans show significantly greater amygdalar activation during the colour naming of the threat words than of the neutral words. Isenberg et al (1999: 10456) take this to demonstrate the amygdala's role in the 'processing of danger elicited by language,' and add that the results confirm the amygdala's role in modulating the perception of, and response to, emotionally salient stimuli. In their view, their study further suggests the 'conservation of phylogenetically older mechanisms of emotional evaluation in the context of more recently evolved linguistic function.'

A further indication of a relationship between the amygdala and language processing can be seen in Lieberman's account of neural activity supporting social cognition. It has long been known, he states, that 'putting feelings into words is an effective strategy for regulating negative emotional responses …[and] the benefits seem to occur whether one is intentionally trying to regulate one's emotions or not' (*op cit* 2007: 270). This process of 'affect labelling' (i.e. giving verbal labels to emotionally evocative visual images) has, he continuesbeen shown by a number of neuroimaging studies to *reduce* the amygdala activity that such images would otherwise produce.

This negative reaction in a neural region of interest – a decrease, rather than an increase, in activity – is arguably as important a feature of neural mapping as its opposite and one to which I will return (see Chapter 6.4). And, more generally, the

whole area of automaticity in linguistic functioning could clearly be a rewarding one for further research.

5.8 Summary and conclusions

After summarising the arguments of the first half of the thesis, I have developed my hypothesis by suggesting that research into the control exercised by an Addressee over a Speaker (see Chapter 4) might help to explain not only why Speakers select the messages they do, but why, in doing so, they must succeed in achieving optimal relevance. Failure to achieve optimal relevance, I argue, impairs the processes underlying message selection. Speaker and Addressee are therefore linked in a feedback loop that enables dialogue – and communication itself – to function as a single, tightly-connected system of personal interaction.

The main part of this chapter explores how Speakers automatically play their part in maintaining this loop under the heavy time pressures of normal dialogue, with particular reference to resources contributing to their success at the level of social interaction.

The exploration begins by examining the core framework and some of the main variations of dual-process theory, described by Evans (2003: 454) as postulating 'two minds in one brain', with two associated processing styles. As generally conceived, one of these processing styles – 'System 1' – consists of a set of rapid, automatic, effort-economical, preconscious processes that work in parallel and are fed by a corresponding memory system that absorbs material slowly and retrieves it fast. The other – 'System 2' – works slowly and effortfully, carrying out consciously-requested functions on behalf of its owner, and is capable of quickly absorbing new material but slow to retrieve it.

One point made in this chapter (Chapter 5.3.6) is that little detailed research seems to have been carried out on the place of language production in dual-process theory. However, there is some evidence from dual-process studies that many language processes indeed belong to System-1, and operate automatically and effortlessly: this

might help to explain how the hypothetical Addressee-Speaker feedback loop functions.

In Chapter 5.4, I consider evidence in favour of this suggestion derived from the rapidly-growing field of social cognitive neuroscience where, amongst areas associated with automatic responses, one – the dorsal anterior cingulate cortex – is of particular relevance to my hypothesis. As described in Chapter 5.4.2, activity in this region has been shown to be associated both with the affective distress caused by physical pain and by distress resulting from social exclusion (i.e. social 'pain'): a finding that, I suggest, may be involved in the damage done to Speaker competence when confronting the social exclusion indicated by Addressee indifference. The section also discusses evidence associating automatically-responding brain regions with face perception in general.

Chapter 5.5 and 5.6 consider two other theoretical frameworks that may offer insight into the connectivity between Speaker and Addressee: mirror neuron research and theory of mind (ToM). Although the implications of mirror neuron theory for human social interaction are still the subject of considerable controversy, research on ToM shows that the mentalizing capacity may be a dual system in its own right, calling on both automatic and controlled neural processes to fulfil different mentalizing tasks. The chapter ends with a discussion of possible interactions between language production and comprehension processes and neural regions associated with automaticity, particularly the basal ganglia and the amygdala.

Some of the results discussed in the chapter are controversial, and many come from young areas of research, accompanied by the warning that such results must be regarded as preliminary (cf Lieberman, 2007: 276). Overall, however, they appear to offer some support for the idea that there is a wide range of automatic social responses that help a Speaker adjust continually to an Addressee's needs. How far this automaticity extends to the verbal level of interaction is discussed in Chapter 6.

CHAPTER 6

The automaticity of verbal interaction

Introduction

In the previous chapter I considered the infrastructure of automatically-achieved successful dialogue. I outlined some of the psychological and neurological means by which a Speaker keeps continuously attuned to an Addressee's attentiveness or lack of it and is thus able to produce coherent, relevant speech. In this chapter, I consider more closely the missing element in this account, the speech production process itself.

After returning to Pickering & Garrod's account of priming-driven alignment between interlocutors, I look at two aspects of message generation and production in which automaticity appears to play an important role. The first (Chapter 6.2) is the use of speech routines, highlighted by Pickering & Garrod themselves, and examined in detail in Kuiper's account of formulaic speech patterns used by auctioneers. The second (Chapter 6.3-5) considers priming from the behavioural – as opposed to linguistic – perspective. Against the background of work by Bargh et al, Dijksterhuis et al, Kay et al and others, it suggests that behavioural priming techniques could automatically affect message production in the same way that linguistic priming affects utterance formulation. I end by outlining some recent experimental work in which I have been involved that appears to confirm a two-way link between behavioural and linguistic representations, with verbal primes affecting social behaviour and behavioural primes also affecting linguistic behaviour.

6.1 Why do people say the things they do?

How far does the theory of automatic Speaker-Addressee rapport outlined earlier extend to the uppermost layer of speech production: the generation of Addressee-

attuned messages, constructed under demanding cognitive and time constraints, and expressed in words the Addressee can readily process within those same constraints?

To some extent, potential solutions are readily forthcoming, as shown by the considerable discussion on the role of interlocutor priming in dialogue described above (see Chapter 2). Indeed, Pickering & Garrod's 2004 model of dialogue uses this 'essentially resource-free and automatic' process to account for the tight coupling between interlocutors' utterances at every level – from phonetic, through phonological, lexical, syntactic, and semantic, to that of the situation model itself – with the computational burden being greatly reduced by the fact that priming processes are automatic.

The automaticity of priming has immediately obvious benefits for interlocutors. However, it also has certain drawbacks, notably in the restriction of Speaker choice. As Bock points out, 'Processes such as [syntactic priming] may limit the flexibility of syntax in the spontaneous, real-time use of language, perhaps contributing to differences in syntactic diversity between planned and extemporaneous speech' (Bock, 1986: 379). It also raises a more serious question about the aspects of dialogue that a priming process seen as primarily linguistic can be expected to explain. Pickering & Garrod acknowledge (*ibid*: 188) that, although their model substantially reduces computational effort 'there are still some conscious decisions [to be made] about what one wants to talk about', and go on to make their major disclaimer about the upper limits of their model's scope:

> 'Our contention is that interactive alignment (and in particular the automatic alignment channels) affects the structures used in production and interpretation rather than directly determining the *content* [authors' emphasis] of production and interpretation. … We assume that alignment provides an explanation of the manner in which interlocutors produce and interpret contributions. So we propose that alignment channels only produce a direct link *between the structures* [my emphasis] that the interlocutors use in language processing. Hence, the alignment process is automatic and direct, even though it does not determine exactly what the speaker produces (as this depends on his long-term knowledge) or how the addressee interprets what he hears "beyond" the level of the situation model.'
>
> (*ibid*: 213)

The manner, not the content; the structures used in production, not the production itself: is this as far as the priming explanation can take us in describing how a hypothesised low-level, automatically-operating system of interaction between Addressee/Speaker and Speaker/Addressee can operate? On the face of it, Pickering & Garrod's argument is compelling, and focuses precisely on limitations of choice. At all the utterance levels described in their model as accessible to the priming mechanism, Speaker fluency and effortlessness are promoted by choice restriction which, in turn, implies predictability and increased ease of comprehension. But at the level (or levels) beyond their model's reach, these considerations cease to apply. A Speaker's long-term knowledge store – his 'encyclopaedic knowledge' in the Levelt model (see Chapter 2.1) – is indefinitely large, and its contents indefinitely varied. Except in the broadest terms, they seem beyond the power of an Addressee to predict while, from the Speaker's point of view, there is no theoretical limit on subject-matter choice at all: he could choose anything at all to talk about and, even within the constraints of context and social norms, his scope for choice is still indefinitely wide.

Agreed, the fact that in practice he usually limits himself to subject-matter that will hold his Addressee's attention in a given context seems to indicate that 'conscious decisions' are indeed being continuously made, utterance by utterance. However, this would involve so much mental effort that it might well undo much of the cost-effectiveness achieved in dialogue by linguistic priming, as discussed above. Nor is this the only paradox. A second problem is the one to which I constantly return: that of speed, now linked with the implications of dual-process theory: however much conscious decision-making is required in the construction of utterances, they are still produced with an address and fluency generally characteristic of a System-1 response. And more challenging still, why does the newsvendor's joke, with its arresting hyperboles (Example 6, Chapter 3.3.4) take just the form it does? Is there any way of explaining what prompted the bystander in Example 9 to choose that particular comment on the unhappy cat, given the indefinite number of variations on the theme that he could have chosen? And, in the emergency-driven and very fast exchanges of Example 11, below, could the man's comment have been in any way predicted?

Example 11

Context: *The scene is a glasshouse at Kew Gardens, where small pools line the walkways. A young couple have halted by a pool: the girl has crouched down and is peering into it closely. AJP is a few paces away.*

a) Girl (*leaps up and screams*): Aaaagh! *(Screaming and laughing at once, she hugs the man; at the same time, she catches AJP's eye)*
b) AJP *(peers into the pool: a large white salamander swims into view):* Oooh!
c) Man: It must have come from Mars
d) AJP: Yes, it must have. (*to girl)* Did it bite you?
e) Girl: Oh no

In summary, just why do people say the things they do? Why do they say things that will, more often than not, capture and retain the attention of their hearers, with the result that they themselves are licensed to continue saying things? And how far is this vital feedback loop supported – as observation seems to indicate – by automatically-conceived speech: System 1 speech? This is a major issue – perhaps the biggest – involving the automaticity or otherwise of speech production. In this chapter, I describe two ways of addressing it, one to some extent established (though still under-researched) and the other speculative but which, I suggest, could offer considerable promise.

6.2 Short-circuiting speech decisions: routines, routinization, and 'speaking appropriately'

Levelt, in his discussion of the processing systems involved in his speech production model, stresses the deliberate nature of conceptualization, as compared to the largely automatic processes of formulation and articulation: 'message construction is controlled processing'. Even here, however, he leaves certain openings for automaticity, acknowledging that an adult's experience with speaking 'is so extensive that whole messages will be available in long-term memory and thus will be [automatically] retrievable' (*op cit*: 21; see Chapter 2.1.1). Pickering & Garrod,

in their alignment-oriented account of dialogue, pick up and considerably develop Levelt's 'whole message' point, and I will argue that this may make it possible to move beyond the Leveltian view of message production as being mainly under executive control (i.e. conscious, serial, slow) to a more automatic model on which messages can be continuously, fluently *and speedily* conceptualised and produced. The following sub-sections explain how this progression from one model to the other might start.

6.2.1 Ease or necessity? The benefits of choice restriction

Pickering & Garrod's speech alignment process implies that interlocutors draw on representations that develop as a dialogue proceeds, thus bypassing the need to construct them from scratch. As they point out, one important implication is that interlocutors develop routines, or set expressions, in the course of an interaction, and go on to use them. 'It is entirely possible,' they suggest, 'that people do not always retrieve each lexical item as a result of converting an internally generated messages into linguistic form… [but] rather that people draw upon representations that have been largely or entirely formed already' (*op cit,* 2004: 181).[1] They also point out that repetition of utterances in dialogue, far from seeming exceptional or unnatural, is standard practice: 'normal dialogue is *highly* [authors' emphasis] repetitive'. Within this overall picture, routines – expressions whose characteristics at every production level are relatively firmly fixed – play a large part[2] and Pickering & Garrod suggest that there may be a second category of repeated utterances whose characteristics depend uniquely on the interlocutors themselves. These are routines that are set up 'on the fly' as the dialogue proceeds. Thus, if an interlocutor uses an expression in a particular way, it may become a routine unique to that dialogue: a process they call 'routinization'. The convergence of terminology in Garrod & Anderson's maze game (1987) is an example of routinization in action, and the same could be said of the

[1] Kempson et al's Dynamic Syntax architecture (2001) explains this outcome in different terms. Pickering & Garrod's proposed alignment patterns across different levels can, in Dynamic Syntax, all be expressed as re-use of the lexical and computational actions used to create particular semantic structures: 'This result is achieved since context, content and lexical actions are all defined in terms of the same tree configurations' (Purver et al, 2006: 317).

[2] As demonstrated by phatic utterances such as those described in Chapter 3, especially by Examples 3 and 5. (The 'project markers' described by Bangerter, Clark & Katz (Chapter 2.3.3) are also routines which they see as vital to the successful navigation of a dialogue.) Pickering & Garrod quote one estimate (Aijmer, 1996) that up to 70% of words in the London-Lund corpus occur within recurrent word combinations.

development of the 'dancer with the big fat leg' label in Schober & Clark's 1989 experiment (see Chapter 2.3.2). Either way, Pickering & Garrod see the use of such expressions as contributing greatly to dialogue fluency by dramatically reducing speaker choice.

As an example, they cite the development of the routinized phrase 'the previous administration' in a political dialogue. On its first appearance in the exchange, they suggest, the expression will have been constructed by accessing the meaning of 'previous' and combining it with the meaning of 'administration' and, when processing it, both Speaker and Hearer will have made important choices in terms of message construction and interpretation. But, they continue,

> … if the expression is repeatedly used, the interlocutors do not have to consider alternatives to the same extent. For example, they do not have to consider that the expression might have other interpretations, or that "administration" is ambiguous (e.g. it could refer to a type of work). Instead, they treat the expression as a kind of name that refers to the last … Government. Similar processes presumably occur when producing expressions that are already frozen… Generally, the argument is that people can "short-circuit" production in dialogue by removing or drastically reducing the choices that otherwise occur during production (e.g. deciding which synonym to use, or whether to use an active or a passive form).'
>
> (*ibid*: 182)

The obvious explanation for this short-circuiting phenomenon, according to Pickering & Garrod, is that routines are generally easier to produce than non-routines.[3] It is obvious, certainly, and appears to be supported by neuroscientific evidence. Posner& Raichle (*op cit*: 125), for example, give evidence of the effect of practice on word choice: once the task has become routinized, the brain regions involved during the original decision-making remain unactivated, and activity only returns when further original word choices are required. But this explanation also

[3] Relevance theory suggests that they also reduce processing effort on the Addressee's part, as explained by Vega Moreno (2005: 250-2): 'If both a familiar expression and a novel one would achieve the same cognitive effects, but using the familiar expression minimises the hearer's processing effort, a speaker aiming at optimal relevance should choose the former… Familiarity with a stimulus may direct the mind to follow a certain familiar route in processing it, minimising the processing effort invested.' This explanation accounts for the fact that routines developed with one interlocutor may not carry over to subsequent dialogues with others; for example, the 'pathway' descriptive format developed by some couples in Garrod & Anderson's maze game (see Chapter 2.4.3) would cause problems if used with players who had developed the 'map coordinates' approach.

has far-reaching and diverse implications. Within the context of Pickering & Garrod's own priming-based model, for instance, 'easy' can be read as a synonym for 'inevitable'. According to this model, previous utterances in a dialogue will activate interlocutors' syntactic, semantic and lexical representations, which leads to repetition at these levels, to links between them, to consequent large-scale alignment, and 'hence the production of routines.' As shown by some of the examples in Chapter 3 (especially Example 3), this automatically-achieved development of an exchange is the very essence of phatic communion – Malinowski's 'flow of language, purposeless expressions of preference or aversion… comments on what is perfectly obvious'. (The halting interchanges in Example 1 between Uta Frith and the autistic Ruth illustrate the problems that occur when, in one of the interlocutors, the repetition/alignment process does not work: arguably an important contributing factor to the social and communicative problems characterising autism itself.) Example 12, quoted by Carter (2002) from the CANCODE spoken English corpus, illustrates the development of a similarly 'purposeless expression of preference', the relevance of which actually lies in the reaffirmation of a relationship between friends after a weekend's separation:

Example 12

Context: Three students – all female – are having Sunday tea in the house they share.

a) Speaker 3: I like Sunday nights for some reason. [*laughs*] I don't know why.

b) Speaker 2: [*laughs*] Cos you come home.
c) Speaker 3: I come home+
d) Speaker 2: You come home to us.
e) Speaker 3: + and pig out.
f) Speaker 2: Yeah yeah.
g) Speaker 3: Sunday is a really nice day I think.
h) Speaker 2: It certainly is.
i) Speaker 3: It's a really nice relaxing day.

(CANCODE/ Carter, 2002)

This exchange neatly illustrates the development, via repetition, of two separate themes ('Coming home' and 'Sunday'). But it also suggests another possible reason why routines are frequently used in natural speech, and this has less to do with ease of production than with outright need. When (as here) the need to say something – anything – is paramount in order to keep the interaction going, the easily-maintained repetitions of routine-laden or routinised speech are an obvious solution. And the same explanation applies to speech contexts far removed from that of easy-going social interaction, where the Speaker must bring a professional – and sometimes legally-binding – degree of precision to his utterances, and where time pressures are particularly unforgiving. This is the situation studied in Kuiper's important analysis of the speech patterns of auctioneers and sports commentators: an analysis whose conclusions parallel those of Smith and DeCoster (see Chapter 5.3.4) as to the underlying memory processes involved, and thus support the hypothesis of System 1 involvement in message generation.

6.2.2 'Off the tussock': the formulaic speech of auctioneers

Kuiper's study of speech production routines, or formulae, starts from the same point as Pickering & Garrod's account: 'Formulae make the business of speaking (and that of hearing) easier… Such expressions likely exist as whole or part utterances within the speaker's dictionary and need not be built up from scratch on every new occasion' (Kuiper, 1996: 3). And the questions underlying his research are very similar to those studied in this thesis: in particular, how do humans speak 'creatively yet appropriately' in particular situations, and manage it at the speed that they do? Unlike Pickering & Garrod, however, he focuses solely on routines: speech that is embedded within particular social contexts, learned and memorised within those contexts, and retrieved when the context re-appears, thus enabling a speaker to produce utterances 'appropriate' to the occasion. 'Speaking is therefore a matter of recalling speech in the context in which it is appropriate' (*ibid*: 4-5).

According to Pickering & Garrod, this is not an accurate reflection of normal dialogue practice, which they see as making standard use of temporary formulae developed between individuals to suit individual circumstances, as in the Garrod & Anderson maze game. But although the lack of such a routinisation element in

Kuiper's analysis might seem to indicate an equal lack of audience-tailored flexibility (and therefore 'creativity') in a routine-user's output, the use of formulae by a practised speaker shows a remarkable degree of variety and versatility, as demonstrated in Kuiper's recordings of the speech of master livestock auctioneers. Excerpts from one of these are given in Example 13 (note the development of the 'hills' theme):

Example 13

Context: New Zealand auctioneer Alistair Hopkinson is conducting a sheep sale in North Canterbury. The lot he is selling comes from a nearby farm, and has been dipped against disease.

They're absolutely sound in the feet, gentlemen,
and they're off the hills straight behind you out there.
They're absolutely natural condition sheep.
They're off the tussock.
I'll make you buy it at fifteen twenty,
Fifteen forty,
Fifteen dollar forty,
Sixty,
Eighty,
Fifteen eighty.
Come on…

They're off the tussock right behind us there, Greg.
You can see the peaks right behind us there.
Fifteen dollars eighty I got…

Who's going to round it?
Yes or no, Sir?
They'll do all right.
They're absolutely guaranteed in the feet.
There's no footrot up there on them there hills.
And I got fifteen eighty for 'em.

At fifteen dollars eighty.
Is that a bid?
Yes or no?…'

(*ibid*: 84-5)

Comparing output like this with that of a less experienced auctioneer, Kuiper emphasises important differences between the two. The formulae used by the junior practitioner are short and more frequently repeated than those of a more experienced performer. A master auctioneer, however, is able to produce a less repetitive and more fluent performance which, while still formula-based, is more idiosyncratic and thus apparently more applicable to a particular context: more 'spontaneous'. The play on the 'gold in them thar hills' formula is particularly felicitous; this is sales talk at its most elegantly persuasive. However, where the *reason* for formula use is concerned, such persuasive skill is only half the story. The key – as with Pickering & Garrod's account – lies in the restriction of choice that formulaic speech imposes on a speaker. 'Formulaic speech,' as Kuiper points out, 'cuts down the number of choices available to the speaker and consequently reduces the amount of linguistic information that must be retained in working memory and linguistic processing that must be done while still allowing for the speaker to maintain a high level of fluency' (*ibid*: 31). And Hopkinson and his colleagues use it, not so much because such use is inevitable – a natural by-product of Pickering & Garrod's alignment process – as because it is essential. Without such automatic message generation, they could not do their job.

It should be stressed here that, although Kuiper's focus is on only the auctioneer's side of the interaction, an auction is inherently a dialogue – a specialised form of dialogue, perhaps, in which the buyers' input is not necessarily verbal, but a dialogue nonetheless, and one involving several partners. Although an auctioneer's speech rate Kuiper quotes is not dissimilar from the overall figure given by Levelt (four to five syllables, as opposed to two to three words, per second[4]), auctioneers themselves

[4] An exception, however, is the speech that Kuiper records of some US tobacco auctioneers, who perform at the 'abnormally fast' rate of eight-plus syllables per second. Here, however, formula use has gone beyond the mere use of set phrases: the words themselves are distorted so as to promote ease of production . As an example, the following 21-word sentence was produced by a North Carolina auctioneer in 4.5 seconds ('bi/bee' = 'five'): '"A dollar seventy bi bi eighty dollar eightly eightly bee bee bee eightely bee eightely bee eightyfivedollar bill now now Virginia." (*ibid*: 51)'

are subject both to the usual pressures felt by interlocutors and also to the heavy pressures unique to their situation. Like any participant in a successful dialogue, they have to be attentive to their partners, but this attentiveness must be especially unremitting: they are constantly scanning buyers for signs of bids, which may come from any direction and may take a variety of forms. Their monitoring of the situation must likewise be unremitting; they must remember exactly where they are in the bidding sequence and, at the same time, be ready to instantly modify this representation: 'the moment they see a bid they must react to drive the market on and signal accurately to all the potential buyers what is currently the highest bid' (*ibid*: 36).. And, above all, the auctioneer must keep the procedure moving at a steady, rhythmic pace, without gaps, even though this rhythm is frequently not matched by the rhythm of the bidding itself. Pressures such as these (together with the similar time and accuracy constraints present in race calling) make auctioneering and sports commentating ideal contexts in which to test the hypothesis that, in Kuiper's words, 'using the linguistic resources of formulaic speech is a response by speakers to particular pressures on them. If they must speak fluently and fast when there is also a lot going on outside to which they must attend, then they will resort to using formulaic speech' (*ibid*: 26).

The test procedure he follows is to compare the speech patterns of auctioneers who are working to different required rates of sale, and are therefore under different types of time pressure (e.g. slowly-proceeding, specialist auctions, such as those at Sotheby's; medium-fast ones, such as livestock and furniture sales; and fast-moving events such as US tobacco sales, where a lot can be sold every five seconds), and his conclusion is that, in the first and third of these categories, formulaic speech is not greatly in evidence. At slow sales, there is no need: 'normal speech will do if auctioneers in a low-pressure tradition become stuck' (*ibid*: 47), while at fast sales – where, in some cases, proceedings are conducted entirely in numbers – there is no time. At medium-fast events, however, the picture is different. Auctioneers here have more time in which to make sales, and they need it, as their audience may include numbers of inexperienced buyers who need encouragement, guidance, and other types of interaction. And this is where features of formulaic speech – which include droned or chanted intonation together with abnormal fluency and formulae themselves – are most in evidence. Kuiper's prediction that the occurrence of

formulaic features in auctioneers' speech varies according to the pressure on the speaker's working memory and processing capacity is thus confirmed: 'auction speech… clearly shows that as cognitive pressure on an auctioneer increases, his ability to say things in an infinite number of ways decreases until, as in the case of the wool auction [in Christchurch, New Zealand, where bid calling is almost entirely done by the buyers] he is left with almost nothing to say' (*ibid*: 73).

6.2.3 'Responses that are made quickly': formula use and the associative processing system

As Example 13 shows, formulaic speech does not have to sound stilted or limited. Indeed, within its limits, it can be extremely varied: for instance, Kuiper's compilation of formulae used by one skilled auctioneer features nine standard descriptions of a lot (e.g. 'Oh, what a beauty', 'Isn't that lovely', 'There's a nice lot for you'), six formulae for attracting the first bid ('How much for that?', 'X dollars I've got for it') and six interpolations, used to keep the action going ('Put your hand up if I can't see you', 'This is something good here'). Nor, very importantly, need the use of formulae be restricted to such extraordinarily demanding professional contexts as the auction room or the commentating box; as Kuiper points out, speech at meetings contains many formulae, as do court proceedings and medical interviews. 'It seems that very many of the things we do socially with words are not, in fact, done with individual words. They could be done with words but they are actually done, most of the time, with formulae' (*ibid*: 94).

Taking this argument further, he suggests that every speaker, or performer in a speech community, uses resources traditional in that community – i.e. formulae – and that formula use may therefore be generally pervasive in society. And this enables him to offer a solution to the other question posed at the start of his analysis (as frequently in my own thesis), about the speeds involved in producing speech that is both fluent and appropriate:

> In answer to [the question] of how it is that humans are able to speak in real time, I have shown that, under many different circumstances some or all of the properties of formulaic speech are resorted to in order to cut down the options made available by the speaker's internalised grammar. This is done largely by replacing syntactic processing by lexical lookup. Cutting down the options must

> assist with real-time processing of speech because having fewer options involves less processing. Less processing provides speakers with more time to do the processing that must be done… Formulaic speech creates opportunities for reducing computation by reading larger than one-word chunks out of memory. This creates more processing time for speech planning. In turn, this makes real-time speech processing easier in some circumstances and perhaps even possible in others…. To be a native speaker under real-time pressures involves the use of formulaic resources.
>
> *(ibid*: 98)

An obvious example of such traditional speech resources in use has already been given: the 'convoys' joke about buses in Example 5 (see Chapter 3.4). But do Kuiper's conclusions offer some insight into the wider question posed in Chapter 6.1: why do people say the things they do – even things that seem as carefully, *deliberately* chosen for contextual applicability as in Examples 9 and 11?

It seems at least possible that the passer-by's comment on the wailing cat ('They don't like it, do they?') is formulaic rather than put together on the spot, carrying as it does faint echoes of Lance-Corporal Jones in television's *Dad's Army*. (It's worth noting that, in terms of age, both he and I are part of this programme's natural audience group.) And it also seems possible that the 'salamander from Mars' comment (Example 11) is not the result of careful decision-making in the young man's Conceptualizer, but a version – an effortlessly, automatically-retrieved version – of the concept MONSTER, expressed in terms of the formula '[thing from] Mars'. As with Chapter 3.4's examples of phatic communion, this too may be speech on auto-pilot, but produced to much subtler effect than the obviously formulaic expressions of agreement or sympathy illustrated earlier.[5]

At a more fundamental level of importance, the correspondences between Kuiper's findings and Smith & DeCoster's dual-process model of memory are also obvious, particularly in the latter's discussion of the demands of rule-based (i.e. sequential, slow) processing characteristic of System 2 (cf Chapter 5.3.4). Adequate motivation and cognitive capacity are both needed for this, where 'capacity' refers not just to attentional resources but also to available processing time. Thus, Smith & DeCoster

---

[5] Several examples of creatively modified formulae are quoted by Vega Moreno (2007: 210): e.g. 'OK there! Now you are *barking up the right tree!*' and 'If it is true he has found out about my affair, he must now be in my house *pouring the beans* to my wife.'

comment (*op cit*: 117), 'responses that are made quickly or when the perceiver is busy or distracted likely will be controlled by the associative system' – in other words, they will be the product of the automatic, preconscious set of processes brought together under the label of 'System 1' in the generic dual-process model. Quick *verbal* responses, made when the speaker is trying to negotiate the delicate opening stages of an interaction, or is prompted into speech by an emergency, or (like the auctioneer) is keeping up with a unremitting torrent of external events: this seems to be the output of associatively-sourced, System-1 message generation, based on routines and formulae that range from the almost ritualised to apparently idiosyncratic.

Kuiper's account of the acquisition and use of formulaic speech is also consistent with another aspect of the associative model: its method of operation. The associative system, according to Smith & DeCoster, is a slow-learning one – even extending over a lifetime – that involves 'small, incremental alterations of representations in ways that facilitate repetition of the same processing' (*ibid*: 110), and Kuiper's comparisons of the speech of master auctioneers with that of their juniors demonstrate just this lifelong progression. The pattern-completion mechanism that drives the associative mode can also been seen operating in both Kuiper's examples and my own, along with its speed of operation. The New Zealand hills visible to Hopkinson from the sheep pens in Example 13 act as the stimulus that elicits – from amongst the mass of information accumulated about hills over his lifetime – the automatically- and speedily-retrieved echo of the gold prospectors' joke. In the same way, the weirdness of Example 11's salamander seems a very plausible trigger for the equally automatic retrieval of the 'Martian' expression of the MONSTER concept.

To sum up, the widespread use of formulaic speech in both casual and formal contexts appears to support the suggestion that the conceptual level of verbal interaction – the level at which people decide to say the things they do – may involve automatic cognitive processes to a greater extent than has been previously acknowledged. And, as the work of Kuiper and Smith & DeCoster implies, speech thus produced by the 'learn slow, retrieve fast' system is almost certainly geared to being easily understood by the Addressee. It was learned slowly, via repetition, in specific social contexts (which, of course, include Addressees themselves). And,

when the context – or some salient element of it – reappears, the Speaker may be stimulated into producing an utterance 'appropriate' to both the occasion and his audience: a slowly-learned, quickly-retrieved formula.

Formula use is therefore an important way of achieving automatic verbal rapport with an Addressee, which merits more investigation that it so far seems to have received. However, it is not the only one. In the following sections, I return to the topic that has already featured several times in my discussion: priming.

6.3 Extending the priming model: behavioural priming and the message level

Priming is essentially a System-1 process. In Schacter et al's words, it involves a '*nonconscious* [my emphasis] form of memory in which an encounter with a stimulus influences the subsequent identification, production or classification of the same or a related stimulus.' (Schacter et al, 2007: 171). Pickering & Garrod's own account of how the activation of a linguistic representation increases the likelihood of that representation's use in a subsequent utterance refers to other characteristics that make priming a System 1 operation, particularly automaticity. And, as noted above (see Chapter 6.1), the benefits to both Speakers and Addressees of linguistic priming are considerable.

For instance, the limitations on syntactic flexibility that Bock (1986) alludes to may be offset by the advantages (discussed above) of choice restriction; as Bock says, 'selecting among the wide array of syntactic options available for the expression of a message in unplanned speech can create problems that lead to hesitations, errors, and other disruptions… Using procedures that are already activated may ease the demands of message formulation and actually contribute to fluency (*op cit*: 379-80). Meanwhile, in a discussion of Garrod & Anderson's 1987 maze game, Branigan et al point out that *both* Speaker and Addressee benefit from the semantic and lexical co-ordination produced (without explicit negotiation) during this experiment:

> The listener benefits because co-ordination increases the likelihood of correctly understanding the speaker's meaning… The speaker benefits because co-

> ordination reduces computational load: By previously interpreting an utterance in a particular way, the speaker has determined a set of semantic procedures (e.g. using 'square' to refer to a node, and 'two, three' to a row position followed by a column position) that can be re-used in production, and do not need to be computed from scratch.
>
> These functional benefits also predict that speakers will tend to be semantically and lexically consistent with their own prior utterances. Listeners benefit from within-speaker consistency because, if the listener resolves ambiguities in the same way as in previous (successfully interpreted) utterances, this resolution is likely to be correct. For speakers the facilitation gained from reusing previous decisions holds irrespective of whether they were originally made by speaker or listener… The same benefits should also hold with respect to other levels of linguistic structure.
>
> (Branigan et al, 2000: B14-15)

Pickering & Garrod's model of priming-enabled alignment between Speaker and Addressee takes this idea of dual benefit several stages further. But, as mentioned above, they conclude that there are limits to what can be achieved this way: alignment channels produce a direct link only between the structures that interlocutors use. So is utterance content – the message level, the decisions behind what people say and how they understand it – a hidden territory as far as priming is concerned, given the indefinite variety of interlocutors' long-term knowledge?

I want to argue that a priming-based model can be taken further, and that it can throw light on the issue of content from which Pickering & Garrod draw back.

One way ahead would be to consider the role of priming, not just in linguistic behaviour, but in behaviour in general. Behavioural priming is described by J.A. Bargh (2005: 38), as behaviour 'induced to occur by environmental factors and not by the individual's conscious awareness and intentions'. Bargh, one of the leading authorities in the behavioural priming field, continues:

> Such factors include, but are not limited to, the presence, features, and behavior of another person or persons (such as the interactive partners). These are the environmental triggers of the behavior, which then occurs without the necessity of the individual forming a conscious intention to behave that way, or even knowing, while acting, what the true purpose of the behavior is.

The scope of this extended conception of priming, and its potential relevance to the study of verbally-expressed, automatic personal interaction, forms the subject of this section.

6.3.1 Think old, walk slow: behavioural priming in action

Although the effects of linguistic priming have been studied for longer[6], there is now a substantial body of evidence on the impact of behavioural priming in a very wide range of contexts, from simple activities such as impression formation to extremely complex ones, such as the formulation, pursuit, and accomplishment of goals. In all cases, the processes involved fall well within the range of operations attributed to dual-process theory's System 1. As Bargh & Chartrand comment in a review of behavioural priming experiments:

> For many years now, researchers have studied two main types of mental processes both in isolation and in interaction with each other. The two types are known by a variety of names – conscious-unconscious, controlled-automatic, explicit-implicit, systematic-heuristic – but it's clear which one is "heavy" and which one is "light". To consciously and willfully [sic] regulate one's own behavior, evaluations, decisions, and emotional states requires considerable effort and is relatively slow. Moreoever, it appears to require a limited resource that is quickly used up, so conscious self-regulatory acts can only occur sparingly and for a short time. On the other hand, the nonconscious or automatic processes we've described here are unintended, effortless, very fast, and many of them can operate at any given time.
>
> (Bargh & Chartrand, 1999: 22)

The experiments they discuss include one of the best-known studies in the field, which focuses on the activation in participants' minds of a social stereotype – ELDERLY –and the observation of subsequent behaviour that fits this stereotype. As the experimental paradigm used is typical of much behavioural priming work, it is worth examining this study in detail.

The procedure is divided into two phases (see Bargh et al 1996). In the first, the ELDERLY prime – here, a verbal one (though visual and other types of prime are also

[6] The term, the concept, and its application to linguistic performance all date back to K. L. Lashley's 1951 paper on serial order in behaviour, which – under the heading '"Priming" of expressive units' – suggests that, before internal or overt enunciation takes place, an aggregate of word units is 'partially activated or readied' (Lashley, 1951: 119).

used in behavioural work) – is administered to the experimental group; in the second, the subsequent behaviour of primed subjects is observed and compared to that of an unprimed control group. The behavioural trait in question is slowness, a quality 'stereotypically associated with elderly people' (Bargh et al, 1996: 14). The prediction is that participants who had been exposed to the ELDERLY prime would walk more slowly than those who had not.

The priming manipulation itself, presented in the guise of a language proficiency test, is a scrambled-sentence task, in which participants are given 30 sets of five words, out of which they have to make grammatically correct four-word sentences. In the prime version, the key stimuli are all words that previous research had identified as components of the ELDERLY stereotype, such as *worried, Florida, grey, sentimental, bingo, ancient, forgetful, cautious* and *alone*; crucially, they do not include the word *slow*. In the control version, these stimuli are replaced with neutral words such as *thirsty, clean* and *private.* After participants[7] have finished this task, they are thanked, partially debriefed, and then directed to the elevator 'down the hall'. As they make their way along the corridor, their progress is surreptitiously observed and timed by a confederate with a stopwatch. Before they finally leave, the experimenter catches up with them to debrief them fully but, before doing so, asks them informally whether they had spotted the ELDERLY words in the scrambled sentences, and whether they thought these might have affected them in any way. Importantly, the answer to both questions is a uniform 'No'.

(This lack of explicit awareness of the prime's content is confirmed by a later experiment, in which the ELDERLY or neutral scrambled sentences are administered to 19 subjects who are afterwards asked about the sentences' content. Only one shows any awareness of a connection between the prime words and the ELDERLY stereotype, and even this one participant is unable to predict how this might have influenced their behaviour. 'Thus, it appears safe to conclude that the effect of the elderly priming manipulation on walking speed occurred nonconsciously' (*ibid*: 17).)

Analysis of the post-experiment walking speed of the primed participants shows a clear difference from the speed of control group: as predicted, the primed group walk

[7] Sixty in all, taking part in two identically-constructed experiments.

more slowly. These results, according to Bargh et al (*ibid*: 16), suggest that 'exposing individuals to a series of words linked to a particular stereotype influences behavior nonconsciously' and that, furthermore, the way this activation influences behavior depends on the content of the activated stereotype itself, not on the stimulus words actually presented. 'Because there were no allusions to time or speed in the stimulus materials, the results of the study suggest that the elderly priming stimuli activated the elderly stereotype in memory, and participants subsequently acted in ways consistent with that activated stereotype.'

6.3.2 Affecting behaviour with verbal and visual primes

Bargh et al's 1996 manipulation of the ELDERLY stereotype is only one of a substantial number of experimental studies of behavioural priming in action, and it falls in mid-range in terms of the complexity of the effects captured. In this case, activating a concept that embodies a whole cluster of stereotypical characteristics – from physical ones like *grey* to social and behavioural ones like *Florida* and *forgetful* – results in the acting-out of a behavioural trait associated with this cluster but conspicuously missing from the priming stimuli.

Other experiments described in the same account have a slightly different focus: one tests the 'stereotype to behavioural trait' hypothesis using a visual rather than a verbal prime; the other focuses on the priming of a single pair of behavioural traits – RUDE vs POLITE. This second study also uses a scrambled sentence task, with participants in the two experimental groups being exposed to terms associated with one or other trait: e.g *bold, disturb, aggravating, interrupt* for rudeness, *respect, appreciate, yield, discreetly* in the case of politeness. After completing the task, participants have to wait for further directions from the experimenter, who appears deeply involved in helping another 'participant' (a confederate); the longer participants are prepared to wait, the more polite their behaviour is taken to be. The priming hypothesis is again confirmed: 'Participants whose concepts of rudeness had been surreptitiously activated in the scrambled-sentence test subsequently were more likely to interrupt the conversation between the experimenter and confederate than were the other participants, and those whose concept of politeness had been activated were least likely to interrupt' (*ibid*: 11).

The third experiment described by Bargh et al (1996) tests the behavioural results of activating a stereotype, but uses a visual prime – subliminally-presented facial images – rather than a linguistic one. The stereotype in this case is that held by White Americans of African-Americans, and the experiment focuses on the activation in participants of the trait of hostility: a trait 'shown by pretesting (and much prior research) to participate in the stereotype' (*ibid*: 13)[8]. In this study, the 41 participants – all non-African-Americans – are asked to work on a tedious computerised task, during which pictures of African-American or Caucasian faces are subliminally displayed. Participants' reactions to a staged computer 'malfunction' are recorded, and their facial expressions assessed for indications of irritability, hostility, anger and uncooperativeness. The findings confirm both the hypothesis and earlier findings by Devine (1989): participants subliminally primed with images of African-American faces react with more hostility to the malfunction than those in the Caucasian group, and participants with low scores on racist attitudes toward African Americans are as likely to demonstrate hostility as those whose racism scores were high.

Overall, Bargh et al conclude, these three experiments demonstrate that activation of both single-trait constructs or stereotypes in one context result in behaviour consistent with them in a subsequent unrelated context – all without the participants' being aware of the influence of the priming events on their behaviour. 'The same priming techniques,' they continue, 'that have been shown in prior research to influence impression formation produce similar effects when the dependent measure is switched to social behaviour' (*ibid*: 21). The importance of this development is demonstrated when Bargh's work described above is compared, for instance, with

[8] Bargh et al are here referring to work by Devine (1989) on stereotypes and racial prejudice, which examined the effects of automatic stereotype activation on participants' subsequent evaluation of behaviour in a race-unspecified person. Both Devine's study and earlier work indicated that, for White Americans, aggression and hostility were traits associated with the stereotype of Black Americans, and Devine showed that was the case regardless of whether participants demonstrated high or low degrees of racial prejudice, as measured on the Modern Racism Scale. A second study by Devine used a lexical priming model (which omitted words directly related to hostility) to examine the extent to which activation of the Black American stereotype appeared to be associated with the ascription of hostility to a fictional, race-unspecified character who displayed behaviour that might, but not necessarily would, be interpreted as hostile (e.g. demanding money back in a shop), Again, the degree to which participants demonstrated a high or low extent of racial prejudice was also measured. The results of this study, Devine commented, suggested that automatic stereotype activation was 'equally strong and equally inescapable for high- and low-prejudice subjects' (Devine, 1989: 19)

one of the earliest studies of behavioural priming, Srull & Wyer's 1979 examination of priming's role in the way impressions are formed of other people: specifically, whether they are seen as hostile or kind. Here the automatic aspect of the process is less in focus: the prime, a scrambled sentence task, features items that directly indicate hostility or kindness (e.g. correctly completed sentences read 'Break his arm/leg' or 'Hug/kiss the boy'), while the experimental task also gives *explicit* directions to the participants – they are asked to form an impression of an ambiguously-presented character in a story and to do so on criteria that explicitly included a hostility/kindness continuum. Even at this early date, however, the key theoretical premise is present: 'Exposure to behavioural instances of a trait in one context may activate a schema associated with this trait, and the schema may then serve as a basis for interpreting subsequent behavioural information that is received in other contexts' (Srull & Wyer, 1979: 1661-2).

6.3.3 From simple actions to complex behaviour: priming for 'intelligence' and 'competitiveness'

Following on from Bargh et al's studies, a major development in behavioural priming research is provided by the work of Dijksterhuis & van Knippenberg, who address the issue of whether the unconscious, unintentional effect of an external trigger on behaviour is confined to 'relatively simple actions', such as walking slowly, or whether more complex behavioural patterns can be evoked in this way (Dijksterhuis & van Knippenberg, 1998: 2-3). They predict that priming test subjects with a social stereotype, e.g. PROFESSOR or FOOTBALL HOOLIGAN, will affect not merely behaviour, but actual task performance – here a paper-based Trivial Pursuit quiz – in line with traits associated with the stereotype (e.g. intelligence for PROFESSOR and stupidity for FOOTBALL HOOLIGAN). The PROFESSOR group is expected to perform better on this general knowledge task than control groups, which are primed either with the stereotype SECRETARY (which is 'supposedly unrelated to knowledgeability and intelligence') or not primed at all. The HOOLIGANS are expected to perform worse. (Only one control group – an unprimed one – was used here.)

The prime used for this series of experiments (unlike those used by Bargh et al) explicitly introduces the stereotype by asking participants to imagine a typical professor/secretary/hooligan and to list their appearance, lifestyle and behaviours. However, none of the participants shows any awareness of the critical factor in the process: the hidden link between the priming manipulation and the task demonstrating its effect.

The results in each case were as predicted. Participants primed with the PROFESSOR stereotype answer more questions correctly than either the SECRETARY or the non-primed groups,[9] while participants primed with the HOOLIGAN stereotype performed worse than the control group. Furthermore, a variant of the experiments, which compares the impact of stereotype primes vs single-trait ones (e.g. *stupid*) shows similar results: 'Participants primed with intelligence (either by priming professor or by priming intelligent) outperformed participants primed with stupidity (either by priming soccer hooligan or by priming stupid). As in the earlier experiments, priming affected behaviour. Participants behaved in line with the activated construct' (*ibid*: 17-18).

While behavioural primes used in experimental work are typically verbal or pictorial. other modes exist, and one – which can be labelled 'material primes' – has been shown to produce behaviours as complex and sustained as those in Dijksterhuis and van Knippenberg's Trivial Pursuit experiments. Inspired by potential connections between behavioural priming theory and the impact of material culture, Kay et al (2004) conducted a set of studies in which the primes are physical objects associated with business (i.e. strategic and competitive) contexts, such as briefcases, suits and boardroom tables. In one experiment, participants are merely primed with pictures of these before being asked to play the 'Ultimatum Game'. In another, however, the focus is again on a version of the ultimatum game, but the pictures are replaced with a real-world scenario in which the physical conditions of the experiment differ markedly. The experimenter in the control group produces a set of experimental materials from a backpack, the participants are each given a pencil with which to fill these in, and the completed documents are stored in a cardboard box. The prime

[9] Intriguingly, the SECRETARY group show an unexpected priming effect of its own: they complete the quiz much faster than the other participants, an effect that may be related to the content of the SECRETARY stereotype.

group, in contrast, are given their questionnaires out of a briefcase, and are given silver, wide-barrelled, 'executive-style' pens for filling them in. The questionnaires are then stored in a black leather portfolio.

The findings of the two experiments mirror each other. In the 'picture prime' study, only four out of the 12 primed participants opt for a co-operative, non-competitive solution to the game, whereas 10 out of 11 of the control group do. The 'physical prime' experiment produces the same result: participants acted in a 'more self-interested and competitive manner' following exposure to business environment primes, even though these are now embedded in the actual environment. 'This result,' Kay et al conclude, 'attests to the ecological validity and/or potential "real-world" significance of implicit priming effects, and further suggests the role that such priming can play in promoting counter-productive behavioural choices in organizational settings' (Kay et al, 2004: 91).

6.3.4 Competitive or cooperative? Behavioural primes, verbal responses

The method which Kay et al use to demonstrate the validity of material priming – the automatic connection between objects linked to business situations and concepts describing those situations – is also of interest here, as it brings my discussion back to the topic of specifically verbal interaction. The 67 participants in this experiment are exposed to descriptions and pictures of either business-related or neutral stimuli, such as a fountain pen, man's suit, boardroom table (business context), or a kite, whale, toothbrush (neutral context). The two groups then perform a standard word-completion task in which a set of word fragments contains a subset that can be completed in either a competitive, 'businesslike' way or otherwise, e.g. *–(p)ower* (vs *(l)ower, (m)ower* etc); *(f)ight* (vs *(l)ight, (t)ight*); *bea(t)* (vs *bea(n), bea(d)*). A particularly important fragment is *c--p---tive,* which can be completed to read either *competitive* or *cooperative*.

The results show that exposure to the business-related objects leads to the completion of significantly more word fragments in 'competitive' mode, with 71% of the business-primed participants opting for the *competitive* completion choice (as against 42% of those in the control group). 'These findings,' Kay et al conclude, 'lend

support to the argument that material objects with a particular social and normative relevance within a given culture – and indeed pictorial representations of such objects – can influence the cognitive accessibility of the broader knowledge structures linked to those representations' (*ibid*: 86). Importantly for my own argument, they also lend support to the idea that primed behaviour, however stimulated, can express itself not just in terms of general action but also as a specifically *verbal* response – a response that describe*s* the behaviour concerned. This could be an indication that the priming relationship between actions and words is bi-directional: not only can verbal primes (as we have seen) elicit behaviour associated with the primed concept, but the behavioural concepts activated by priming can elicit words associated with these concepts. As I hope to show, there may be further indications of this bi-directionality, in which case behavioural priming could prove to be as powerful an automatic stimulus in speech production as linguistic priming, and one that operates at the Speaker's message level.

6.4 What mechanisms support behavioural priming?

The experiments described above cover a wide spectrum of complexity and an even wider one of activities, from simply walking down a corridor to financial decision-making and – ultimately – selecting vocabulary that expresses the (primed) competitive ethos. What are the mechanisms that produce such variety?

Several explanations have been put forward for the phenomenon of behavioural priming, ranging from simple associative linkages between situations and behavioural responses (an account advanced and then rejected by Bargh et al, 1996; see below) to DeMarree et al's 'active-self account' (2005), which suggests that the effect of a behavioural prime can be absorbed into a person's *self*-concept, which then guides behaviour in a manner consistent with this (temporarily altered) self. However, the explanation currently most generally accepted involves a combination of ideomotor theory, which proposes that 'merely thinking about a behavior is sufficient to produce that behavior' (De Marree et al, *ibid*: 658), and an automatic link between perception and behaviour, as suggested by Carver et al. This argument rests on the close parallel between knowing how a certain behavioural trait presents

itself and knowing what it means to behave in the trait-like manner. In their investigations of such a link, Carver et al suggest that behaviour-specifying information contributes to schematic knowledge structures in much the same way as does purely perceptual information. 'We have assumed that many interpretive or conceptual schemas (though perhaps not all) either incorporate, or are closely associated with, information that specifies behavior or qualities of behavior… If so, activating the conceptual schema should also render the behavioral information more accessible. This in turn would make the behavioral information more likely to be incorporated into ongoing action' (Carver et al, 1983: 407). And this assumption is borne out by Carver et al's own experiments, which examine the effects both of direct physical perception (e.g. of hostile behaviour) and of the activation of a conceptual schema (scrambled sentences containing 'hostile' words). In both cases, there is a priming effect. 'Observing a model do a behavior seems to enhance one's own access to the mental records specifying the behavior, thereby making it more likely to occur in one's own action' (*ibid*: 414), and the same applies when the observation of behaviour is conceptual rather than directly physical.

Pointing to the consistency between this finding and their own experimental results, Bargh et al (1996: 22) amplify Carver's comments: 'If one has just perceived another person acting in a generous or an aggressive way… one's behavioural schema for generosity or aggression is activated and accessible, and so one is more likely to behave that way oneself in subsequent situations for which generosity or aggression is a relevant response.' Importantly, they stress the fact that this activation appears to take place *directly*, via the priming mechanism, rather than via the experiencer's conscious awareness of the social situation. Thus, primed participants in the RUDE/POLITE experiment behaved rudely even though – as a post-experimental survey showed – they had not perceived the experimenter as rude. 'The fact that the behavioural measure showed quite strong effects of the priming manipulation, whereas the effect on the judgment measure was nonexistent, argues against [this] alternative interpretation of our findings. It was not the case that the priming manipulation affected consciously made judgments about the experimenter, which then determined behavioural responses to him. The results instead point to a direct effect on behavior that is not mediated by conscious perceptual or judgmental processes' (*ibid*: 12).

Although the perception-behaviour link is generally seen as the mechanism underpinning behavioural priming, the addition of goal priming to the picture both complicates and extends it. Goal priming – evoking conceptual representations with an element of intentionality – seems on the face of it a process that involves System 2 cognition rather than the automatic responses of standard priming; as Bargh et al themselves point out (2001: 2), most theories of goal pursuit emphasize conscious choice. They suggest, however, that these representations can in fact become activated 'without an act of conscious will, such that subsequent behaviour is then guided by these goals within the situational context faced by the individual… [it is] possible that goal representations do not need always to be put into motion by an act of conscious choice.' [10] This hypothesis, in their view may 'take some of the mystery' out of behavioural-priming effects:

> The explanation of [such] findings in terms of the nonconscious activation of single behavioural responses as a consequence of related perceptual activity is not entirely free of difficulty – that is, what is the "behaviour" that is activated by the professor stereotype that causes superior test performance? However, the present findings of nonconscious goal pursuit… provide a possible additional mechanism for such effects. For example, the professor stereotype might include goals typically ascribed to professors, such as a desire to do intellectual work and the determination to solve problems and find correct solutions, and these activated goals might then guide performance over time on the experimental task.
>
> (Bargh et al, 2001: 19)

Another perspective on behavioural priming, the 'social functional account', takes the joint activation of dynamic and perceptual concepts suggested by Bargh et al a stage further. As summarised by De Marree et al (*ibid*: 658), this starts from the idea, originating in evolutionary research, that people may adjust their attitudes to fit in with their social environment. As a result, activation of stereotypes should produce stereotype-consistent attitudes as well as behaviour – an effect that has already been

[10] To support this claim, they present a sequence of experiments that build on the conceptual complexity studied in Dijksterhuis & van Knippenberg's work with the PROFESSOR stereotype by focusing on the motivation leading to a primed outcome. In one of these, a word-search puzzle is used to prime participants with the goal embodied in the concept HIGH PERFORMANCE. (Words appearing in the puzzle include *win, compete, succeed, strive,* and *achieve*.) Task performance in further puzzles is 'substantially better' than that achieved by a control group, and a reliable priming effect is also shown in experiments that test the readiness of the primed group to persist in a task, and to resume an interrupted activity.

observed. It was found, for instance, that participants primed with an ELDERLY stereotype endorsed more conservative attitudes, and those primed with a SKINHEAD stereotype endorsed more racist attitudes, compared with control participants. 'The social functional account holds that activating social constructs causes people to want to fit in with their social environment. Any expressed opinions, judgments, and behaviors that would help the person fit in could therefore be modified.' It is worth asking at this point what verbal form these modified opinions and judgments might take: faced with (or thinking of) an elderly person, does a Speaker use words more characteristic of an elderly person's vocabulary than his own? Does he 'talk old' as well as 'think old'?

A further perspective also has to be considered. When priming – behavioural or otherwise – takes place, what is happening at the neural level, and what light does this throw on the issues raised above? Interestingly, findings here appear to confirm the essential automaticity of the prime-to-outcome process. In a review of neurological priming studies, Schacter & Buckner (1998:191-2) report that findings converge to suggest 'one neural correlate of priming: repeating items during performance of the same task, or even during performance of different tasks, can lead to decreases in the amount of activation present in specific brain areas.' Cabeza & Nyberg's review (2000: 29) also points to an association of priming with a decrease in neural activity, in this case in the left prefrontal area. They also point out that this effect can be observed whether the stimulus takes the form of words or pictures.

The implications of these findings for dual process theory are indicated in the 2007 paper by Schacter et al quoted above in Chapter 6.3: 'Priming is a nonconscious form of memory…. Cognitive and neuropsychological evidence indicates that priming reflects the operation of implicit or nonconscious processes that can be dissociated from those that support explicit or conscious recollection of past experiences' (*op cit*: 171). Defined like this, priming is unarguably a System-1 process. And Chein & Schneider suggest further possible implications, by focusing on the predictions of dual processing theory in the area of learning. In a meta-analysis of neuroimaging studies of practice-related change, they show that practice – i.e. repetition to the point where processing becomes automatic – leads to a reduction in extent and activity across a distributed cortical network, including prefrontal, anterior cingulate,

parietal and occipito-temporal regions. In further work, they show that this same network is engaged during early associative learning of both verbal and nonverbal material, but disengaged once the associations in question have been well-learned; they interpret these practice-related changes as 'reflections of the waning contributions of a domain-general control system as a task becomes well learned, an interpretation consistent with the central tenets of the dual-processing framework' (Chein & Schneider, 2005: 621). This is also consistent with Smith & DeCoster's description of the process by which information is gradually transferred from the fast-learning memory system to the slow (associative) one: 'The two memory systems interact in several ways… Most important is the process of *consolidation* [authors' emphasis], by which newly formed memory is transferred by repeated presentations from the fast-binding to the slow-learning system.' (Smith & DeCoster, 2000: 110).

6.5 Behavioural primes, linguistic outputs: can the two traditions converge to shed light on the message level?

In some ways, the account of behavioural priming given above is complementary to the account of linguistic priming given in Chapter 2 (see, especially, Chapter 2.4.4). Both focus on an automatic process – a 'spreading-activation retrieval mechanism' (Dell, 1986: 283) that is assumed to handle representations in the same way, whatever the differences between them. Both draw heavily on the work of two particular research groupings, one centred on Pickering & Garrod's 'mechanistic' theory of dialogue, and the other on the work of Bargh and others in social psychology. And they can be seen as offering complementary answers – explicit in one case, implicit in the other – to my question 'Why do we say the things we do?'

It is now time to consider how far the two research traditions can *jointly* take my inquiry further.

Importantly, both sides of the linguistic/behavioural divide acknowledge the importance of the other's contribution. Pickering & Garrod, for example, suggest that their linguistic model could serve as 'the basis for predominantly automatic

accounts of social interaction more generally' (*ibid*: 188). Citing the work of Bargh, Dijksterhuis and others, they point out that the apparent strength of the automatic perception-behaviour link suggests that the great majority of social acts do not involve decision-making. 'Our contention is somewhat related, in that we argue that the process of alignment [via linguistic priming] allows the reuse of representations that are constructed during comprehension in a way that removes the need to make complex decisions about how to represent the mental state of the interlocutor. Of course, there are still some conscious decisions about what one wants to talk about…'

Meanwhile, on the behavioural priming side, there is an awareness that there is more to research here than simply finding out what can be primed, substantial though these discoveries have been. As Bargh (2006: 148) comments, 'To some… these impressive empirical demonstrations have become an embarrassment of riches; our empirical knowledge has outstripped our ability to understand and conceptualise just what is going on here – what exactly is being primed, and how are these impressive effects produced?' Answers to a range of 'second-generation' questions[11] are now needed and, Bargh suggests, models of how ideas and thoughts are expressed – language models – may offer insights into how 'other forms of behaviour' are expressed. 'These models have already tackled the problem of how parallel processes (thoughts, ideas, intentions) are transformed into serial speech acts…And if theorists such as Clark, Lakoff and Johnson, and others are on the right track as far as language being in the service of behaviour (a sub-goal if you will), then it may be that serial, real-time behavior *in general* [author's emphasis] follows the same principles – even, perhaps, the same mechanism' (*ibid*: 161).

He supports his suggestion by pointing out that many of the key concepts in social cognition – particularly priming and automaticity – themselves originated in psycholinguistic theory and research. Lashley's 1951 introduction of the concept of priming (see above, Note 5) is only one of these; others include the concept of automatic activation (derived from Cherry's 1953 observation of the automatic attention pre-empted by the sound of one's own name in a hitherto-disregarded

[11] Such as whether behavioural primes are controllable, how do they operate in parallel, which ones are more likely to occur in natural, stimulus-rich environments, and which prime 'wins' if conflicting responses are activated.

conversation), and work on word-associations and the spreading-activation phenomenon in the early 1970s. In all of these, according to Bargh, the study of language comprehension and production has provided social cognition with highly useful models and, given this record, 'it might be the case that the underlying mechanisms of language production and of social behavior *are one and the same*' (*ibid*: 162).

As indicated by the emphasis that Bargh puts on his own words here, this is a statement of major importance – not least because it shows how just how close the linguistic and social cognitive research traditions are on the priming issue. If language use *is* action, he adds, then how language is produced 'may well be how behavior in general is produced…We do not usually formulate sentences in our mind prior to saying them. Rather, we may have some vague ideas about what we want to say, the ideas or points we want to make, and these guide what we say, but the ways in which we express these intentions, out loud, are opaque to us.'

I would like to propose that the reverse of Bargh's suggestion is also true: the way behaviour in general is produced is also the way language is produced. Or, to put it another way, language is just one form of behaviour amongst others and it can therefore equally readily influence – that is, prime – social behaviour and *be influenced by it*. Moreover, as much of the work on behavioural priming shows, the priming stimulus need not be a directly-perceived piece of behaviour, but may be a conceptual representation of this behaviour, and language is highly effective in evoking/eliciting such representations, whether in the form of explicit primes ('Break his arm', for HOSTILITY) or indirect ones ('Florida… grey… cautious' for ELDERLY). And here I return to the issue first raised at the end of Chapter 6.3.4 above: if language use can activate a conceptual representation that nonconsciously affects behaviour, can behaviour or the representation of it nonconsciously affect language use? Put another way, is the link between language use and behaviour bi-directional – can a behavioural prime produce a linguistic output? For instance, could the word 'friendly' produce both unconconsciously-adopted friendly behaviour *and* the message-level decision to express friendliness, with appropriate (and perhaps formulaic) semantic, lexical and other choices?

It is this possibility that leads me to suggest that the work of Bargh and others can offer some answers – implicit though they are as yet – to the questions still surrounding the message level of speech production, just as Pickering & Garrod's interaction model helps explain the effortlessness and speed of speech production at lower levels. This is also implied in Bargh's own answer to the question he raises in the title of his 2006 paper ('What have we been priming all these years?'). Perhaps, he says, what we have been priming 'is a role, a conceptual structure that contains not only the nuts and bolts of how to act within that persona, but, at essence, the *perspective* [author's emphasis] a person in that role would have on the world – the purposes and goals and values that person… would have' (*ibid*: 155). Add this perspective to the 'nuts and bolts' – which include the means of expressing this perspective – and one of the products could be, in Pickering & Garrod's words, 'decisions about what one wants to talk about'. Since, by hypothesis, these decisions are an effect of behavioural priming, they may not result from conscious choice, but from processes functioning below the level of consciousness: System-1 processes, heavily dependent on slowly-accumulated and quickly retrieved elements from a Speaker's long-term knowledge, and accompanied – as we have seen – by marked reductions in effort, right down to the neural level.

The implications of this suggestion are considerable, for philosophy and ethics as well as social cognition and pragmatics. (Thus, if Conceptual-level speech decisions are affected by external influences without a Speaker's knowledge, then who – to repeat the question first raised in Chapter 2.1.1 – is doing the decision-making?[12]) They certainly help to explain the speed and flexibility with which a Speaker can produce speech that is appropriate and relevant to a given context. Very importantly, they also help to explain how a Speaker can achieve his essential goal of producing speech that will retain the Addressee's attention, because the Addressee – whether a live interlocutor, a mental representation of him, or a mixture of the two – *is himself part of the prime*.

[12] Though easy to raise, this is not a question that is easy to answer. As Uleman says in his introduction to *The New Unconscious,* 'It is not yet clear… whether the new unconscious must challenge our fundamental notions of free will and social responsibility, or merit and blame, and of the self and religious beliefs' (Uleman, 2005: 15)

At the level of simple social interaction, this outcome can be seen as a manifestation of the self-fulfilling prophecy effect. As Bargh et al themselves point out (1996), empathic social relations can be created in this way, with the priming mechanisms nonconsciously producing behavioural and emotional responses corresponding to those perceived in another person: 'The perception-behaviour link may be an important ingredient in the "glue" that binds two (or more) interaction partners, keeps them on the same wavelength, and helps to bring each partner a sense of validation by others of their experience' (*op cit*: 26). However, if the dimension of verbal interaction can be added, behaviourally-primed effects become more precise, more closely attuned to what will seem relevant to an Addressee, because they have originated at least partly from the Addressee, or a representation of him. At its most stereotypical level, this sort of communicative success can be seen in the detailed directions to Jordan Marsh given by Bostonians to the 'out-of-towner' with a non-local accent (Kingsbury, quoted by Krauss & Pardo, 2004; see Chapter 2.3.3). Another possible instance – that relies on the perception of age-group rather than locality for its prime – is the choice of joke in Example 9, formulaic though it may be ('They don't like it, do they?'): behind the automaticity involved in a formula lies *another* level of automaticity, the selection of what formula to use. And in both cases, the triggers for these choices are those described by Bargh above (cf Chapter 6.3): the 'presence, features, and behavior' of the interactive partners.

It seems possible that this hypothesis, if it can be substantiated, could approach from another angle the phenomenon that Horton & Gerrig (2005a: 10) term memory 'resonance', by which other people function as highly salient cues to activate information with which they are associated (see Chapter 2.4.5). Taking this further, Garrod & Pickering (2007: 447) suggest that routinized referential expressions become associated with the person who used them, resulting in 'partner-specificity effects during comprehension and production.' Bringing these suggestions together, is it possible to make a case for a special type of mental representation: an Interlocutor Profile ('I-profile' for short), or stock of knowledge about specific Addressees (or stereotypical versions of them) that includes both their behavioural and speech patterns, and their history in terms of both past and current dialogues? Its key characteristic is that it might be available for use in both System 2 and – crucially – System 1 operating mode: 'The information required can be accessed in

one of two ways – consciously and reflectively [as in formal social settings], or automatically and at speeds compatible with the production and interpretation of normal fluent speech,' (Pollard, 2005b: 32).

However, this is a highly speculative suggestion. There are other major questions to be answered first, among which the overall nature of the representations activated in joint behavioural-linguistic priming is only one. There are also issues of detail, such as the relationships between the language used to evoke a behavioural concept, the behavioural concept itself, and the language that this concept elicits. Furthermore, when the prime is a real-life Addressee, what exactly does the priming consist of: is it the representation of your interlocutor that you have in your head, or the real interlocutor in front of you (who is constantly updating your mental representation)? Put another way, how does priming interact with protagonist-dominated situation model theory, as outlined in Chapter 2.2.2?

Perhaps most fundamental of all, what evidence is there that, just as linguistic representations can influence social behaviour, behavioural representations can influence linguistic processing? Clearly, finding such evidence is a top priority, and one suggestive experiment has already been described in Chapter 6.3.4 (Kay et al's combination of behavioural priming with a word-completion task). Some further evidence is considered in the following two sub-sections: the first draws on neuroscientific work showing connections between words expressing action and action itself, and the second returns to the field of behavioural priming itself.

6.5.1 The two-way prime: actions and words

A possible starting-point here is the work on mirror neurons carried out in the 1990s by Rizzolatti, Iacoboni and others (see Chapter 5.4), which suggests that a 'direct matching' mechanism – similar to that found in monkeys – exists in humans that maps a visual or kinematic description of the action observed onto an internal motor representation of the same action. Evidence that Broca's area (BA 44) is amongst those activated during physical imitation has broadened the potential scope of this theory: as Rizzolatti & Arbib suggest (1998), it seems to indicate a long-standing neural connection between action and communication. Approaching the action-communication link from a different perspective, other researchers have shown that

processing action words activates areas involved in the processing of movement, with the result that there is now a 'large body of psychological and neuroimaging experiments that have interpreted their findings in favor of a functional equivalence between action generation, action simulation, action verbalisation, and perception of action' (Grèzes & Decety, 2001: 1). For example, Martin et al (1995: 102), in a PET study of neural activation involved in the generation of colour and action words, show that generating the latter activates a brain region (the middle temporal gyrus) just anterior to the area involved in the perception of motion. Their findings, they suggest, indicate that 'object knowledge is organized as a distributed system in which the attributes of an object are stored close to the regions of the cortex that mediate perception of those attributes.' In another PET study that compares the neural areas activated by reading nouns and verbs, Perani et al (1999) confirm Martin et al's findings on action words and also – uniquely for the verbs in the study – highlight the activation of the left superior parietal lobule, an area associated with a range of functions that include spatial orientation.

While this research explores the representational link that leads from language to physical action, other work takes action itself as a starting-point, and Jeannerod (1999) suggests that the link might also run in the other direction. In a review of research on action representation (and in particular on comparing neural responses to 'meaningful' and 'meaningless' action), he notes the repeated involvement of part of Broca's area and continues: 'Activation of [this area] during object recognition raises the problem of the involvement of language during the observation of meaningful actions. Indeed, this area is also activated in several situations related to language for action, such as generation of action words… It is a possibility that during observation the subjects *automatically associate the action with action verbs* [my emphasis].' (Jeannerod, 1999: 12)

The mental simulation of actions also seems to have a neural association with verbalisation, as Grèzes & Decety's meta-analysis shows. The studies covered in this review show simulation leading to activation in both parts of Broca's (Brodmann Areas 44 and 45), and in the adjoining BA 46. These results, according to Grèzes & Decety, 'support the view that mental simulation engages implicit verbal mediation' (*ibid*: 12).

Overall, Grèzes & Decety urge some caution in interpreting the results of verb-action research, partly because there is rather little overlap between regions activated by silent verbalisation and the other target processes studied, and partly because of the impact that verbalisation itself has on neural analysis: 'Language dominates other cognitive processing and … it is difficult to clearly distinguish what belongs to what' (*ibid*: 13). However, there is evidence from other areas of investigation that may shed further light on the mechanisms supporting an association between action and language (as opposed to between language and action). A series of experiments by Matlock (2004) considers the simulation-verb link from the perspectives of mental modelling and language use – in particular, of loose or figurative language use – and the outcome appears to confirm the connection noted in the neurological studies between imagined actions and the language describing them.

The studies focus on 'fictive motion' (FM) verbs, which describe no explicit action or state change, as in 'The road *goes* through the desert' or 'The fence *runs* along the property line', and the aim is to assess whether representing the implicit type of motion conveyed by such verbs involves representing the physical action concerned. Thus, 'do people trying to understand these sentences mentally simulate motion?' (Matlock, 2004: 1389). Mental simulation research, she comments, suggests that people construct mental models that resemble physical space, and represent movement within these models in a way that to some extent resembles the perception of physical movement.

> Given this, simulation might also be involved in fictive motion processing… For instance, if people read a story about travel through a spatial region, such as a desert, they ought to construct a model with properties consistent with what they know about deserts… If [they] read a fictive motion sentence at the end of the story, such as *The trail goes through the desert*, their processing ought to be affected by the model they constructed, including motion in that model. Reading about slow motion should result in longer processing times than reading about fast motion. The same should occur when people are reading about long distance (vs short distance) and when reading about a difficult terrain (vs an easy terrain).
>
> (Matlock, 2004: 1391).

Subjects in Matlock's experiments read short stories along the lines described above which, in the case of the slow vs fast scenario, features exercise taken by an elderly man ('Bad heart… slowly walks… totally exhausted') and an athlete ('Excellent shape… sprints… totally exhilarated'). Both end in the fictive-motion sentence, 'The path follows the creek'; subjects are asked whether this sentence relates to the story, and their response times are recorded. If, Matlock argues, fictive motion processing involves simulation, there should be differences in how the motion was represented in the story (slow or fast) and these differences should be reflected in the reading times. In other words, qualities associated with the representation of certain actions should feed through into the language process – here considered in terms of comprehension.

The studies all produce the same outcome: latencies were shorter after the subjects had read about fast travel, short distances, and easy terrains, longer after reading about the reverse. Overall, the data suggests that processing fictive motion verbs does indeed involve simulation of actual motion. Matlock rules out the possibility that linguistic priming is involved – on the ground that in a control study without the FM sentences, slow/fast differences do not appear – so it appears that it is the presence of the FM verb alone that has the effect on language processing. Is this a subtle (and high-speed) form of behavioural priming at work? Further work would here be interesting.

6.5.2 The two-way prime: think old, think forgetful

Another approach to the problem of the two-way behavioural prime can be found in the behavioural priming literature itself and, in particular, in the substantial work on the priming power of stereotypes. The focus of interest here, as in Bargh's 'Think old, walk slow' experiment described above, is the stereotype of the elderly person: worried, grey, cautious, alone, slow, and *forgetful*. As Dijksterhuis et al (2000: 535) comment, this last stereotypical trait – impaired memory function – has been shown by a vast body of literature to be 'at least to some degree accurate,' since the elderly score consistently lower than others on memory tasks such as free recall tasks. So what happens when younger people are primed with the ELDERLY behavioural stereotype? Does the effect of the behavioural prime modify their cognitive

capacities, including their linguistic capacities, as clearly as it modifies physical behaviour such as walking speed?

Dijksterhuis et al show that such modification can indeed take place, and to a degree that depends on the amount of real-world contact with the elderly that the priming subjects have had. In an experiment with 75 undergraduates, participants are given a word-recognition task in which they have to distinguish between existing and nonsense words. Every word is subliminally preceded by another: neutral for the controls, and an ELDERLY prime (e.g. *old*, *gray*, *bingo*, though not *forgetfulness*) in the case of the experimental group. Afterwards, all subjects complete a memory task, involving the recall of as many of the existing words from the word-recognition task as possible. Before the results are checked, however, both the control and the priming group are divided into two further groups, based on questionnaire answers filled in earlier: one group consists of those who have spent some time amongst the elderly and the other of those who have not.

The findings confirm Dijksterhuis et al's prediction that activating the ELDERLY stereotype leads to impaired memory performance, to an extent determined by the amount of real-world contact with the elderly that the participants have had: the mean number of words remembered by the prime/high contact group is 3.7, as against 5.7 for the prime/low contact group (and also 5.7 for the no prime/much contact group).

A second experiment confirms Dijksterhuis et al's further hypothesis, that the strength of the post-priming effects is predicted by the strength of the *association* formed by the high-contact group between the ELDERLY stereotype and forgetfulness. And this association, in turn, 'has behavioural consequences once the social category of the elderly is activated: One starts to display memory impairment oneself. The conclusion is that more past contact leads to behavioural (or memorial) changes during present contact' (*ibid*: 540). As will be obvious, the implications of this connection between amount of past exposure and primed present behaviour could be extremely important in the overall context of my own inquiry, and especially in connection with my hypothesized Interlocutor Profile (see Chapter 6.5, above). If contact with the elderly means that you readily adopt ELDERLY-stereotypical speech patterns, you are automatically responding to (being primed by) an I-profile of an

elderly person – and, whatever the detriment to your own normal speech processes, you may as a result find yourself more attuned to an elderly interlocutor. But of more immediate importance here is the question of what types of representation are affected by such exposure. The area chosen by Dijksterhuis et al for the study of post-prime memory problems is language and, although this is not the focus of their two studies, one effect is to show clearly that behavioural priming can impact language processing, albeit in a negative direction. This raises the further question: what aspects of language processing are affected?

Another pair of ELDERLY prime experiments specifically addresses this question and, at the same time, contributes further insight into the issue raised in Chapter 6.4 about how priming works. According to Catmur et al (in preparation), the relationship between stereotype priming and cognition is poorly understood, and a particular cause of confusion is the nature of the mechanisms involved. Does the activation of a stereotype have a direct effect on cognitive processes, or might it affect cognition only indirectly, by directly altering mood or motivation? 'On many tasks,' according to Catmuret al, 'these factors are hard to disambiguate: for example, effects of motivation and mood can influence cognition via effects on attention or strategy selection.'

The two studies by Catmur et al (on which I was part of the research team) address this problem using a combination of neural activity observation and the 'Subsequent Memory' paradigm. This paradigm features scrambled-sentence primes designed to activate the ELDERLY stereotype, followed by a lexical classification task in which participants are presented with words describing living or inanimate entities and asked to classify them accordingly. The classification task is followed by a memory test designed to assess how many words the subjects remembered. But, in order to rule out motivational effects such as anxiety, the test in both experiments is a 'surprise' one, which we present to the participants only after the other procedures have taken place. They are therefore relaxed during the priming and encoding (language input/lexical classification) processes – a crucial factor in the first experiment, where both the prime and the lexical classification (LC) task are administered to participants in an fMRI scanner, and where the neural activation accompanying the LC task forms a key part of the findings.

A second important difference between the two experiments concerns the memory test itself. In the fMRI study, this is a recognition test in which participants are exposed to a mix of new words and those they have seen in the decision task, and asked to state how certain/uncertain they are that they have seen them before. In the second experiment, which features the same priming and input procedures but administers them via computer rather than in the scanner, participants complete an uncued recall test, in which they write down all the words they remember from the encoding phase.

Although conditions leading to the recall test are more accommodating than the recognition one – fewer words are presented, each one is presented twice, and the test follows immediately on the encoding phase (rather than 10 minutes later), the results of the two experiments are in marked contrast with each other. As predicted by the work of Dijksterhuis et al (2000), the primed group in the recall experiment remember significantly fewer words than those in the control group. In the recognition experiment, however, there are no significant differences between the two groups – again as predicted by further work by Dijksterhuis and others[13]. As Dijksterhuis et al note (2000: 541), their study showed that 'elderly-priming negatively affected recall, but not recognition. Recognition performance was the same among primed participants and no-prime control participants.'

According to Dijksterhuis et al, these findings 'clearly show' that it is retrieval rather than encoding that is being impaired. However, this conclusion contrasts with a further outcome of Catmur et al's study. Analysis of neural activity captured during the encoding phase of our fMRI experiment, when correlated with the memory test results, shows that the experimental group experience reduced *encoding*-related activity, in comparison with the control group.[14]

Overall, these findings give some important insights into the way that behavioural priming and language processes interact. At a fundamental level, the exclusion of anxiety-producing factors from the experiments suggests that the impact of

[13] Dijksterhuis, Bargh & Miedema (2000), cited by Dijksterhuis et al (2000) in their paper discussed here.

[14] The areas where these differences are observed are the bilateral inferior temporal gyrus and inferior frontal gyrus – regions that have been reliably shown to be involved in encoding.

stereotype priming on cognitive operations is a direct one: as participants do not know that a memory test is coming, mood and motivation are not involved in the neural activities observed while they do an undemanding LC task.
The findings themselves show that encoding – language intake – has a crucial effect on language output and that, when influenced by a stereotype prime that includes an 'impaired memory' trait, it is the encoding process, rather than actual retrieval, that can be affected by priming (or by real-world contact). Moreover, the outcome affected is word recall rather than word recognition, and Catmur et al suggest two possible explanations for this. One is that priming results in a weaker encoding process; consequently, if stronger memory representations are quired for recall rather than for recognition, the recall process should be affected more strongly by the priming manipulation. An alternative account could be offered by dual-process theories of recognition proposing that recall depends on recollection, while recognition depends on recollection and familiarity. 'By such an account, priming may have selectively impacted upon the encoding processes associated with recollection but not familiarity.'

Conclusions such as those outlined in this section point the way to further work on the relationships between priming and memory, possibly involving primes that might achieve the reverse effect to memory impairment: for example, can encoding activity be *enhanced* by priming with the stereotype YOUNG or even – following the Dijksterhuis example – PROFESSOR? The main thrust, however, needs to be on continuing investigation of how the System-1, automatic phenomenon of behavioural priming affects speech decision-making up to and including the message level.

6.6 Summary and conclusions

Using the interlocutor feedback loop and its social infrastructure as a background, in this chapter I finally address the question with which this thesis opens: why do people say the things they do? And, in particular, how far do automatic processes of the type envisaged in dual-process theory's System 1 enable them to make decisions about message content that maintain the coherence and relevance of dialogue, while complying with the heavy time and attentional pressures involved?

In an attempt at a partial answer, I approach the issue from two perspectives: linguistics (speech routines) and social psychology (behavioural priming). In discussing speech routines and routinisation, I return to Pickering & Garrod's point that such 'short-circuits' in speech decision-making are both widespread and highly economical in terms of speaker's effort. Nor do such routines necessarily lack audience-tailored flexibility and creativity, as demonstrated by the speech formulae of master auctioneers collected by Kuiper. Using examples from my own collection of naturally-occurring exchanges, I go on to suggest that formulaic speech, combined with such audience-tailored creativity, can produce a wider, more nuanced range of effortlessly-retrieved routines than has been previously supposed. I also underline the applicability to speech routines of Smith & DeCoster's dual-process model of memory (cf Chapter 5.3.4): in keeping with their description of associative (i.e. System-1 style) memory, formulae are learned slowly and retrieved fast, and their ease of retrieval makes them natural candidates for use when the Speaker is under time or other pressures.

While speech routines and their relationship to System-1 processes offer obvious support for a theory of automatic message generation, the connection between message generation and behavioural priming is less clear. Indeed, as I point out, work on possible convergences of the linguistic and behavioural priming traditions is still ongoing but, as outlined in Chapter 6.3, the evidence for behavioural priming itself is both extensive and robust. It therefore seems arguable that elements of an Addressee's past or present social behaviour could act as a prime – an automatic trigger – for a Speaker's message-generation processes. Because it originated partly with the Addressee himself, the resulting verbal output should dovetail closely with what the Addressee is cognitively and affectively equipped to process, and should therefore be relevant in terms of both effort and effect.

In the final section of the chapter, I consider an important issue raised by this proposal: the availability or otherwise of the representations involved to bi-directional processing. If, is already well established, verbal representations can be act as primes to influence social behaviour, can the influence run in the opposite direction, from social behaviour to verbal output? What evidence is there that a concept activated by a behavioural prime may go on to be verbally expressed?

Research cited in the chapter to support this suggestion includes studies both from the behavioural priming field and from neural and other explorations of associations between language and physical action. In the behavioural priming field itself, recent work – in some of which I took part – indicates that a behavioural prime such as the ELDERLY stereotype can affect word recall (though not recognition), and that the process it affects is not language retrieval but language intake: confirmation that behavioural representations can indeed influence linguistic processing, together with a possible indication of some of the mechanisms involved.

Taken together, the results of work on speech routines and on behavioural priming appear to offer good grounds for continuing investigation into the automaticity of the message-generation process, both in order to shed light on the process itself and to help explain how a Speaker may 'often enough' achieve optimal relevance to his Addressee.

Chapter 7

Speech production, dual-process theory, and the attentive Addressee: conclusions

My inquiry opened with two questions on speech production, about the nature of the conceptualization process and its ability to form messages that tend to be relevant to the Addressee. It ends with a single conclusion that addresses both issues. Basing my claim on evidence drawn from a wide range of disciplines, I suggest that the Addressee – hitherto somewhat neglected in many theoretical approaches – plays a decisive part in the Speaker's message generation process, and that this influence is exerted automatically, below a Speaker's level of consciousness. I also argue that the key factor involved is the degree of attentiveness shown by the Addressee to what the Speaker has to say, which can be seen as indicating the degree of relevance to the Addressee that the Speaker has achieved. I show that lack of attentiveness, in particular, appears to damage speech production at the conceptualization level, from which it follows that Speaker and Addressee are linked in a feedback loop. Unless a Speaker achieves and continually maintains relevance to an Addressee, the Addressee's attentiveness will falter, the Speaker's performance will consequently be impaired, and the communication process itself will be endangered.

Achieving relevance to an Addressee is therefore not an optional extra for a Speaker, but a crucial prerequisite. However, it is a formidably difficult one to meet if, as envisaged in Levelt's speech production model, the decisions made by the Conceptualizer are largely under executive control, and have to be taken as a result of conscious deliberation. Normal, everyday speech is – in addition to being optimally relevant – self-evidently capable of keeping up with the heavy demands that time, context and the Speaker's other preoccupations make on it, and it is at this point in my argument that the automatic nature of this Addressee-Speaker relationship becomes significant. Using the framework and terminology of dual-process theory, I claim that the Conceptualizer operates to a much greater extent than

previously recognised as a System-1 set of processes: fast, automatic, capable of massive parallel processing, relatively undemanding of effort, and working below the level of consciousness.

Nor is the Conceptualizer the only mechanism that operates in this way to support the Addressee-Speaker feedback loop and ensure continuing Speaker sensitivity to the Addressee's processing needs. Other System-1 resources that a Speaker can rely on for this purpose include a range of neural-level reactions (including one linked to the emotional pain of social exclusion – as valuable a danger signal as the sensation of physical pain itself); aspects of theory of mind; and intuitive inference and memory systems. The last of these is envisaged in dual-process theory as a slow-learning memory system that enables preconscious, fast retrieval of information and past associations connected with a present stimulus**,** and is suitable for use in producing a type of 'ready-made' utterance which demands little effort of both Speaker and Addressee, and whose use is extremely common: verbal routines and formulae.

A further cognitive mechanism that helps ensure automatic dovetailing between Speakers' productions and Addressees' needs is priming. Linguistic priming is already a well-established phenomenon. However, I suggest that behavioural priming – not hitherto considered in connection with linguistic performance – may also be an important source of Speaker-Addressee 'mesh', especially at the conceptualization level. If the Addressee himself can act as a prime (as behavioural priming theory suggests), then behaviourally-primed effects on the speech production process can be achieved that are more likely to yield utterances which seem relevant to an Addressee, as they have originated at least partly from him.

Overall, application of the dialogue model I describe – one based on a system of low-level, automatic interactions between Addressee and Speaker – sheds new light on both the issues raised at the start of this thesis. By demonstrating the degree of automaticity involved in message-generation (and message-inhibition), it extends the account of the Conceptualizer presented by Levelt, and also implicates a wide variety of mechanisms by which an Addressee can shape a Speaker's output, thus increasing its chances of optimal relevance. At the same time, the use this model makes of dual-process structures and arguments has perhaps extended the scope of dual-

process theory itself, given that there are few recent dual-process accounts which focus specifically on language processes. (Mercier & Sperber's discussion of the role of intuitive and reflective inferences in argumentation is a notable exception.)

I am suggesting, then, that the Addressee may be a powerful source of preconscious, relevance-shaping influence over the Speaker, not only by activating in him knowledge of how the Addressee behaves, talks and thinks, but by priming certain types of linguistic behaviour. The case of the 'think old, walk slow' experiments (cf Chapter 6.3.1) shows how this might work. Here, subjects exposed to the ELDERLY prime end up walking more slowly than controls. This behaviour might be explained on the assumption that the primed subjects have in their repertoire pre-existing plans – behavioural schemata – for different manners and speeds of walking, and that the effect of the ELDERLY prime is to activate schemata which match the 'elderly' stereotype, and are therefore more likely to be implemented. It may be that the presence of the Addressee primes the decision to convey one particular type of message rather than another in a similar way. Thus, in the presence of a friendly Addressee, a Speaker may access a behavioural schema for FRIENDLY that is realised both in social and linguistic behaviour: by speaking rather than remaining silent and, moreover, by speaking in a certain way, on a certain topic, using certain constructions and vocabulary[1]. In this way, a general intention, which may itself be primed by the presence of the Addressee, might get fleshed out in context via lower-level instances of spreading activation or priming.

A concrete example of this – the case of phatic communion, discussed in Chapter 3 – illustrates another interesting aspect of automatically-driven conceptualization. Typically, phatic utterances such as 'Those flowers – they're so pretty there… Really lovely' (Example 3.3.) express a proposition that is low in relevance to the Addressee (who has noticed this information for herself), and are, in Malinowski's phrase, 'comments on what is perfectly obvious'. In order to satisfy the Addressee's expectations of relevance, they must therefore carry implications derivable not from the proposition expressed – which is simply weakly confirmed – but on another level: for instance, from the fact that they are uttered, in these circumstances, in a

[1] Perhaps, if the Addressee is known to the Speaker, these topics, constructions, and vocabulary could be among those that form part of the Addressee's own verbal repertoire, and are therefore included in the Speaker's mental model of this particular interlocutor.

friendly tone of voice, on a topic that Speaker and Addressee are obviously agreed on rather than finding controversial, and so on. Their overall relevance to the Addressee is therefore established and, with it, the continued ability of the Speaker to continue talking: something demonstrated in the example quoted.

Here as elsewhere in phatic communion, an utterance can achieve interpersonal relevance even though the proposition expressed is not new to either the Speaker or the Addressee, and could well be retrieved ready-made. But, as I have shown, the use of ready-made utterances – with the advantages they offer in terms of speed and effortlessness – can also occur in cases where the goal is precisely to achieve relevance by informing the Addressee of something that is *new*: of, for instance, the current state of the livestock bidding conducted by New Zealand auctioneers (cf Chapter 6.2.2). Here, the overall communicative intention is determined by the nature of the situation – selling as many sheep as possible – and the Speaker's other intentions are subordinate to that. I have suggested that, in this type of case, a Speaker may have developed a schema or routine for utterance production which, thanks to the automatic processes by which it can be retrieved, permits speed under considerable pressure – and which still leaves room for a certain degree of opportunistic creativity, such as the auctioneer's references to the hills behind the auction pen. Here, relevance is achieved by a blend of automaticity of form with novelty of content.

This is not to deny, of course, that a Speaker who is relatively free from time and other situational constraints may take a more reflective stance to utterance production: planning and rehearsing his contribution in advance, developing strategic goals, drafting and redrafting to find the best possible way of achieving them. What I am claiming is simply that not *all* contributions follow this pattern, and may fall more on the intuitive, automatic rather than the effortful, reflective side. As I point out earlier (Chapter 2.4.5), it is an empirical question how much of utterance production – and comprehension – is the product of priming-driven activation, and how much is the result of intuitive or reflective inferences about an Addressee's mental state. In fact, even a reflective utterance is likely to have a substantial intuitive element: the Speaker may reflect in detail on an overall strategic goal but let lower linguistic levels look after themselves, or may produce an intuitive first

attempt at an utterance and refine it by redrafting – perhaps when promoted to do so by a visible decrease in Addressee attentiveness.

What directions can be suggested for future research? As I point out in Chapter 6.5, a top priority is to extend the evidence of the relationship between behavioural and verbal representations: in particular, evidence that behavioural primes can find expression in verbal outputs, just as verbal primes are capable of translation into behavioural outputs. One of the experimental methods described by Kay et al (Chapter 6.3.4) could be particularly useful here, as it specifically focuses on a verbal output from a behavioural – in this case, a 'material' – prime. Devising additional versions of their word completion paradigm, in which behavioural primes could be used to test the activation of other verbally-expressed concepts besides COMPETITIVE/COOPERATIVE, could give important information on the role played by priming as a whole in language production. The work carried out by Dijksterhuis et al and by Catmur et al (Chapter 6.5.2) on the impact of priming on language recall also marks important progress in this area, and extending these inquiries by varying the stereotype involved would again be extremely interesting.

Another suggested area of further work concerns the core of my hypothesis: the power of the Addressee to encourage or eclipse a Speaker. Variants on the 'narrative' method used by Bavelas et al (Chapter 4.2.4-5) could be valuable here, while the experimental opportunities offered by virtual reality technology, such as that used by Garau et al and Boker, Cohn et al (Chapter 4.3.2-3) are enormously exciting. They would also shed more light on an area that, while demonstrably important, I have neglected later in my narrative to some extent: the role of affect in naturally-occurring dialogue. The expression-damping methods described by Boker, Cohn et al are particularly relevant, as they could offer insight into the mechanisms of negative affect at work in the adult communicator (an effect already well-established in child development studies). There could, moreover, be an intriguing link here with the neuropragmatic work of van Berkum et al (2009), which shows that mood can have a testable effect on comprehension. For instance, a happy mood increases the amount of processing effort that an audience is prepared to put into the comprehension process, enabling him to activate more contextual information and

form anticipatory hypotheses[2] which may alter the assignment of reference to pronouns. Meanwhile, a depressed mood alters the effort-effect balance in the opposite direction, reducing the amount of effort the audience is prepared to put in, leading to different assignments of reference. This suggests an intriguing explanation for the effect of Addressee inattentiveness on the language production process: the distress that results from 'blanking' may alter the effort-effect balance in the same way, by reducing the amount of effort available for both production and comprehension, with even the simplest utterance feeling like hard work. It would be interesting to investigate these possible parallels between production and comprehension in more detail.

Continued multidisciplinary research along these lines could, I suggest, help expand our knowledge of Sperber's 'next frontier' in pragmatics, and contribute to the development of a unified body of understanding about the relationships between Speakers and their Addressees.

[2] E.g. a sentence starting 'Joe feared Sarah because…' invites the anticipatory hypothesis that the person to be feared is female (Sarah). Following on with the pronoun 'he' disconfirms this expectation: an outcome that can be observed and documented.

References

Abercrombie, D (1994), 'Phatic communion', *The Encyclopedia of Language and Linguistics* (Asher, R E, ed-in-chief), Vol 6. Pergamon, Oxford: 3006-7

Adamson, L B & Frick, J E (2003), 'The still face: a history of a shared experimental paradigm', *Infancy 4/4*: 451-473Adolphs, R (1999), 'Social cognition and the human brain', *Trends in Cognitive Sciences 3/12*: 469-479

Adolphs, R (2001, 'The neurobiology of social cognition', *Current Opinion in Neurobiology 11*: 231-239

Adolphs, R, Tranel, D, & Damasio, A R (1998), 'The human amygdala in social judgment', *Nature 393* (4 June): 470-474

Aijmer, K (1996), *Conversational Routines in English: Convention and Creativity*. Longman, London and New York (quoted by Pickering & Garrod, 2004: 181)

André, E, Rehm, M, Minker, W. & Bühler, D (2004), 'Endowing spoken language dialogue systems with emotional intelligence'' *Affective Dialogue Systems (ADS 2004)*, *Lecture Notes in Computer Science* (Vol 3068): 178-187

Argyle, M & Dean, J (1965), 'Eye-contact, distance and affiliation', *Sociometry 28/3*: 289-304

Baker, R E, Gill, A J, & Cassell, J (2008), 'Reactive redundancy and listener comprehension in direction-giving'. *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, Association for Computational Linguistics: 37-45

Bangerter, A, Clark, H H, & Katz, A R (2004), 'Navigating joint projects in telephone conversations', *Discourse Processes 37/1I*: 1-23

Bargh, J A (2005), 'Bypassing the will: toward demystifying the nonconscious control of social behavior', *The New Unconscious* (Hassin, R R, Uleman, J S, & Bargh, J A, eds). Oxford University Press, Oxford: 37-58

Bargh, J A (2006), 'What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior', *European Journal of Social Psychology 36*: 147-168

Bargh, J A , Chen, M, & Burrows, L (1996), 'Automaticity of social behavior: direct effects of trait construct and stereotype activation on action', *Journal of Personality and Social Psychology 71/2*: 230-244 (**http://ovidsp.tx.ovid.com.libproxy.ucl.ac.uk**/; Ovid full text pagination)

Bargh, J A & Chartrand, T L (1999), 'The unbearable automaticity of being', *American Psychologist 54/7*: 462-479 (**http://ovidsp.tx.ovid.com.libproxy.ucl.ac.uk/**; Ovid full text pagination)

Bargh, J A, Gollwitzer, P M, Lee-Chai, A, Barndollar, K, & Trötschel, R (2001), 'The automated will: nonconscious activation and pursuit of behavioural goals', *Journal of Personality and Social Psychology 81/6*: 1014-1027 (**http://ovidsp.tx.ovid.com.libproxy.ucl.ac.uk/**; Ovid full text pagination)

Baron-Cohen, S (1995, paperback edition, 1997), *Mindblindness: an Essay on Autism and Theory of Mind*. MIT Press, Cambridge Mass and London

Baron-Cohen, S, Wheelwright, S, Hill, J, Raste, Y & Plumb, I (2001), 'The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism', *Journal of Child Psychology and Psychiatry and Allied Disciplines 42*: 241-251

Bavelas, J B & Chovil, N (2000), 'Visible acts of meaning: an integrated message model of language in face-to-face dialogue', *Journal of Language and Social Psychology 19*: 163-194

Bavelas, J B, Coates, L, & Johnson, T (2000), 'Listeners as co-narrators', *Journal of Personality and Social Psychology 79/6*: 941-952

Bavelas, J B, Coates, L, & Johnson, T (2002), 'Listener responses as a collaborative process: the role of gaze', *Journal of Communication* (September 02): 566-580

Baylor, A L (2009), 'Promoting motivation with virtual agents and avatars: role of visual presence and appearance', *Computation of Emotions in Man and Machines* (Robinson, P & el Kaliouby, Rana, eds), Philosophical Transactions of the Royal Society B (vol 364 no 1535). Royal Society Publishing, London: 3559-3565

Bickmore, T & Cassell, J (1999), 'Small talk and conversational storytelling in embodied conversational interface agents', *AAAI Fall Symposium on Narrative Intelligence 1999* (http://www.psu.edu; pdf pagination)

Bock, J K (1986), 'Syntactic persistence in language production', *Cognitive Psychology 18*: 355-387

Boker, S M, Cohn, J F, Theobald, B-J, Matthews, I, Brick, T R & Spies, J R (2009), 'Effects of damping head movement and facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars', *Computation of Emotions in Man and Machines* (Robinson, P & el

Kaliouby, Rana, eds), Philosophical Transactions of the Royal Society B. Royal Society Publishing (vol 364 no 1535), London: 3485-3495

Branigan, H P, Pickering, M J, & Cleland, A A (2000), 'Syntactic co-ordination in dialogue', *Cognition 75*: B13-25

Brass, M, Schmitt, R M, Spengler, S & Gergely, G (2007), 'Investigating action understanding: inferential processes versus action simulation', *Current Biology 17*: 2117-2121

Brass, M & Heyes, C (2005). 'Imitation: is cognitive neuroscience solving the correspondence problem?', *Trends in Cognitive Sciences 9*/*10*: 489-495

Brennan, S E & Metzing, C A (2004), 'Two steps forward, one step back: Partner-specific effects in a psychology of dialogue' (commentary on Pickering & Garrod, 2004), *Behavioral and Brain Sciences 27***:**192-3

Brockmann, C, Isard, A, Oberlander, J, & White, M (2005), 'Modelling alignment for affective dialogue', *Proceedings of the 27th Annual Meeting of the Cognitive Science Society 2005* (http://www.r-brockmann.de/cb/publications/biow05_maad.pdf; pdf pagination)

Brown, P M & Dell, G S (1987), 'Adapting production to comprehension: the explicit mention of instruments', *Cognitive Psychology 19*: 441-472

Brown, P & Levinson, S C (1978/1987), 'Politeness: some universals in language usage', *Questions and Politeness* (Goody, E N, ed). Cambridge University Press, Cambridge UK; re-issued in book format 1987, CUP (book format pagination given here)

Cabeza, R & Nyberg, L (2000), 'Imaging cognition II: an empirical review of 275 PET and fMRI studies', *Journal of Cognitive Neuroscience 12/1*: 1-47

Calder, C (2006), personal communication

Camras, L A, Sullivan, J & Michel, G (1993), 'Do infants express discrete emotions? Adult judgments of facial, vocal, and body actions', *Journal of Nonverbal Behavior 17/3*: 171-186CANCODE (Cambridge and Nottingham Corpus of Discourse in English) (1994-2001). Cambridge University Press, Cambridge UK (quoted by Carter, 2002)

Carreiras, M, Garnham, A, Oakhill, J, & Cain, K (1996), 'The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish', *Quarterly Journal of Experimental Psychology 49A/3*: 639-663

Carruthers, P (2006), *The Architecture of the Mind*. Oxford University Press, Oxford

Carruthers, P (2009), 'An architecture for dual reasoning', *In Two Minds: Dual Processes and Beyond* (Evans, J St B T & Frankish, K, eds). Oxford University Press, Oxford: 109-127

Carter, Rita (1998/2000), *Mapping the Mind*. Weidenfeld & Nicolson/Phoenix, London

Carter, Ronald (2002), 'Language, creativity and creating relationships', *Humanising Language Teaching 4*/6 (http://www.hltmag.co.uk/nov02/idea.htm; online pagination)

Carver, C S, Ganellen, R J, Froming, W J & Chambers, W (1983), 'Modeling: an analysis in terms of category accessibility', *Journal of Experimental Social Psychology 19*: 403-421

Catmur, C, Walsh, V & Heyes, C (2007), 'Sensorimotor learning configures the human mirror system', *Current Biology 17*: 1527-1531

Catmur, C, Mars, R B, Rushworth, M F & Heyes, C (2011), 'Making mirrors: premotor cortex stimulation enhances mirror and counter-mirror motor facilitation', *Journal of Cognitive Neuroscience 23/9*: 2352-2362

Catmur, C, Berry, C J, Dijksterhuis, A, Otten, L, Pollard, A J, Tosto, M G & Bird, G, 'Thinking of older adults affects memory encoding' (in preparationChein, J M, & Schneider, W (2005), 'Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning', *Cognitive Brain Research 25*: 607-623

Cherry, E C (1953), 'Some experiments on the recognition of speech, with one and two ears', *Journal of the Acoustic Society of America 25*: 975-979 (quoted by Bargh, 2006: 161)

Cisek, P & Kalaska, J F (2004), 'Neural correlates of mental rehearsal in dorsal premotor cortex', *Nature 431* (21 October 2004): 993-996

Clark, H H (1992), *Arenas of Language Use,* University of Chicago Press. Chicago Ill

Clark, HH (1996), *Using Language*. Cambridge University Press, Cambridge UK (quoted by Pickering & Garrod, 2004: 176)

Clark, H H & Krych, M A (2004), 'Speaking while monitoring addressees for understanding', *Journal of Memory and Language 50*: 62-81

Cohn, J F & Tronick E Z (1983), 'Three month old infants' reaction to simulated maternal depression', *Child Development* 54: 185-193 (quoted by Boker, Cohn et al, 2009)

Cohn, J F & Tronick, Edward Z (1987), 'Mother-infant face-to-face interaction: the sequence of dyadic states at 3, 6, and 9 months', *Developmental Psychology 23/1*: 68-77

Coupland, J, Coupland, N, & Robinson, J D (1992), '"How are you?": negotiating phatic communion', *Language in Society 21*: 207-230

Cowie, R (2009), 'Perceiving emotion: towards a realistic understanding of the task', *Computation of Emotions in Man and Machines* (Robinson, P & el Kaliouby, Rana, eds.), Philosophical Transactions of the Royal Society B (vol 364 no 1535). Royal Society Publishing, London: 3515-3525

Csibra, G (2007), 'A communicative-referential account of gaze following in human infants', seminar given at the Department of Phonetics & Linguistics, University College London, 2007

Csibra, G (2010), 'Recognising communicative intentions in infancy', *Mind & Language 25/2*: 141-168

Dapretto, M, Davies, M S, Pfeifer, J H, Scott, A A, Sigman, M, Bookheimer, S Y, & Iacoboni, M (2005), 'Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders', *Nature Neuroscience,* December 05 (http://www.nature.com/natureneuroscience; online pagination)

Decety, J, Grezes, J, Costes, N, Perani, D, Jeannerod, M, Procyk, E, Grassi, F, & Fazio, F (1997), 'Brain activity during observation of action: influence of action content and subject's strategy', *Brain 120*: 1763-1777

Delgado, C E F, Messinger, D S, Yale, M E (2002), 'Infant responses to direction of parental gaze: a comparison of two still-face conditions', *Infant Behavior and Development 25*: 311-318

Dell, G S (1986), 'A spreading-activation theory of retrieval in sentence production', *Psychological Review 93/3*: 283-321

DeMarree, K G, Wheeler, S C, & Petty, R E (2005), 'Priming a new identity: self-monitoring moderates the effect of nonself primes on self-judgments and behavior', *Journal of Personality and Social Psychology 89/5*: 657-671

Devine, P G (1989); 'Stereotypes and prejudice: their automatic and controlled components', *Journal of Personality and Social Psychology 56/1*: 5-18 (**http://ovidsp.tx.ovid.com.libproxy.ucl.ac.uk/**; Ovid full text pagination)

Dijksterhuis, A & van Knippenberg, A (1998), 'The relation between perception and behavior, or how to win a game of Trivial Pursuit', *Journal of Personality and Social Psychology 74/4*: 865-877 (**http://ovidsp.tx.ovid.com.libproxy.ucl.ac.uk/**; Ovid full text pagination)

Dijksterhuis, A, Aarts, H, Bargh, J A & van Knippenberg, A (2000), 'On the relation between associative strength and automatic behavior', *Journal of Experimental Social Psychology 36*: 531-544

Dijksterhuis, A, Bargh, J A & Miedema, J (2000), 'Of men and mackerels: attention and automatic behavior', *Subjective experience in social cognition and behavior* (Bless, H & Forgas, J P, eds). Psychology Press, Philadelphia: 36-51 (quoted in Dijksterhuis et al, 2000)

Eggins, S & Slade, D (1997), *Analyzing Casual Conversation*. Cassell, London and Herndon VA

Eisenberger, N I, Lieberman, M D, & Williams, K D (2003), 'Does rejection hurt? An fMRI study of social exclusion', *Science 302*: 290-292

Ellsworth, Phoebe C, Carlsmith, J Merrill & Henson, Alexander (1972), 'The stare as a stimulus to flight in human subjects: a series of field experiments', *Journal of Personality and Social Psychology 21/3*: 302-311

Evans, J St B T (2003) 'In two minds: dual-process accounts of reasoning', *Trends in Cognitive Sciences 7/10*: 454-459

Evans, J St B T (2009), 'How many dual-process theories do we need? One, two, or many?', *In Two Minds: Dual Processes and Beyond* (Evans, J St B T & Frankish, K, eds). Oxford University Press, Oxford: 33-54

Farroni, T, Csibra, G, Simion, F & Johnson, M H (2002), 'Eye contact detection in humans from birth', *Proceedings of the National Academy of Sciences of the USA 99*: 9602-5

Ferguson, C A (1976), 'The structure and use of politeness formulas', *Language in Society 5*: 137-151

Ferstl, E C & von Cramon, D Y (2002), 'What does the frontomedian cortex contribute to language processing: coherence or theory of mind?', *NeuroImage 17*: 1599-1612 (quoted in Frith & Frith, 2003/4)

Frankish, K (2009), 'Systems and levels: dual-system theories and the personal-subpersonal distinction', *In Two Minds: Dual Processes and Beyond* (Evans, J St B T & Frankish, K, eds). Oxford University Press, Oxford: 89-107

Frankish, K & Evans, J St B T (2009), 'The duality of mind: an historical perspective', *In Two Minds: Dual Processes and Beyond* (Evans, J St B T & Frankish, K, eds). Oxford University Press, Oxford: 1-29

Frith, Uta (1989/2003), *Autism: Explaining the Enigma* (2nd ed). Blackwell, Oxford

Frith, U & Frith, C D (2003/4), 'Development and neurophysiology of mentalizing', *The Neuroscience of Social Interaction: Decoding, Imitating, and Influencing the Actions of Others* (Frith, C D & Wolpert, D M, eds). Royal Society, London (2003)/Oxford University Press, Oxford (2004): 45-75

Fussell, S R & Kraut, R E (2004), 'Visual copresence and conversational coordination' (commentary on Pickering & Garrod 2004), *Behavioral and Brain Sciences 27*: 196-7

Gallese, V (2003/4), 'The manifold nature of interpersonal relations: the quest for a common mechanism', *The Neuroscience of Social Interaction: Decoding, Imitating, and Influencing the Actions of Others* (Frith, C D & Wolpert, D M, ed). Royal Society, London (2003)/Oxford University Press, Oxford (2004): 159-182

Garau, M, Slater, M, Bee, S, & Sasse, M A (2001), 'The impact of eye gaze on communication using humanoid avatars'; *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York: 309-316 (http://www.psu.edu/; pdf pagination)

Garau, M, Slater, M, Pertaub, D-P, & Razzaque, S (2005), 'The responses of people to virtual humans in an immersive virtual environment', *Presence 14/1*: 104-116 (http://www.psu.edu/; pdf pagination)

Garau, M, Slater, M, Vinayagamoorthy, V, Brogni, A, Steed, A, & Sasse, M A (2003), 'The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment', *CHI* 2003 *April 5-10*. ACM, New York (online pagination used here)

Garrod, S & Anderson, A (1987), 'Saying what you mean in dialogue: a study in conceptual and semantic co-ordination', *Cognition 27*: 181-218

Garrod, S & Pickering, M J (2007), 'Alignment in dialogue', *The Oxford Handbook of Psycholinguistics* (Gaskell, G, ed). Oxford University Press, Oxford: 443-451

Garrod, S & Pickering, M J (2009), 'Joint action, interactive alignment, and dialog', *Topics in Cognitive Science (2009)*: 292-304

Goffman, E (1963), *Behavior in Public Places.* Free Press of Glencoe, London (quoted by Argyle & Dean, 1965)

Grèzes, J & Decety, J (2001), 'Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis', *Human Brain Mapping 12*: 1-19

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

Harley, T A (2001), *The Psychology of Language: from Data to Theory* (2nd ed). Psychology Press, Hove and New York

Haxby, J V, Hoffman, E A, & Gobbini, M I (2000), 'The distributed human neural system for face perception', *Trends in Cognitive Sciences 4*/*6*: 223-233

Heim, S, Eickhoff, S B, & Amunts, K (2008), 'Specialisation in Broca's region for semantic, phonological, and syntactic fluency?', *NeuroImage 40*/3: 1362-1368

Hickok, G (2009), 'Eight problems for the mirror neuron theory of action understanding in monkeys and humans', *Journal of Cognitive Neuroscience 21/7*: 1229-1243

Hill, T (2009), 'Feel like you're the writer type (fiction)?'; *The Times (Saturday Review),* 3.10.09: 12

Horton, W S & Gerrig, R J (2005a), 'Conversational common ground and memory processes in language production', *Discourse Processes 40/1*: 1-35

Horton, W S & Gerrig, R J (2005b), 'The impact of memory demands on audience design during language production', *Cognition 96*: 127-142

Iacoboni, M, Woods, R P, Brass, M, Bekkering, H, Mazziotta, J C, & Rizzolatti, G (1999), 'Cortical mechanisms of human imitation', *Science 286*: 2526-2528

Iacoboni, M, Molnar-Szakacs, I, Gallese, V, Buccino, G, Mazziotta, J C & Rizzolatti, G (2005), 'Grasping the intentions of others with one's own mirror neuron system', *Public Library of Science (PLoS) Biology 3*/*3*: 529-535

Indefrey, P & Levelt, W J M (2004), 'The spatial and temporal signatures of word production components', *Cognition 92*: 101-144

Isenberg, N, Silbersweig, D, Engelien, A, Emmerich, S, Malavade, K, Beattie, B, Leon, A C, & Stern, E (1999), 'Linguistic threat activates the human amygdala', *Proceedings of the National Academy of Science USA/96*: 10456-10459 (http://www.pnas.org; online pagination)

Izard, C (1979), *The maximally discriminative facial movement coding system (MAX)*, University of Delaware, Instructional Resources Center; Newark DE

Izard, C, Dougherty, L & Hembree, E (1983), *A system for identifying affect expressions by holistic judgements (AFFEX)*, University of Delaware, Instructional Resources Center; Newark DEJacob, Pierre (2008), 'What do mirror neurons contribute to human social cognition?', *Mind & Language 23/2*: 190-223

Jakobson, R (1960/1995), 'Linguistics and poetics', *Selected Writings III,* republished in 'The speech event and the functions of language', *On Language* (Waugh, Linda R & Monville-Burston, Monique, eds). Harvard University Press (paperback edition, 1995), Cambridge, Mass and London: 69-79

Jeannerod, M (1999), 'To act or not to act: perspectives on the representation of actions (25th Bartlett Lecture)', *The Quarterly Journal of Experimental Psychology Section A 52/1*: 1-29

Johnson-Laird, P N (1983), *Mental Models: Toward a Cognitive Science of Language Inference and Consciousness.* Harvard University Press, Cambridge Mass

Kampe, K K W, Frith, C D, & Frith, U (2003), '"Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality', *The Journal of Neuroscience 23/12*: 5258-5263 (http://www.jneuroscience.org.libproxy.ucl.ac.uk; full text pagination)

Kay, A C, Wheeler, S C, Bargh, J A, & Ross, L (2004), 'Material priming: the influence of mundane physical objects on situational construal and competitive behavioural choice', *Organizational Behavior and Human Decision Processes 95*: 83-96

Kempson, R (2004), 'Grammars with parsing dynamics: a new perspective on alignment', (commentary on Pickering & Garrod, 2004), *Behavioral and Brain Sciences 27***:** 202-3

Kempson, R, Meyer-Viol, W, & Gabbay, D (2001), *Dynamic Syntax: the Flow of Language Understanding.* Blackwell, Oxford

Kendon, A (1967), 'Some functions of gaze-direction in social interaction', *Acta Psychologica 26*: pp 22-63

Kendon, A & Cook, M (1969), 'The consistency of gaze patterns in social interaction', *British Journal of Psychology 60/4*: 481-494

Kingsbury, D (1968), 'Manipulating the amount of information obtained from a person giving directions', unpublished honours thesis, Department of Social Relations, Harvard University (quoted by Krauss & Pardo, 2004)

Kleinke, C L (1986), 'Gaze and eye contact: a research review', *Psychological Bulletin 100/1*: 78-100

Kleinke, C L, Staneski, R A & Berger, D E (1975), 'Evaluation of an interviewer as a function of interviewer gaze, reinforcement of subject gaze, and interviewer attractiveness', *Journal of Personality and Social Psychology 31/1*: 115-122

Krauss, R M & Pardo, J S (2004), 'Is alignment always the result of automatic priming?' (commentary on Pickering & Garrod, 2004), *Behavioral and Brain Sciences 27***:** 203-4

Kraut, R E, Fussell, S R, & Siegel, Jane (2003), 'Visual information as a conversational resource in collaborative physical tasks', *Human-Computer Interaction 18*: 13-49

Kuiper, Koenraad (1996), *Smooth Talkers: the Linguistic Performance of Auctioneers and Sportscasters*. Lawrence Erlbaum, NJ

Lashley, K S (1951), 'The problem of serial order in behavior', *Cerebral Mechanisms in Behaviour: the Hixon Symposium* (Jeffress, L A, ed.). Wiley, New York and Chapman & Hall, London: 112-146

Laver, J (1975), 'Communicative functions of phatic communion', *Organization of Behavior in Face-to-Face Interaction* (Kendon, A, Harris, Richard M & Key, M R, eds.). Mouton, The Hague: 215-240 (also published in Laver, John (1974), *Semiotic Aspects of Spoken Communication*. Edward Arnold, London)

Leighton, J, Bird, G, Charman, T, & Heyes, C (2008), 'Weak imitative performance is not due to a functional "mirroring" deficit in adults with Autism Spectrum Disorders', *Neuropsychologia 46*: 1041-1049

Levelt, W J M (1989), *Speaking: from Intention to Articulation*. MIT Press, Cambridge Mass & London

Levelt, W J M , Roelofs, A, & Meyer, A S (1999), 'A theory of lexical access in speech production', *Behavioral and Brain Sciences 22*: 1-75

Lieberman, M D (2007), 'Social cognitive neuroscience: a review of core processes', *Annual Review of Psychology 58*: 259-289

Lieberman, M D (2009), 'What zombies can't do: a social cognitive neuroscience approach to the irreducibility of reflective consciousness', *In Two Minds: Dual Processes and Beyond* (Evans, J St B T & Frankish, K, eds). Oxford University Press, Oxford: 293-316

Lieberman, M D, Jarcho J M & Satpute, A B (2004), 'Evidence-based and intuition-based self-knowledge: an fMRI study', *Journal of Personality and Social Psychology 87*/*4*: 421-35

Malinowski, B (1923), 'The problem of meaning in primitive languages', *The Meaning of Meaning* (Ogden, C K & Richards, I A, eds). Kegan Paul, London and Harcourt Brace, New York: 451-481

Martin, A, Haxby, J V, Lalonde, F M, Wiggs, C L, & Ungerleider, LG (1995), 'Discrete cortical regions associated with knowledge of color and knowledge of action', *Science 270*/*5233*: 102-105

Matias, R & Cohn, J F (1993), 'Are Max-specified infant facial expressions during face-to-face interaction consistent with differential emotions theory?', *Developmental Psychology 29*/*3*: 524-531Matlock, T (2004), 'Fictive motion as cognitive simulation', *Memory & Cognition 32*/*8*: 1389-1400

Mercier, H & Sperber, D (2009), 'Intuitive and reflective inferences', *In Two Minds: Dual Processes and Beyond* (Evans, J St B T & Frankish, K, eds). Oxford University Press, Oxford: 149-170

Mercier, H & Sperber, D (2011), 'Why do humans reason? Arguments for an argumentative theory', *Behavioral & Brain Sciences 34*: 57-111

Metzing, C & Brennan, S E (2003), 'When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions': *Journal of Memory & Language 49*: 201-13

Morrow, D G (1985), 'Prominent characters and events organize narrative understanding', *Journal of Memory and Language 24/3*: 304-319

Murray, L & Trevarthen, C (1985), 'Emotional regulation of interactions between two-month-olds and their mothers', *Social Perception in Infants* (Field, Tiffany M & Fox, Nathan A, ed). Ablex, Norwood NJ

Myers, J L, O'Brien, E J, Albrecht, J E, & Mason, R A (1994), 'Maintaining global coherence during reading', *Journal of Experimental Psychology: Learning, Memory, and Cognition 20/4*: 876-886

Nagy, E (2008), 'Innate intersubjectivity: newborns' sensitivity to communication disturbance', *Developmental Psychology 44/6*: 1779-1784

Neal, D T & Chartrand, T L (2011), 'Embodied emotion perception: amplifying and dampening facial feedback modulates emotion perception accuracy', *Social Psychological and Personality Science* (online publication; doi:10.1177/19485506 11406 138): http://spp.sagepub.com

Nishitani, N, Schürmann, M, Amunts, K & Hari, R (2005), 'Broca's region: from action to language', *Physiology 20*: 60-69

Oster, H, Hegley, D & Nagel, L (1992), 'Adult judgments and fine-grained analysis of infant facial expressions: testing the validity of a priori coding formulas', *Developmental Psychology 28/6*: 1115-1131Perani, D, Cappa, S F, Schnur, T, Tettamanti, M, Collina, S, Rosa, M M, & Fazio, F (1999), 'The neural correlates of verb and noun processing: a PET study', *Brain 122*: 2337-2344

Perrett, D, & Mistlin, A (1990) 'Perception of facial characteristics by monkeys', *Comparative Perception vol 2: Complex Signals* (Stebbins W & Berkely M, eds). Wiley, Oxford (quoted in Baron-Cohen, 1995)

Pertaub, D-Pl, Slater, M, & Barker, C (2001), 'An experiment on public speaking anxiety in response to three different types of virtual audience', *Presence: Teleoperators & Virtual Environments 11/1*: 68-78 (http://www.psu.edu/; pdf pagination)

Pickering, M J (2006), 'The dance of dialogue', *The Psychologist 19/12*: 734-737

Pickering, M J & Garrod, S (2004), 'Toward a mechanistic psychology of dialogue', *Behavioral & Brain Sciences 27***:** 169-226

Piwek, P (2006), 'Perspectives on dialogue: introduction to this special issue [on dialogue processes]', *Research on Language and Computation 4*: 143-152

Pollard, A J (2005a), 'Beyond politeness: phatic communication and relevance theory', unpublished MS submitted as part of MA in Pragmatics, University College London,

Pollard, A J (2005b), 'The one who leads: relationships between mind-reading, mindblindness, and the role of the speaker in relevance theory', dissertation

submitted in partial fulfilment of requirements for MA in Pragmatics, University College London

Posner, M I & Raichle, M E (1994), *Images of Mind*. Scientific American Library, NY

Puce, A & Perrett, D (2003/4), 'Electrophysiology and brain imaging of biological motion', *The Neuroscience of Social Interaction: Decoding, Imitating, and Influencing the Actions of Others* (Frith, C D & Wolpert, D M, eds); The Royal Society. London (2003)/Oxford University Press, Oxford (2004): 1-21

Purver, M, Cann, R & Kempson, R (2006),'Grammars as parsers: meeting the dialogue challenge', *Research on Language and Computation 4*: 289-326

Reeves, B & Nass, C I (1996), *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Publications, Chicago Ill, and Cambridge University Press, Cambridge

Rilling, J K, Sanfey, A G, Aronson, J A, Nystrom, L E & Cohen, J D (2004), 'The neural correlates of theory of mind within interpersonal interactions', *NeuroImage 22*: 1694-1703

Rizzolatti, G & Arbib, M A (1998), 'Language within our grasp', *Trends in Neuroscience 21/5*: 188-194

Robinson, P & el Kaliouby, R (2009), 'Computation of emotions in man and machines: introduction'; *Computation of Emotions in Man and Machines* (Robinson, P & el Kaliouby, Rana, eds), Philosophical Transactions of the Royal Society B (vol 364, no 1535). Royal Society Publishing, London: 3441-3447

Sacks, O (1994), 'A neurologist's notebook: an anthropologist on Mars', *New Yorker,* December 27 1993-January 3 1994 (quoted by Baron-Cohen (1995: 140)

Sanford, A J, Moar, K, & Garrod, S C (1988), 'Proper names as controllers of discourse focus', *Language and Speech 31/1*: 43-56

Satpute, A B & Lieberman, M D (2006), 'Integrating automatic and controlled processes into neurocognitive models of social cognition', *Brain Research 1079*: 86-97

Schacter, D L & Buckner, R L (1998), 'Priming and the brain', *Neuron 20/2*: 185-195

Schacter, D L, Wig, G S, & Stevens, W D (2007), 'Reductions in cortical activity during priming', *Current Opinion in Neurobiology 17/2*: 171-176

Schilbach, L, Wilms, M, Eickhoff, S B, Romanzetti, S, Tepest, R, Bente, G, Shah, N J, Fink, G R, & Vogeley, K (2010), 'Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry', *Journal of Cognitive Neuroscience 22/12*: 2702-2715

Schober, M F & Clark, H H (1989), 'Understanding by addressees and overhearers', *Cognitive Psychology 21*: 211-232

Sebastian, C L, Fontaine, N M G, Bird, G, Blakemore, S-J, De Brito, S A, McCrory, E J P, & Viding, E (2011), 'Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults', *Social Cognitive and Affective Neuroscience 2011*. Advance Access published April 4, 2011; doi:10.1093/scan/nsr023

Shamay-Tsoory, S G, Aharon-Peretz, J & Perry, D (2009), 'Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions', *Brain 132*: 617-627

Singer, T, Wolpert, D M, & Frith, C D (2003/2004), 'Introduction: the study of social interactions', *The Neuroscience of Social Interaction: Decoding, Imitating, and Influencing the Actions of Others* (Frith, C D & Wolpert, D M, eds). The Royal Society, London (2003)/Oxford University Press, Oxford (2004): xiii-xxvii

Slater, M, Pertaub, D-P, Barker, C, & Clark, D (2006), 'An experimental study on fear of public speaking using a virtual environment', *Cyberpsychology & Behavior* (online version; online pagination used here)

Sloman, S A (1996), 'The empirical case for two systems of reasoning', *Psychological Bulletin 119/1*: 3-22

Smith, E R & DeCoster, J (2000), 'Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems', *Personality and Social Psychology Review 2000/4*: 108-131

Smith, E R & Collins, E C (2009), 'Dual-process models: a social psychological perspective', *In Two Minds: Dual Processes and Beyond* (Evans, J St B T & Frankish, K, eds). Oxford University Press, Oxford: 197-216

Sperber, D (2005), 'Modularity and relevance: how can a massively modular mind be flexible and context dependent?', *The Innate Mind* ((vol 364 no 1535). (eds). Oxford University Press, Oxford: 53-68

Sperber, D (2008), *Relevance in Communication*, lecture at University College London

Sperber, D & Wilson, D (1982), 'Mutual knowledge and relevance in theories of comprehension', *Mutual Knowledge* (Smith, N V, ed.). Academic Press, London: 61-87

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

Sperber, D & Wilson, D (2002). 'Pragmatics, modularity and mindreading', *Mind & Language 17/1-2*: 3-23

Srull, T K & Wyer, R S (1979), 'The role of category accessibility in the interpretation of information about persons: some determinants and implications', *Journal of Personality and Social Psychology 37/10*: 1660-1672

Stanovich, K E (2004), *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. University of Chicago Press, Chicago Ill (quoted by Thompson, 2009)

Stanovich, K E (2009), 'Distinguishing the reflective, algorithmic, and autonomous minds: is it time for a tri-process theory?', *In Two Minds: Dual Processes and Beyond* (Evans, J St B T & Frankish, K, eds). Oxford University Press, Oxford: 55-88

Striano, T (2004), 'Direction of regard and the still-face effect in the first year: does intention matter?', *Child Development 75/2*: 468-479

The Times 2 (2010), 'Young Times: jumbled up', 18.10.10: 21

Thompson, V A (2009), 'Dual-process theories: a metacognitive perspective', *In Two Minds: Dual Processes and Beyond* (Evans, J St B T & Frankish, K, eds). Oxford University Press, Oxford: 171-195

Tronick, E, Adamson, L B, Als, H & Brazelton, T B (1975), 'Infant emotions in normal and perturbated interactions', paper presented at the biennial meeting of the Society for Reseearch in Child Development

Tronick, E, Als, H, Adamson, L, Wise, S & Brazelton, T B (1978), 'The infant's response to entrapment between contradictory messages', *Journal of the American Academy of Child Psychiatry 17*: 1-13

Uleman, J S (2005), 'Introduction: becoming aware of the new unconscious', *The New Unconscious* (Hassin, R R, Uleman, J S, & Bargh, J A, eds.); Oxford University Press, Oxford 3-15:

Ullman, M T (2006), 'Is Broca's area part of a basal ganglia thalamocortical circuit?', *Cortex 42*: 480-485

Van Berkum, J J A, De Goede, D, Van Alphen, P, Mulder, E, & Kerstholt, J (2009), 'Mood and heuristic anticipation in language comprehension', Third Biennial Meeting of Experimental Pragmatics (XPRAG 2009); Lyons, France

Van der Henst, J B & Sperber, D (2004), 'Testing the cognitive and communicative principles of relevance', *Experimental Pragmatics* (Noveck, I & Sperber, D, eds). Palgrave Macmillan, Basingstoke: 229-279

Vega Moreno, R E (2005), *Creativity and convention: the pragmatics of everyday figurative speech*, PhD Thesis in Linguistics, University College London

Vega Moreno, R E (2007), *Creativity and Convention: The Pragmatics of Everyday Figurative Speech*. John Benjamins, Amsterdam/Philadelphia

Wagenaar, W A (1986), 'My memory: a study of autobiographical memory over six years', *Cognitive Psychology 18*: 225-252

Walker, J, Payne, S, Smith, P, Jarrett, N (2007), *Psychology for Nurses and the Caring Professions* (3rd ed.). McGraw-Hill International, Maidenhead

Warren, T & Rayner, K (2004), 'Top-down influences in the interactive alignment model: The power of the situation model' (commentary on Pickering & Garrod 2004), *Behavioral and Brain Sciences 27***:** 203-4

Webster, B (2007), 'Hands-free mobiles "just as dangerous"', *The Times,* 5.4.07: 38

Weinberg, K M & Tronick, E (1994), 'Beyond the face: an empirical study of infant affective configurations of facial, vocal, gestural and regulatory behaviors', *Child Development 65*: 1503-1515Wieser, M J, Pauli, P, Alpers, G W, & Mühlberger, A (2009), 'Is eye to eye contact really threatening and avoided in social anxiety? – An eye-tracking and psychophysiology study', *Journal of Anxiety Disorders 23/1*: 93-103

Žegarac, V & Clark, B (1999a), 'Phatic interpretations and phatic communication', *Journal of Linguistics 35*: 321-346

Žegarac, V & Clark, B (1999b), 'Phatic communication and Relevance Theory: a reply to Ward & Horn', *Journal of Linguistics 35*: 565-577

Zwaan, R A & Radvansky, G A (1998), 'Situation models in language comprehension

and memory'; *Psychological Bulletin 123 (2)*, 162-185
(http://ovidsp.tx.ovid.co+m.libproxy.ucl.ac.uk/; Ovid full text pagination