# The Social Evolution of Pragmatic Behaviour

Thomas C. Scott-Phillips

MMath. MSc.

PhD.

University of Edinburgh

2008

I declare that I have composed this thesis and conducted the research presented herein, except where indicated otherwise.

Thomas C. Scott-Phillips

19th November 2008

So many people deserve thanks that I dare not list them all individually, lest I forget some, so I will name none. You all know who you are.

The LEC has been an outstanding place for me to conduct my postgraduate education. Thanks are due to all there for the many stimulating conversations, and for thoughtful responses to my numerous queries and questions. The same is true of the academic staff of LEL and PPLS, and the institutional support I have received extends even beyond there; it includes several individuals at the IEB, for example. Undoubtedly, the university has provided an excellent intellectual environment for my work. Thanks are also due to the many administrative staff, especially those of PPLS, who have made my postgraduate life significantly less stressful than it might otherwise have been.

The contribution of the many friends I have made in the last four years has been equally important. In particular, Edinburgh's swing dance community has provided me with much joy. I am also very grateful to my old friends who supported my decision to return to education, and who have remained in touch while I have pursued that goal.

Five people deserve special thanks. It must be frustrating to finance one's son through an undergraduate education in mathematics and into a viable career only to see him leave all behind for postgraduate education in the evolution of language and the far less secure world of academia. To their immense credit my parents have never challenged my life choices, even for a moment. I am under no illusions about how much difference such support makes when one takes such decisions.

I am also aware of how serendipitous I have been to have had the supervisors that I did. The breakdown of the student-supervisor relationship is sadly commonplace. In contrast, I have been fortunate enough to have had not one but two mentors who have treated me with respect, and who have challenged, supported, questioned and inspired me at the appropriate times. Of the many people who have influenced this thesis, it is they that have had the most direct impact, which is how it should be.

Finally, there is one person who has managed to combine all of the above: not only has she given me enormous amounts of support, friendship and joy, she has also read every word of this thesis. Thank you.

# Abstract

Pragmatics is the branch of linguistics that addresses the relationship between language and its external environment – in particular the communicative context. Social evolution (or sociobiology) is the branch of the biological sciences that studies the social behaviour of organisms, particularly with respect to the ecological and evolutionary forces with which it must interact. These two disciplines thus share a natural epistemic link, one that is concerned with the relationship between behaviour and the environment. There has, however, historically been no dialogue between them. This thesis attempts to fill that void: it examines pragmatics from the perspective of social evolution theory.

Chapter 1 gives a brief introduction to the two fields and their key ideas, and also discusses why an evolutionary understanding of pragmatics is crucial to the study of language origins.

In chapter 2 the vexed question of the biological function of language is discussed. Responses are given to the claims, common in the evolutionary linguistics literature, that the processes of exaptation, self-organisation and cultural transmission provide alternatives to natural selection as a source of design in nature. The intuitive conclusion that the function of language is communication is provisionally supported, subject to a proper definition of communication.

Chapter 3 reviews previous definitions and consequently argues for an account predicated on the designedness of signals and responses. This definition is then used to argue that an evolutionarily coherent model of language should recognise the pragmatic realities of ostension and inference and reject the code-like idealisation that is often used in its place.

Chapter 4 observes that this fits the argument that the biological function of language is communication and then addresses the key question faced by all evolved communication systems – that of evolutionary stability. The human capacity to record and remember the past behaviour of others is seen to be critical.

Chapter 5 uses the definition of communication from chapter 3 to describe a very general model of evolved communication, and then uses the constraints of that model to argue that Relevance Theory, or at least some theory of pragmatics with a very similar logical structure, must be correct.

Chapter 6 then applies the theoretical apparatus constructed in chapters 2 to 5 to a crucial and topical issue in evolutionary linguistics: the emergence of learnt, symbolic communication. It introduces the *Embodied Communication Game*, an experimental tool whose basic structure is significantly informed by both social evolutionary and, in particular, pragmatic theory. The novelty of the game is that participants must find a way to communicate not just the content that they wish to convey, but also the very fact that a given behaviour is communicative in nature, and this constraint is found to fundamentally influence the type of system that emerges.

Chapter 7, which concludes the thesis, recounts and clarifies what it tells us about the origins and evolution of language, and suggests a number of possible avenues for future research.

The question of the origin of language… must be settled before linguistics takes its final form

– Charles S. Peirce, 1902 (reproduced in Buchler, 1955, p.69)

# One

# Social Evolution and Pragmatics

Although it is true that the last 20 years or so have witnessed an explosion in language evolution research, it is wrong to think that previous generations of scholars did not consider the matter. On the contrary, it has historically generated more than its fair share of academic discussion (for a book-length treatment see Stam, 1976), and the present activity does not mark the first flourishing of interest since Darwin; the ape-language experiments of the 1960s and 1970s (e.g. Gardner & Gardner, 1969; Premack, 1971; Terrace et al., 1979), for example, were in significant part motivated by a concern with the origin(s) of language (Hauser et al., 2007). In at least one important respect, however, little has changed: because the evolution of language is an inherently interdisciplinary exercise, different scholars inevitably approach the matter from quite different theoretical perspectives, sometimes dramatically so. Moreover, they sometimes disagree, even if they do not know it themselves, on what language is and consequently what is to be explained: "considerable confusion has resulted from the use of 'language' to mean different things… positions that seemed absurd and incomprehensible, and chasms that seemed unbridgeable, were rendered quite manageable once the misunderstandings were cleared up" (Fitch et al., 2005, p.180). As a result, "some confusion in debates in the field actually arises from the fact that different questions are being asked" (Kirby, 2007, p.671).

Presumably a similar sentiment was at least in part responsible for the Société de Linguistique de Paris's mistimed 1866 ban on the subject, issued just seven years after Darwin's Origin of Species and five years before Descent of Man. Because of the conceptual difficulty of the problem some commentators still consider such an edict "defensible" (Edwardes, 2007, p.91) and others are happy to damn the research agenda of those that do not see things as they do: "there is perhaps no other field of human inquiry which has been so vitiated by a failure to get priorities straight" (Bickerton, 2007, p.510). Both these criticisms, it is worthwhile to note, come from scholars who are actively involved in language evolution research. Outside that group, some have concluded that questions regarding the origins and subsequent evolution of language simply cannot be answered. For example: "there are no facts of the matter, and therefore the Société de Linguistique de Paris was probably well-advised" (Lightfoot, 1991, p.69); and "explanations of the origin of human thought and language are simply speculations lacking the kind of detailed historical information required for an evolutionary explanation" (Richardson, 1996, p.541). While these issues can be debated (e.g. Durrant & Haig, 2001), what is not in doubt is that while Darwinism has spread, from within its original domain of biology to nearly all other corners of human inquiry, including psychology, anthropology, ethics, mathematics, archaeology, computer science, physics and even literature, it has had little impact in linguistics, at least until the recent renaissance. The broad and often profound impact that Darwinism has had elsewhere has given rise to the metaphor of a "universal acid" that "eats through just about every traditional concept, and leaves in its wake a revolutionized world-view, with most of the old landmarks still recognizable, but transformed in fundamental ways" (Dennett, 1995, p.63). The field of linguistics has to some degree been an exception to this otherwise inexorable march. It is true that historical linguistics has frequently made use of evolutionary thinking (and also that Darwin was heavily inspired by processes of language change) but the study of synchronic language has until relatively recently largely ignored Darwin's ideas – or, if it has considered them, generally deemed them not relevant (e.g. Chomsky, 1972, 1982, 1988; Mehler, 1985; Piatelli-Palmarini, 1989). As a result, "language evolution has been an intellectual orphan in linguistics" (Givón & Malle, 2002, p.vii). Such neglect may reasonably be termed a "disgrace" (Bickerton, 2007, p.524).

This situation is, however, currently being rectified. The very existence of language evolution as a field of study implies that there is a sizable body of individuals that would subscribe to the view that evolution can be a source of explanation for language and linguistics (Carstairs-McCarthy, 2007). Representatives of this community often argue that the insights offered by contemporary disciplines like computational modelling, neuroscience and genetics are sufficiently sophisticated to render the study of language origins scientifically respectable (e.g. Christiansen & Kirby, 2003a). Some of the findings that arise from these approaches, like the FOXP2 gene (Enard et al., 2002) or the discovery of mirror neurons (Rizzolatti et al., 1996), are clearly of relevance. However, debate about exactly what they tell us often takes us back to the problems discussed above, with researchers talking past each other as they seek different answers to different sorts of questions. In fact, such breakthroughs have thus far provided few constraints on theorising. For example, three chapters (Corballis, 2003; Lieberman, 2003; Pinker, 2003) in the same edited book (Christiansen & Kirby, 2003b) point towards FOXP2 as support for their theories, even though those same theories are in some important respects incompatible with each other. Consequently, journals and conferences on language evolution frequently make explicit pleas for further data in a deliberate attempt to elevate the field above the realm of theoretical speculation. This development is undeniably welcome, but we should not over-congratulate ourselves on our scientific credentials, for such demands are not new. Two hundred and fifty years ago Rousseau discussed language origins in his Discourse on the Origin of Inequality in Men and later wrote an (unfinished) Essay on the Origin of Languages. This work was sufficiently influential that post-Rousseau "it was no longer satisfactory for writers simply to submit hypotheses on the subject [of language origins], mustering whatever support they could. It became necessary, in addition, to weigh the possibility of any answer at all and to assess the ultimate usefulness of such speculations" (Stam, 1976, p.93). Unfortunately we still cannot take for granted that such demands are met as a matter of course.

Or perhaps that is unnecessarily harsh, for a degree of confusion is inevitable in a field that demands interdisciplinarity. It is no surprise that researchers will sometimes talk past each other when contributions necessarily come from a range of disciplines that is by any measure diverse: linguistics, psychology, genetics, neuroscience, artificial intelligence, computer science, philosophy, archaeology, anthropology and several other fields have all contributed to the growing language evolution literature. There is, however, a curious omission from this list: sociobiology or, to use the more contemporary term, social evolution, the study of an organism's social behaviour, particularly in the light of the evolutionary pressures placed upon it (see §1.2 below). (The term sociobiology is still used in places, but has fallen out of widespread use. This is presumably a response to the unfortunate and incorrect political connotations it derived after the publication of *Sociobiology: The new synthesis* (Wilson, 1975). *Social evolution* has sprung up in its place, and will be used in this thesis. It should not be confused with *cultural evolution* or similar terms.) To be sure, the work of many scholars in the field of language evolution has been influenced by social evolutionary principles, but a full examination of language from that perspective does not yet exist. Moreover, while some linguists interested in evolutionary matters have lamented that there are not more who share their background (e.g. Bickerton, 2003; Newmeyer, 2003), there are no such explicit demands for biologists of any stripe to join the field, despite their relative absence. In fact, those same individuals who bemoan the lack of linguists have been known to attack those who come from other disciplines on the grounds that they are ill-qualified to discuss the evolution of language (see Bickerton, 1996).

The goal of this thesis is, then, to partially address the omission. More specifically, it will ask what social evolution theory can tell us about *pragmatics*, the branch of linguistics that addresses how language is influenced by context and other factors external to the language itself (see §1.3 below). Regrettably, pragmatics has been largely neglected by the language evolution community. For example, the index to a reasonably recent edited collection of writings that was specifically commissioned to provide a comprehensive overview of the field (Christiansen & Kirby, 2003b) contains not a single reference to *pragmatics* or *relevance*. In contrast, there are eight pages indexed by *semantics* and/or *meaning*, 23 by *phonology*, *phonetics* and derivative terms, and 68 by *syntax* and/or *grammar*. With the exceptions of a few individual chapters (e.g. Dessalles, 1998; Franks & Rigby, 2005), the collections of work that have arisen from the international conferences on the evolution of language (Cangelosi et al., 2006; Hurford et al., 1998; Knight et al., 2000; Smith et al., 2008; Tallerman, 2005; Wray, 2002) show a similar neglect, and there are very few published works that buck this general trend (exceptions are Dessalles, 2007; Hoefler & Smith, 2008; Hurford, 2007; Origgi & Sperber, 2000; Smith, 2008; Sperber & Origgi, 2004; Pinker et al., 2008). This bias is reflective of the emphasis of contemporary linguistics more generally, in which syntax is seen as core and pragmatics peripheral. Accordingly, this thesis is mostly concerned with the fact of language (or language use), rather than its form. That is, its principle question is not why language takes the form that it does, but why we bother to speak at all. (Having said that, questions of form are touched upon, in particular in chapter 6.) As observed above, this is relatively unexplored territory in evolutionary linguistics, yet it is the question that springs most immediately to the mind of the social evolutionist, who, upon observing communication between organisms, "could not rest happy with the idea of animals performing such a time-consuming, and above all complex and statistically improbable, activity for nothing" (Dawkins, 1982, p.31).

Why pragmatics rather than any other branch of linguistics? Because it is the aspect of language that social evolution theory most immediately addresses. The two disciplines share a natural epistemic link: social evolution studies the behaviour of organisms in relation to each other, and pragmatics addresses interpersonal linguistic phenomena, particularly with respect to context. While syntax, semantics, phonetics and phonology are, in general, about matters internal to language itself, pragmatics asks how such phenomena interact with the outside world. It thus shares with social evolution an interest in the relationship between behaviour and the environment. Indeed, it is not possible to do pragmatics without reference to interlocutors, nor to do social evolution without reference to the interactions between organisms; nor, indeed, to do evolutionary biology without reference to the environment. It is this commonality that makes the project of this thesis an intellectually coherent one. After all, if we are to meaningfully talk about the (biological) evolution of any trait then we must talk about its interaction with the environment. In the case of language that interaction must take the form of linguistic use, otherwise there is nothing for natural selection to act on. Put another way, if we wish our definition of language to be evolutionarily meaningful (and why wouldn't we?) then we cannot define it only with reference to its internal features; we must also make reference to the interlocutors and their shared context. This perspective hence demands that, contra its present neglect, pragmatics be of pivotal importance to evolutionary linguistics. Yet as we have already observed, this is not the case. Similarly, social evolution should be of interest to pragmaticians who wish to place their discipline on a Darwinian footing, but there is very little published work on this topic (exceptions are Origgi & Sperber, 2000; Sperber, 2000, Sperber & Wilson, 2002). Furthermore, whilst it is true that the basic theories of social evolution (ideas like kin selection (Hamilton, 1964), reciprocity (Trivers, 1971), the handicap principle (Grafen, 1990; Zahavi, 1975) and the like) have been invoked by researchers in evolutionary linguistics, a systematic enquiry is lacking; and as will shortly be explained, where they have been used these ideas have too often been misapplied.

Whilst the core objective of this thesis is to consider what social evolution can tell us about pragmatics, we will also find that some valuable lessons pass in the opposite direction. That is, we will see how insights from pragmatics can help us to explore issues of general social evolutionary interest. Specifically, the problems of evolutionarily stable communication (in chapter 4) and of the emergence of communication (in chapter 6) come into sharper focus when the light of pragmatics is shone upon them. This thesis can thus also be read as an attempt to create a dialogue between two disciplines that have not previously engaged with each other, despite a natural epistemic link. With that in mind, the next two sections introduce the basic tenets and the key ideas of the two fields in question.

## 1.1 What is social evolution?

In the most general terms, social evolution is concerned with the evolution of behaviours that involve the interaction of more than one organism. This includes, then, sex, conflict, parent-offspring interactions, cooperation and, most relevantly for this thesis, communication (this is by no means a complete list, of course). Social evolutionary thinking is implicit in Darwin's own writing and that of virtually all the major figures in evolutionary biology over the following 100 years. Nevertheless, we can date the systematic enquiry into the study of animal social behaviour to, approximately, the development in the 1950s and 1960s of ethology as a fully-fledged academic discipline, as pioneered by the likes of Tinbergen, Lorenz and von Frisch, an endeavour for which they shared a Nobel prize. Ethology was and is concerned with animal behaviour in all its forms, but given that no organism exists in a vacuum, much animal behaviour is necessarily social. A slightly more substantial difference between social evolution and ethology is methodological: the former emphasises the value of controlled experimentation, the latter observation in natural environments. This is, however, a very approximate distinction, and it would be fair to say that a department or individual researcher's preferred term is as much the result of historical factors as any clear dividing line between the two. In particular, one should certainly not draw the conclusion that since ethology values naturalistic observation it is an unscientific discipline in which one may pick and choose from a range of explanatory frameworks with few theoretical constraints. On the contrary, ethology produced, in the form of Tinbergen's four whys (Tinbergen, 1952), a framework for understanding the nature of different evolutionary explanations that was and still is recognised as fundamental to all biological investigation (Mayr, 1963). So much so, in fact, that it would be impossible to provide an introduction to social evolution without detailing, at the outset, what the four whys are and the relationships between them.

Function, mechanism, phylogeny and ontogeny can be grouped into a two-by-two grid, as per figure 1.1. One axis defines the type of explanation – either a single form or the historical/developmental process behind it. The other axis defines the level of explanation – either ultimate (why it exists) or proximate (how it works). Thus in the top-right corner we have phylogeny, which refers to the evolutionary development of the trait in question, and in the bottom-right we have ontogeny, the development of the trait within the organism. Function, in the top-left, refers to the ultimate, evolutionary level purpose of the trait – what, if anything, it was selected for – and in the bottom-left is mechanism, which addresses precisely how the functional logic is implemented.

| | | different types of explanation | |
|---|---|---|---|
| | | single form (explains the makeup of one particular form) | development /history (explains the historical process behind the form) |
| different types of question | ultimate (explains why organisms are the way they are by describing how natural selection shaped their current form) | function – what, if anything, is it adapted for? | phylogeny – what is its evolutionary history? |
| | proximate (explains how organisms work by describing structures, process of development, etc.) | mechanism – how does it work? | ontogeny – how does it develop? |

Figure 1.1: Tinbergen's four whys.

An example is perhaps necessary here (taken from Krebs & Davies, 1993). In a lion pride the females come into oestrus at the same time as each other. Why? The ultimate, functional answer is that it ensures that the pride's various litters of lion cubs are produced at the same time, and this enables communal suckling, female companionship when the young males leave the pride and other benefits that increase the likelihood that the young will survive to adulthood. A proximate, mechanistic answer to our question would be to do with how female pheromones influence the oestrus cycle of other females and hence achieve the functional objectives described. The ontogenetic answer will chart how oestrus develops through the female's lifespan, and the phylogenetic answer will do the same for the evolution of oestrus in the species. These four explanations are not alternatives to each other. They are complementary, and together they give us a complete picture of biological traits (Tinbergen, 1963). However, it would be fair to say that social evolution theory is principally concerned with the two explanations that address the current form of social behaviour: function and mechanism. A sharp line should be drawn between these two types of explanation (Mayr, 1963): although complementary, they are fundamentally different, and there are no shades of grey between the two. Explanations are either one or the other. The former are concerned with the selection pressures that shaped the trait in question, and the latter with how that selection is implemented. As figure 1.1 suggests, the difference is analogous to that between phylogeny and ontogeny: one deals with the evolutionary story; the other with the present form of the organism. (Wherever terms function and mechanism are used in this thesis it will, unless explicitly stated otherwise, be in this sense. Any additional prefix like 'Tinbergenian' is used only for emphasis.)

The distinction between the two is crucial. The conflation of function and mechanism is unfortunately commonplace outside social evolutionary circles. Consider, for example, the suggestion that our syntactic abilities are based upon the same cognitive mechanisms that must underpin reciprocity (Bickerton, 1998). The assumption here is that because reciprocity is about who did what to whom and when, then there must be some cognitive representation of the corresponding social calculus. From this assumption follows the suggestion that such representations are utilised in syntax: "[for this reason] thematic analysis must have been selected for long before the hominid-pongid split" (ibid., p.351, italics added). This is akin to saying that because ant behaviour can be described in terms of kin selection then ants must have some of the cognitive machinery required to do arithmetic (see Dawkins, 1979). In fact (of course, even) nothing of the sort is entailed by the theories of kin selection or reciprocity. (For the record, ants appear to recognise kin through smell (Billen & Morgan, 1998).) This is because reciprocity and kin selection are ultimate, evolutionary explanations of behaviour rather than proximate, mechanistic ones. As such, they are unable to say anything about how the logic they describe is manifested in any given case, though they will of course guide our intuitions in that regard. To be sure, one way to implement the logic of reciprocity would be for natural selection to install in organisms sophisticated cognitive machinery of this nature, and this may even be the case with humans. However this does not follow as a matter of course from the observation of reciprocal behaviour.

The difference between the proximate and the ultimate is often characterised as the difference between the *how* and the *why*: ultimate processes explain the evolutionary logic behind certain behaviours – i.e. *why* they exist – and proximate mechanisms explain *how* that logic is enforced. However, for some scholars outside of social evolution, mechanistic descriptions of behaviour are precisely what is sought (this is often the case in both psychology and linguistics, for example) and are therefore given the status of why. An alternative is to describe the difference as that between *process* and *product*: ultimate explanations describe the logic behind the evolutionary process while proximate explanations describe the product of that process. Put into these terms, the claim that syntactic abilities must have existed because our social behaviour is describable in hierarchical terms is seen to be a claim that a product of evolution utilises the very same machinery as an evolutionary process. When phrased this way it becomes clear that the claim is epistemically incoherent.

Of course, it may in fact be true that syntactic abilities and social cognition utilise the same psychological machinery. The point being made is not to dispute or endorse that view; it is rather that the assumption that this *must* be the case demonstrates confusion about the nature of different evolutionary explanations. Neither is this a matter of petty point scoring; such misunderstanding is sadly commonplace within evolutionary linguistics. Here are two further examples, both concerned with the question of why we are such enthusiastic talkers – an evolutionarily ultimate problem. The first is the assertion that strong reciprocity "shows how a weakness in reciprocal altruism can be overcome" (Hurford, 2007, p.477). There is no need to go into the details of exactly what strong reciprocity is; all that matters is that we note (following West et al., 2007) that it is defined (e.g. Gintis, 2000) in mechanistic terms, whereas reciprocity is a functional explanation. Hence it is not enough to say that strong reciprocity will do the work where reciprocity alone is insufficient, because such an response says nothing about why such a set-up would be stable – which is precisely the (ultimate) problem that reciprocity is sought to explain. The second example makes the same mistake; the functional problem of why we talk cannot be overcome by a "profound coalitionary restructuring [of human sociality]" (Knight, 1998, p.75) because this, too, is a proximate phenomenon. The structure of human sociality is a product of natural selection, not a process. (A note on terminology: the term reciprocal altruism is in common use, but is unfortunately ambiguous, as it suggests that the exchanged behaviours are altruistic. They are not, precisely because they form part of a mutually-beneficial exchange. Indeed, the very point behind reciprocal altruism is to explain how such a process might be evolutionarily stable. Reciprocity is therefore a better term (West et al., 2007), and will be used it in this thesis.)

So, social evolution is concerned with the derivation of corresponding functional and mechanistic explanations of social behaviour; that is, how organisms do things to one another. In the human case that includes, most dramatically, language. We now turn to pragmatics.

## 1.2 What is pragmatics?

Although the metaphor of the "wastebasket" of linguistics (Bar-Hillel, 1971, p.405) is unfairly disparaging, it is true that pragmatics is notoriously difficult to define (Levinson, 1983). Indeed, it is concerned with a range of phenomena that appear at first glance to be relatively unrelated. One view would be to say that pragmatics is the study of how more is communicated than is actually said. Another would be to list the central topics of enquiry: phenomena like deixis (the way in which reference can depend upon who is speaking), speech acts (the study of language use as a series of acts, each designed to do something) and implicature (whereby utterances can have meanings that go beyond their literal meaning). A third is to say that while syntax, semantics and phonetics are concerned with language-internal phenomena, pragmatics studies those aspects of language that are external to the linguistic system itself: "if in an investigation explicit reference is made… to the user of the language, then we assign it to the field of pragmatics… if we abstract from the user of the language and analyze only the expressions and their designata, we are in the field of semantics" (Carnap, 1958, p.79). Although this is a very approximate characterisation, it does emphasise that pragmatics is concerned with the relationship between the language and its external environment, and hence it hints at why pragmatics can be understood as being concerned, at some level, with same sort of phenomena as social evolution theory. Whilst none of these characterisations is wholly satisfactory, all are correct to some degree. In an effort to capture all of these in one definition a good first approximation is to say that pragmatics is concerned with the study of how context and use affect meaning (Huang, 2007).

If any single idea can be said to be at the core of pragmatics it is the thesis of linguistic underdeterminacy: the (often huge) gap between the literal meaning of an utterance and the meaning that is actually conveyed (Atlas, 2005; Austin, 1955; Carston, 2002; Recanati, 2004; Searle, 1979; Sperber & Wilson, 1986). For example, the utterance 'It's raining' may well mean 'Get the umbrella', 'I don't want to go out after all' or indeed any one of an infinite number of possible interpretations (Quine, 1960). Pragmatics takes seriously the existence of this gap between what are termed utterance meaning (the literal meaning of the utterance) and speaker meaning (the meaning the speaker intends to convey). One consequence of this is that we are forced to recognise the messy realities of inference and ostension (the provision of evidence that leads the listener towards a conclusion) rather than abstract away from such complications to an idealised model in which language is seen as a code to be translated. This gives rise to what is termed the ostensive-inferential model of language, which is placed in contrast to the code model (a distinction that will be discussed in §2.2 and §3.4). It also leads to one further way to characterise what pragmatics is: the study of how the gap between utterance and speaker meaning is bridged.

The discipline is sometimes said to have originated with the publication of Grice's Logic and Conversation (1975). However, its roots can be traced back to the philosophy of Pierce and other semioticians of the early 20th-century, and it would do an injustice to the work of several other scholars, in particular Wittgenstein, Strawson, Austin and Searle, to credit only Grice for the inception of the field as we know it. Nevertheless, Grice's contribution was of particular significance. It included, for example, the distinction between communicative and informative intent, which takes a central role in many theories of pragmatics and which is used, in chapters 4 and 6, to inform matters of evolutionary interest. A speaker's informative intent is, as the term suggests, the intention to inform their interlocutor of something; i.e. that the interlocutor understand the content of the utterance. The communicative intent, on the other hand, is the speaker's intention that the interlocutor understand that the speaker has an informative intent (see figure 1.2). That is, Grice recognised that before listeners can address the question of meaning they must first recognise that the speaker wishes to inform them of something. Moreover, he showed that because informative intentions are embedded within communicative intentions then if a speaker's communicative intention is satisfied then their informative intention will also be satisfied as a matter of course.

| | |
|---|---|
| communicative intent: | I intend that you understand that… |
| informative intent: | ..I intend that you understand X |

Figure 1.2: The distinction between communicative and informative intent

A second major contribution concerns the question of how listeners come to understand speaker meaning when it is underspecified, as it so often is: implicature, and indirect speech more generally, is rife in natural language. Grice suggested that speakers obey what he termed the *cooperative principle*, which can be summarised thus: "Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice, 1975, p.45). This general goal is then given flesh in the form of four maxims of conversation: quality (tell the truth), quantity (do not say too much or too little), relevance (be relevant) and manner (be clear). Although the cooperative principle is phrased prescriptively, it is actually intended as a descriptive analysis of what speakers actually do. Consequently, Grice argued, listeners assume that speakers follow the cooperative principle, and by doing so are able to solve the problem of underdeterminacy described above. Consider, for example, the following exchange (from Levinson, 1983, p.102):

A: Where's Bill?

B: There's a yellow VW outside Sue's house

When taken literally, B's utterance fails to address A's question. In doing so it thus violates the maxims of quantity (the presence of the yellow VW is more information than was requested) and relevance (what has the yellow VW to do with where Bill is?). There is, then, an apparent failure of cooperation. However, rather than draw this conclusion, the listener assumes that this is not what has occurred, and hence searches for some non-literal interpretation of B's utterance which does satisfy the four maxims.

Whilst Grice's ideas had a profound effect on the development of pragmatics, his seminal work was clearly programmatic and tentative in nature. Numerous refinements, additions and extensions of his work have been produced, but even the most high-profile instances (e.g. Horn, 1984; Levinson, 1983) take Grice's central ideas, and in particular the maxims of conversation, as the bedrock for their later work. Perhaps the most radical challenge to the maxims that has gone on to play a central role in pragmatic research comes in the form of *Relevance Theory* (Sperber & Wilson, 1986) (hereafter RT), which supplants the four maxims with a single notion of *relevance*. Since relevance is one of Grice's four maxims, a common misunderstanding (e.g. Blutner, 1998; Clark, 1996; Kopytko, 1995; Roberts, 1996) is that Relevance Theory collapses the four maxims into one, but in fact the relevance-theoretic notion of relevance is more precisely and technically defined than the corresponding Gricean maxim. The basic idea is that there is a trade-off between the cognitive effects of an utterance and the processing effort required to achieve those effects, and relevance is the degree to which that trade-off is optimised. RT asserts that human cognition is geared up to maximise relevance, and its radicalism arises from the claim that this is sufficient to capture all pragmatic behaviour – no further maxims or constraints are necessary. RT plays a major role in this thesis: the central concern of chapter 5 is to illustrate how RT captures the Darwinian realities of an evolved communication system.

Pragmatics, then, is concerned with how interlocutors do things to one another using language, and how context affects meaning. As already observed, this means that it shares with Darwinism a concern with how factors external to the focal actor influence that actor's choice of behaviour. The goal of this thesis is to take advantage of this commonality to help develop a naturalistic, Darwinian foundation for pragmatics. The next section outlines how this will be done.

## 1.3 Thesis outline

Before I outline the thesis, I should comment briefly on what it is not. Specifically, I wish to play down any expectation that this is a thesis in pragmatics. There is, for example, no analysis of corpus data, nor much discussion of, say, felicity conditions or implicatures. There is, in short, little in it that a typical practioner of the discipline would recognise as the normal material of pragmatic research. (Perhaps the nearest thing to 'traditional' pragmatics in the thesis is §5.1, which introduces RT.) Instead, this is an interdisciplinary exercise in the extension of a successful theoretical apparatus, that of social evolution, to a phenomenon that it has until now left alone – natural language. Moreover, it is conducted in the context of contemporary research in the origins and evolution of language. As we have already observed, this is a field in which there is a significant risk of conceptual confusion, and it is for that reason that much of this thesis is theoretical in nature. If we are to do science we must have a clear conception of what it is that we wish to study and what our explanatory tools can and cannot tell us, and in many ways large parts of this thesis (for example, chapter 3, which asks how we should define communication; and chapter 5, which describes the minimal constraints that a general theory of evolved communication must work within) can be seen as an attempt to be as clear as possible on these matters. The emphasis on pragmatics only arises because, as already discussed, it is, of the traditional topics of linguistic enquiry, the one that is most immediately addressed by social evolution theory. That is not to say that other aspects of language are off limits to such an enquiry, but those targets are by-and-large left for another day. Given all of this, it could be said that this thesis is more about the evolution of communication (in humans) than it is about the evolution of language.

What, then, is our agenda? As already suggested a, if not the, central question for social evolution is that of biological function: what, if anything, is a trait adapted for? This is therefore my starting point. It is frequently assumed (following Pinker & Bloom, 1990) that with respect to language the answer to this question is communication. Many computational and mathematical models of the evolution of language, for example, use communicative success as a direct measure of fitness (e.g. Batali, 1998; Hurford, 1989, 1991; Nowak et al., 2000; Nowak et al., 1999; Smith, 2004). Nevertheless, the claim is disputed, not least by Chomsky, who has argued that communication is at best a secondary function of language, and also that adaptation by natural selection is not the only explanation of design in nature. Chapter 2 examines the issues around these questions, in particular the debates between, on the one hand, Chomsky and those who have endorsed his views (Gould, Piatelli-Palmarini) and on the other hand Pinker, Bloom and Jackendoff. It concurs very much with Pinker et al. on the importance of adaptation, but asks whether the models of language and communication that are (implicitly) accepted by both sides are appropriate. Specifically, it raises the question of whether the code model (Shannon & Weaver, 1949) is the correct way to conceptualise human communication.

Chapter 3 addresses that issue. The first half of the chapter defends a definition of communication grounded in social evolution, which is then applied, in the second half, to human communication. More precisely, it is argued that communication should be defined with reference to the design of both the signaller's and the receiver's behaviour, not just because such a definition correctly captures various prima facie instances of communication and non-communication, but also because it contains within it the evolutionary logic that all stable communication systems must satisfy. It is also argued that the obvious alternative approach, to define communication in terms of information transfer, is at best derivative on this adaptationist account. It is then shown that the ostensive-inferential alternative to the code model that is central to a pragmatic conception of language follows quite naturally from this definition.

We then return, in chapter 4, to the matter of function and conclude that the biological function of language is indeed communication, but only when communication is appropriately defined. The relationship between this conclusion and two other adaptive hypotheses about the origins of language (gossip (Dunbar, 1993, 1997) and sexual selection (Burling, 2005; Miller, 2000)) is discussed, as is the relationship between the social evolutionary notion of function and the school of linguistic thought termed functionalism. We then address the immediate problem faced by any naturally selected communication system, that of evolutionary stability. If the signaller can benefit from unreliable or dishonest signals then they should evolve to do so, but if that scenario is costly for the receiver then they will evolve not to attend to the signal. Now receiving no benefit, the signaller will evolve not to signal, and the system will have collapsed in an evolutionary retelling of Aesop's fable of the boy who cried wolf. The defining question of animal signalling theory is what evolutionary processes might prevent this scenario from occurring (Maynard Smith & Harper, 2003; Searcy & Nowicki, 2007). Possible answers are critically evaluated and classified, and this analysis benefits from insights in pragmatics. In particular, the pragmatic distinction between communicative intention and informative intention highlights the difference between honesty and reliability. A high-level, three way classification of the processes that can stabilise communication is suggested, and that is used to argue that a concern for reputation is the most likely explanation of why humans are honest.

Chapters 2, 3 and 4 are thus concerned, in various ways, with the evolutionary function of language. In chapter 5 we turn to the question of mechanism, and ask how the evolutionary logic of communication is manifested in human linguistic behaviour. Using the definition of communication from chapter 3, a set of minimal conditions that all evolved communication systems must satisfy is derived. It is then shown that these constraints predict, with remarkable precision, the two principles of relevance that lie at the heart of RT, and it is thus argued that if RT itself is not correct then something with a very similar logical structure must be. It should be made clear that although this chapter provides evolutionary support for RT, it does so not by virtue of a critical assessment of current pragmatic theory, but by virtue of the development of a general social evolutionary model of communication which is then observed to map onto RT. This is in keeping with the point made above that this is not a thesis in pragmatics, but instead one that explores what a social evolutionary conception of language might look like. One review of RT, by one of the major figures in contemporary pragmatics, admits that as a discipline pragmatics is an "untidy collection of usage principles, accrued over decades of careful observation… it may be a bit ramshackle, but it delivers the goods (or at least some of them)" (Levinson, 1989, p.469). The same review also notes that RT is an alternative rather than an addition to this body of work, one which offers "a simple unifying vision" (ibid., p.456). Correspondingly, chapter 5 does not attempt to add a social evolutionary dimension to the current disorder, or even use social evolution to tidy the mess; rather, it argues that social evolutionary principles predict RT, and thus supports the "paradigm-change in pragmatics" (ibid.) that RT aspires to achieve.

In all, then, chapters 2 to 5 offer an account of precisely what social evolutionary thinking can tell us about our linguistic behaviour. Chapter 6 then uses this foundation to address a major explanandum in evolutionary linguistics – the emergence of learnt symbolism. Specifically, an experimental approach is introduced that, unlike previous work, recognises the distinction, central to pragmatics, between informative and communicative intent. The crucial novelty in this work is that there is no a priori distinction between communicative and non-communicative behaviour. Moreover, the constraint of embodiment is seen to fundamentally impact on the form of emergent communication systems. The way in which communication emerges under such circumstances provide insights into the cognition of communication and the possible process by which humans came to use learnt symbolism. Moreover, the definition of communication derived in chapter 3 is used to show that this process is not analogous to any other instance of the emergence of communication, but is in fact qualitatively different. Chapter 7, which concludes the thesis, recounts and clarifies what it tells us about the evolution of language, and suggests a number of possible avenues for future research.

If there is a general theme that runs through the thesis it is that if we define language only with reference to its internal features then we cut the link with the environment that is necessary for a definition to be evolutionarily meaningful. Of course, that is not to say that the link with the environment is all that is important about language. Far from it: there are many non-pragmatic features that mark language out as being different in some important way from other forms of communication (Hockett, 1960), and in chapter 6 we will focus quite directly on two of them – learning and symbolism. However to even approach these matters we must first give naturalistic foundation to our conception of language, and indeed convince ourselves that, from a biological perspective, language is a tool for communication. This may appear "obvious" (Blackmore, 1999, p.99; Clark, 2000, p.405; Dunbar, 1993, p.727; Hurford, 2008, p.253; Nowak & Komarova, 2001, p.288; Pinker, 1994, p.367) to many, but there are voices to the contrary. These include the pre-eminent figure of 20th-century linguistics and one of the most prominent evolutionary theorists of the last 50 years, and so the arguments should be taken seriously. The next chapter does so.

# Two

# The argument from design

Can the argument from design be usefully applied to language? Pinker and Bloom (1990) argued that it can, and thus concluded that language evolved for the purpose of communication, a claim that was later upgraded to a "truism" (Pinker & Jackendoff, 2005, p.224). It is, however, disputed by others, including Chomsky and Gould. We will eventually conclude that Chomsky and Gould are incorrect in this respect, but the investigations presented here are not redundant, since although they arrive at the same conclusion as Pinker and Bloom, the reasons for doing so are different. Moreover, our analysis will lead us to question what exactly is meant by the term communication, and once that question is answered, not only will we be able to understand in more detail exactly what is entailed by the claim that language evolved for the purpose of communication, but it will also provide us with a platform that will be used to discuss several other issues later in the thesis.

The argument from design is simply enough stated: natural selection is the only source of design in nature, and language is well designed for the purposes of communication; therefore language was selected for that reason. If we accept the two premises the conclusion must logically follow. However, both premises are disputed, and this chapter will assess those arguments in some detail. Chomsky in particular has questioned both. In fact, it would be easy to conclude, on the basis of quotes like the following (and these are by no means isolated examples) that Chomsky is hostile to any evolutionary explanation of language:

> "The processes by which the human mind achieved its present stage of complexity and its particular form of innate organization are a total mystery… it is perfectly safe to attribute this development to 'natural selection', so long as we realize that there is no substance to this assertion, that it amounts to nothing more than a belief that there is some naturalistic explanation for these phenomena." (1972, p.97)

> "It does seem very hard to believe that the specific character of organisms can be accounted for purely in terms of random mutation and selectional controls. I would imagine that the biology of a 100 years from now is going to deal with the evolution of organisms the way it now deals with the evolution of amino acids, assuming that there is just a fairly small

space of physically possible systems that can realize complicated structures… Evolutionary theory appears to have very little to say about speciation, or about any kind of innovation. It can explain how you get a different distribution of qualities that are already present, but it does not say much about how new qualities can emerge." (1982, p.23)

"Evolutionary theory is informative about many things, but it has little to say, as of now, of questions of this nature [the evolution of language]… in the case of such systems as language or wings it is not easy even to imagine a course of selection that might have given rise to them." (1988, p.167)

Such pronouncements are "baffling" (Maynard Smith, 1995, p.47) to the evolutionary biologist. When pressed, Chomsky insists that this hostility is not directed at evolution *per se*. That would be a "misreading" (Chomsky, 1996, p.41). It is more specifically at adaptationist approaches to evolution, particularly when applied to the language faculty:

"What kind of biological evolution [is language the result of]? Well, here you have to look at the little bit we know. We can make up lots of stories. It is quite easy: for example, take language as it is, break it up into fifty different things (syllable, word, putting things together, phrases and so on) and say: 'OK, I have the story: there was a mutation that gave syllables, there was another mutation that gave words, another one that gave phrases… another that (miraculously) yields the recursive property (actually, all the mutations are left as miracles).' OK, maybe, or maybe something totally different; stories are free and, interestingly, they are for the most part independent of what language is… the story you choose is independent of the facts, pretty much." (2002, p.146)

It was in response to these and similar views (e.g. Mehler, 1985; Piatelli-Palmarini, 1989) that Pinker and Bloom's influential article *Natural language and natural selection* (1990) was written. There they laid out the adaptationist case, which rests on the two simple premises of the argument from design. The first of these, that natural selection is the only source of design in nature, is utterly uncontroversial to the evolutionary biologist: "Natural selection is the only plausible explanation for adaptive design. What other explanation could there be?" (Maynard Smith & Szathmáry, 1995, p.290). However Chomsky (and indeed, as we shall see, some language evolution researchers) has appealed to two alternatives: exaptation and self-organisation. A third alternative, also popular in the evolutionary linguistics literature, is cultural transmission. Chomsky also disputes the second premise, that language is well designed for communication. All of these arguments will be discussed, but a final conclusion will be deferred, since a proper assessment of the claim that language is well designed for communication will obviously depend on what we take communication to be, which will be addressed in detail in chapter 3. In the meantime, however, we can make significant headway on the other matters, starting with the first premise. Three alternatives to natural selection as a source of design in nature will be considered: exaptation, self-organisation and cultural transmission.

## 2.1 Premise I: Natural selection is the only source of design in nature

### 2.1.1 Alternative I: Exaptation

Exaptation is the process by which some biological feature evolves a different function than that for which it was originally selected (Gould & Vrba, 1982). Chomsky has suggested this as a plausible alternative to adaptation with respect to language:

> "In some cases it seems that organs develop to serve one purpose and, when they have reached a certain form in the evolutionary process, became available for different purposes, at which point the processes of natural selection may refine them further for these purposes… possibly human mental capacities [including language] have in some cases evolved [in this way.]" (1988, p.167)

This is a view endorsed by Gould:

> "Yes, the brain got big by natural selection. But as a result of this size, and the neural density and connectivity thus imparted, human brains could perform an immense range of functions quite unrelated to the original reasons for increase in bulk… the universals of language are so different from anything else in nature, and so quirky in their structure, that origin as a side consequence of the brain's enhanced capacity, rather than as a simple advance in continuity from ancestral grunts and gestures, seems indicated. (This argument about language is by no means original with me, though I ally myself fully with it; this line of reasoning follows directly as the evolutionary reading of Noam Chomsky's theory of universal grammar.)" (1989, p.14)

> "Noam Chomsky, on the other hand, has long advocated a position corresponding to the claim that language is an exaptation of brain structure. (Chomsky, who has rarely written anything about evolution, has not so framed his theory [in terms of exaptation], but he does accept my argument as a proper translation of his views into the language of my field – Chomsky, personal communications.) Many adaptationists have so misunderstood Chomsky that they actually suspect him of being an odd sort of closet creationist. For them, evolution means adaptive continuity, and they just cannot grasp the alternative of exaptive seizure of latent capacity that is present for other reasons." (1991, p.61-62, italics added)

The claim here, then, is that the capacity for language evolved for some reason other than adaptation to the function of communication, but was later exapted for use in communication. Perhaps this is true. A common suggestion is that there is a propositional language of thought (Fodor, 1975) that could have been exapted for communicative use. This is a plausible hypothesis, and some version of it may even be likely. Pinker and Bloom endorse it in their response to the commentaries of their 1990 article, as do a number of other scholars who take the problem of language origins seriously (e.g. Bickerton, 1991; Newmeyer, 1991). Moreover, the empirical data are consistent with such a hypothesis: it is becoming increasingly clear that some animals, including many non-human primates, and especially apes, have rich and structured cognitive representations of the world (Hurford, 2007). So let us suppose, for the sake of argument, that language has exapted devices already employed in this language of thought. Does this mean that an adaptationist approach to language is inappropriate?

No, not at all. Exaptations are adaptations. Bird wings exapted heat-regulation devices, and these in turn exapted feathers, and so on and so on. Chomsky recognises this general point: "organs develop to serve one purpose and, when they have reached a certain form in the evolutionary process, become available for different purposes, at which point the process of natural selection may refine them further" (1988, p.167). He does not, however, seem to take it to its logical conclusion: that every adaptation is an exaptation at some level or another. The only difference between the two is a matter of emphasis: the former focuses on historical contingencies; the latter on adaptive design. Yet ultimately they are the same thing (Mayr, 1982). With specific reference to language, this means that "language could be a modified or an unmodified spandrel" (Maynard Smith & Szathmáry, 1995, p.290). A spandrel is a trait that has been thrust into action for a purpose for which it was not designed. It was originally an architectural term used to refer to the triangular-shaped blocks that appear at the intersection of the dome and its supporting columns as the inevitable consequence of placing a dome on a square base. They are often elaborately decorated and hence appear to be part of the architectural aesthetic, even though that is not the purpose for which they were designed (1979). But then:

> "if the claim is only that language is a modified version of a structure that once served some other function [i.e. an exaptation – TSP], the answer is an (almost trivial) yes: the claim is true of most complex structures. But if language is modified, then natural selection was the modifying force. The only interesting non-selectionist alternative would be that language is an unmodified spandrel… [However,] no serious theory has been built upon these lines." (Maynard Smith & Szathmáry, 1995, p.290, italics added)

The general point was recognised by Gould: "the same principle [exaptation] applies when we consider the specific history of any lineage… most major transitions are exaptations" (1991, p.60). (Note that this was written before the publication of The Major

Transitions in Evolution (Maynard Smith & Szathmáry, 1995); the term major transitions is been used here in a different sense than it is there, and no equivalence is suggested.) Why, then, is he (and Chomsky, and others) so vehemently against the idea that language even be considered as an adaptation? It should be stressed that Gould did not see language as a special case: he considers adaptationism an inappropriate approach for much evolutionary biology. That was the central thrust behind the landmark paper he published with Richard Lewontin, in which they coined the term spandrel (1979), Gould thought that not enough emphasis is placed on spandrels in contemporary evolutionary biology, and it is this matter of emphasis that explains his hostility to an adaptationist approach to human psychology in general (as seen in e.g. Gould, 1991).

How so? Gould was, by training, a paleobiologist, and hence comes from a field that is principally concerned with the ancestral form of organisms and which routinely deals with timescales in the millions of years. When evolutionary processes are considered over such vast distances, historical contingency inevitably announces itself as the most important influence over present form (indeed, Gould's landmark paper on exaptation (Gould & Vrba, 1982) was published in the journal Paleobiology). Viewed over shorter (but not yet short) timescales, other factors, specifically adaptation, become more salient. In general, the timescale on which we view the evolutionary process has significant implications for the relative importance we place on the various factors that contribute to the form of a given trait or behaviour (Wilkins & Godfrey-Smith, in press). For example, the question 'Why, functionally-speaking, is the human spine built the way it is?' could be answered in number of different ways, possibly all of them true. Chomsky recognises this point:

> "Is the spine the best way it could be, given the options allowed by Nature? Probably not. It may be the best way it is from another point of view: history. Maybe if you take the long millions of years through which this thing gradually evolved, OK, that's the best thing Nature could come up with, on the various paths it's taken. But if you started over and you just ask: 'What's the best possible way to construct an object

> which will have the functions of the spine?', it's alleged that you could do it much better. So it's not perfect." (2004, p.156)

This insistence that we take the longest possible view is fine – but only up to a point. If there is no ancestral form that we take as basic, then the argument collapses in a *reductio ad absurdum*: "if there were no constraints on what is possible, the best phenotype would live forever, would be impregnable to predators, would lay eggs at an infinite rate, and so on" (Maynard Smith, 1978, p.32). On a similar note, Pinker points to Haldane's observation that "there are two reasons why humans do not turn into angels: moral imperfection, and a body plan that cannot accommodate both arms and wings" (Pinker, 1994, p.358). We must constrain our domain of explanation somehow. Indeed, it is precisely the constraints of historical contingency that make most biological questions interesting; without them organisms can only be seen as suboptimal versions of the perfect creature described above. That is not to downplay the importance of historical contingency or, indeed, factors like drift that come into play when we view evolution over short timescales. The point being made is simply that *all* interesting phenomena are exaptations at one level or another, and that the term refers to the same phenomena as adaptation does, only with a different emphasis. In recognition of this, some authors, including some in evolutionary linguistics (e.g. Kinsella, in press), prefer (following Dawkins, 1982) to talk of natural selection as a process of *meliorisation* rather than optimisation.

This is a straight-forward enough point that will not have escaped Chomsky, Gould and other critics of adaptationism. Their disagreement must be more fundamental than this. Here is my best guess: the issue is one of *explanation*. Sure, we can accept that the specific shape of the human spine is adaptive given that we evolved from a quadrapedic species, but adaptation cannot on its own *explain* why the spine takes the form it does; instead much of the explanatory load must be carried by historical contingency. This seems reasonable enough, and helps explain the central thrust of Gould and Lewontin's famous paper (1979), which emphasised the role of an organism's 'Bauplan' and attempted to play down the importance of adaptation.

However, this does not challenge the main point: that whatever design work is done, it must have been natural selection that did it. The claim is not that adaptation, on its own, explains the entirety of the natural world. Such an assertion would suffer from a similar reductio ad absurdum to that which we saw above: no creature is perfect. Instead, the claim is that adaptation by natural selection is the only source of any design that we do witness. Exaptation helps account for the phylogeny of that design, and in some cases that means that it will carry more explanatory load than the process of adaptation by natural selection. However that does not make it an alternative source of design; it remains the case that whatever design work has been done must have been done by natural selection.

### 2.1.2 Alternative II: Self-organisation

Although he does not use the term himself, Chomsky has also appealed to self-organisation as an alternative to natural selection. Self-organisation occurs when a system that consists of many separate entities demonstrates order and/or design by virtue of a multiplicity of relatively simple rules followed by each of the individual entities, with no independent source of organisation or design. Any design witnessed is hence an emergent property of the system, rather than the product of any organisational power. Classic biological examples include the flight of bird flocks, the structure of ant colonies and the formation of colour patterns on the skins of animals (see Ball, 2001). The idea can be illustrated with a simple parlour game. Each player chooses, at random, one other player to be their 'predator' and another to be their 'protector'. Every player makes such a choice, and once these choices are made all the players move within the room in accordance with one simple rule: keep the protector in a direct line between yourself and the predator. Each individual should move if and only if this rule is not satisfied, and when they move they should do so in such a way as to satisfy this rule. It does not matter how far each individual is from their predator; only that the protector is directly between the two. The net movement of all the players' behaviour describes a system, one which we might expect to be essentially random. However, that is not the case; instead, a stable state is quickly reached. Rather than describe that state here, I have put it in the next paragraph so that you may guess for yourself before reading the answer. This stable state is a good solution to the 'problem' of arranging the individuals in such a way that they all have their protector between themselves and their predator. Yet it appears to be arrived at without any independent, external source of organisation or design. That is, it is self-organised. There is no designer here. Order is instead simply the incidental consequence of the behavioural rule that the players follow.

(The solution is a line. There are some quite specific conditions under which this may not happen, but they are unlikely to occur, particularly as the number of players get larger. For example, if two individuals select the same predator and are each other's protector then they will compete to put the other individual between them and the shared predator.)

This is a simple example, but the idea can be taken much further. There is now an entire industry, and at least one research centre (the Santa Fe Institute), dedicated to the creation of models that utilise self-organisation and emergence, and the phenomena that are studied in this way are not limited to biology. In particular, self-organisation has been applied in evolutionary linguistics, most profitably to the emergence of phonological structure (de Boer, 2001; Oudeyer, 2006). It is the claim of some proponents of self-organisation that it is an alternative to natural selection as a source of design in nature, sometimes with explicit reference to language (e.g. de Boer, 2001; Longa, 2001). Chomsky appears to agree:

> "There seems to be no substance to the view that human language is simply a more complex instance of something to be found elsewhere in the natural world. This poses a problem for the biologist, since, if true, it is an example of true 'emergence' – the appearance of a qualitatively different phenomenon at a specific stage of complexity of organization." (1972, p.70)

> "We know very little about what happens when $10^{10}$ neurons are crammed into something the size of a basketball, with further conditions

imposed by the specific manner in which this system developed over time." (1975, p.59)

"The answers may well lie not so much in the theory of natural selection as in molecular biology, in the study of what kinds of physical systems can develop under the conditions of life on earth and why, ultimately because of physical principles." (1988, p.167)

To understand Chomsky's enthusiasm for self-organisation it is instructive to consider his approach to language and indeed cognitive science more generally. Specifically, Chomsky's defining contribution to linguistics was to insist that language, and specifically syntax, is most profitably studied as a formal, computational system of symbol manipulation, whereby phonological input is translated into propositional form. This ultimately means that the proper Chomskyan is committed to the view that communicative function and other pragmatic aspects of meaning cannot, by definition, form any part of a formal grammar (Lakoff, 1991). More generally, the linguistic system is considered to exist independently of the environment in which it operates. The opposite thesis, that the system and the environment cannot be coherently decoupled, is axiomatic of Darwinism and is also central to the selectionist view of human psychology described by Spencer (Leslie, 2006) at around the same time that Darwin first proposed his ideas. Indeed, it was Spencer who coined the phrase 'survival of the fittest'. Spencer can be seen as the original proponent of the idea that behaviour and environment must be studied in the round, so much so that this view has been termed 'Spencerian' (Godfrey-Smith, 1993). Evolutionary biology's modern synthesis is "Spencerian to its core" (Dennett, 1995, p.394), as are empiricism and behaviourism, which Chomsky famously mauled (1959) in his review of Skinner's Verbal Behaviour (1957). In a preface to a reprint of his original review, Chomsky insists that "I do not see how his [Skinner's] proposals can be improved upon… I do not, in other words, see any way in which his proposals can be substantially improved within the general framework of behaviourist or neobehaviourist, or, more generally, empiricist ideas that has dominated much of modern linguistics, psychology and philosophy" (1967, p.142). This, like many of his other views on evolution, is as anti-Spencerian as you like. It seems natural to conclude, then, that Chomsky's antipathy to an adaptationist account of language is a manifestation of a larger antipathy to Spencerian explanations in general (Dennett, 1995), a commitment that is entirely in keeping with his work in linguistics; "hostility to evolution as an explanation of the origin of language in the species goes hand in hand with hostility to experience as an explanation of the origin of linguistic form in the individual" (Hymes, 1991, p.49). Interestingly, a visceral reaction to the idea that outside influences might shape individual behaviour can also be observed in Chomsky's anarchist political views; "central to anarchism is the celebration of spontaneity and self-organization" (Knight, 2004, p.592). External influences, then, cannot (should not?) be used to explain behaviour, according to Chomsky, and instead the supposed alternative of self-organisation is to be invoked, be it in politics or linguistics, where the influence could be evolutionary, developmental, or cultural.

However, self-organisation is not necessarily an alternative to adaptation as an explanation of design in nature. In particular, natural selection can select for genotypes that utilise self-organisation to build the phenotype. Consider again the predator/protector parlour game described above. It was described in terms of separate human players interacting with each other, but is intended as a general illustration of self-organisation. What sort of entity do the players stand for? Any. In the case of biological systems, this could mean separate organisms or parts of organisms, for example cells or neurons. In the former case this would mean that the organisms had behavioural rules (adaptive or not) that they each follow, and that their resultant geographical structure is an emergent phenomenon. In the latter case the emergent phenomenon is some part of a particular organism's phenotype, for example an organ, which is made up of the cells or neurons (or whatever) that follow the predator/protector rule. Either way, the resultant trait is still subject to selection pressure. For example, if there is some other way to arrange the cells or neurons into the required line, and if that alternative is more adaptive than the self-organised solution (for example, it may be more energy efficient), then we should expect it to evolve (subject to historical constraints, of course). In other words, the process of self-organisation is simply one way that natural selection may use to implement its desired outcome. Self-organisation is of biological interest not because it provides an alternative to natural selection but because it allows natural selection to act with a particularly light hand.

This fact illustrates a general point that is not sufficiently well appreciated: that natural selection frequently (indeed, has to) makes use of the material provided to it by the environment and the contingent make-up of the organism. As one of the pioneers of the study of self-organised systems puts it, "the structure of fitness landscapes inevitably imposes limitations on adaptive search" (Kauffman, 1993, p.118). More generally, the role of the genome in evolution is not to pre-specify every aspect of an organism's behaviour but to equip the organism with the tools necessary to survive in an often changing environment. These tools "can be seen as a set of calibration devices that help to fit an organism to its not-quite static environment" (Dickins & Dickins, 2007b, p.46) and we can imagine possible hierarchies of such tools, in which each level shifts more and more of the burden of design from the genome to the organs and other structures built by the genome (e.g. Dennett, 1995; Lorenz, 1939). So to say that some trait is selected for is not to state that every aspect of that trait is hard-coded into the genome; only that the genome has been selected to build organs (of which brains are arguably the foremost example) that make use of the material provided to them to create the traits in question (Dawkins, 1982). As an example, consider the way in which humans walk. We make great use of gravity, so much so that walking is better conceived of as controlled falling: we simply put our leg out, transfer our weight away from the stationary foot, and let gravity do the rest. Researchers in robotics call this passive dynamic walking, and have shown that this solution is more efficient and often more stable than the alternative: to place the foot, using muscular energy, in front of the body at every step (McGeer, 1990). This is presumably the reason why it is not used by any vertebrate; it is maladaptive in comparison to a solution that transfers much of the workload to physics. Self-organisation makes similar use of the environment (which includes gravity), but that does not take it outside of the Darwinian world.

However, like exaptation above, this much is surely reasonably straight-forward. Also like exaptation, Chomsky's real point may be more about, to use his own phrase, "explanatory adequacy" (1965, p.25) than about the nature of the relationship between natural selection and self-organisation. A pair of examples from the evolutionary linguistics literature serves to illustrate. Two major studies (de Boer, 2001; Oudeyer, 2006) have suggested that self-organisation may explain the structure of phonological systems. The two studies differ in several important respects, but both take the form of computational models in which phonological structure emerges as agents modify their systems so as to replicate the vocalisations of other agents. There is no selection for phonological structure in these models, yet the systems that emerge resemble those observed in natural languages, specifically in the non-random distribution of the vowel and consonant spaces. Let us suppose that these models are to some degree accurate. Does this mean that natural selection is not the source of the design? No. As the author of one of the studies observes: "These properties [of self-organisation] enable a genome to generate complex, highly organized forms without the need for precise specification of each detail in the genome" (Oudeyer, 2006, p.42, italics added). So, whereas in biology the design work is done by natural selection which can select for genomes that utilise self-organisation to build organs and phenotypes, in these models the design work is done by the scientists themselves, who specify the 'genome' that allows self-organisation to run its course. What is common to both is that the 'genome' has been selected precisely because it makes use of self-organisation. Thus whilst it may be necessary to invoke the process of self-organisation as part of our explanation, that does not take it outside the conventional neo-Darwinian paradigm; it is an enrichment of rather than an alternative to natural selection.

That is not, however, to play down the importance of self-organisation. To invoke natural selection as the sole explanation of design and to let the matter rest there is unquestionably an "incomplete and therefore unsatisfactory" (Oudeyer, 2006, p.43) strategy, since this says "nothing about the way in which natural selection was able to find such a solution" (ibid.). As before, natural selection is not (and never is) the exclusive locus of explanation, but that is not the adaptationist claim. Functional accounts are insufficient on their own; we must also "identify how the form was generated, and… understand the structure of the relationship between genotypic space and phenotypic space" (ibid., p.43-44). We have already met these two demands in chapter 1: they are ontogeny and mechanism, two of the 'four whys' that constitute evolutionary explanation. That is, how forms are generated is the question of ontogeny, and how functional design is implemented in the phenotype is the question of mechanism. Students of self-organisation in biology tend not to express the point in these terms, of course; perhaps that explains why they are sometimes prone to make over-zealous claims about the explanatory power of self-organisation. For example, the author of the other study of self-organisation in speech states that: "universal tendencies of vowel systems can be predicted as the result of self-organization, and therefore *do not need* to be explained as the result of biological evolution" (de Boer, 2001, p.37, italics added). We observed in chapter 1 that there is pluralism inherent in the four whys: the different explanations are complementary and all are necessary for a complete explanation of biological phenomena. Thus the total removal of natural selection as a possible source of explanation is as justified as the total removal of any of the other whys; that is: not at all. Just as adaptation by natural selection is not the sole explicator, neither is self-organisation. On the contrary, the two are explanatorily symbiotic: neither functional nor mechanistic accounts are sufficient on their own.

Again, however, the matter at hand is not about explanation but about the source of design. What these models show is that universals in phonological structure can emerge as the result of simple but efficient heuristics, and that the degree and type of innate specification necessary to generate such universals is not as large or as specific as might initially be assumed. Nevertheless, if the self-organised approach had proven to be maladaptive then natural selection would still be able to select against it. In other words, there is still selection here for the type of agents that can allow self-organised solutions to flourish. As noted above, in these models that selection is performed by the investigator rather than nature, but the principle is the same. So what studies like these reveal is not the independence of self-organisation as a source of design, but rather how the "imperatives of meta-engineering" (Dennett, 1995, p.222) might work. The nature of these imperatives, which take restraining forces and turn them into building tools, must be explored, but in doing so we should not lose sight of the fact that natural selection is still able to choose between these and other possible mechanisms. Just as constraints of time, material or location often serve to inspire great feats of human creativity, environmental constraints sometimes cause natural selection to find cunning solutions to design problems, with self-organisation a particularly ingenious example.

### 2.1.3 Alternative III: Cultural transmission (and individual learning)

The models of the self-organisation of phonological structure discussed in the previous chapter could arguably, depending on your definition, also be considered models of cultural evolution. Other models are more explicitly cultural. Language evolution has given rise to many studies, mostly computational (e.g. Brighton, 2002; Hurford, 2000; Kirby, 1999; Kirby et al., 2004), but more recently numerical (Smith & Kirby, 2008), mathematical (Kirby et al., 2007), and even experimental (Kirby et al., 2008) that show, quite conclusively, that cultural transmission can indeed produce the appearance of design in language without the explicit natural selection of that design. The basic idea is that the language produced by one generation is then learnt by the next, who then produce it themselves. Thus one generation's output becomes the next generation's input, and so on and so on, in a repeating cycle of production and acquisition; hence the name iterated learning model (ILM) (see Kirby & Hurford, 2002 for a review). Through this cyclical process languages can evolve from an unstructured to a highly structured state, even in the absence of apparent selection for that structure. Instead the change occurs because the agents (the ILM has typically been studied using computational models, hence the term agents, although as mentioned above human participants have also been used) do not learn from the full range of utterances possible in the language, just as human infants are not exposed to the infinite number of utterances that are possible in any given language (the poverty of the stimulus (Chomsky, 1965)). They are hence forced to make generalisations based upon the restricted language that they do hear (Hurford, 2000; Zuidema, 2003). Moreover, in an iterated model the strength of the (innately determined) preferences for generalisation makes no difference to the final language, under a range of reasonable assumptions (Kirby et al., 2007). It is consequently claimed that cultural transmission offers an alternative to natural selection as a source of design in nature. In fact, under what is known as the complex adaptive systems view of the evolution of language, it is not just culture that offers an alternative source of design, but also individual learning (Kirby, 2000, 2002; Kirby et al., 2007). Is this the correct way to conceive of the relationship between (biological) natural selection, cultural transmission and individual learning – as overlapping but essentially separate processes?

Let me address learning first. A degree of phenotypic plasticity exists in perhaps every creature on this earth; indeed, to have no such plasticity will almost certainly be maladaptive unless a species' environment is extraordinarily predictable. The contrast between innately specified and more plastic behaviours is nicely illustrated by analogy with two different approaches to espionage (Dennett, 1995): need-to-know, in which agents are told precisely and only what is necessary for them to complete their part of the mission, and commando-team, in which agents are told as much as possible so that they are as well prepared as possible to deal with any unanticipated eventualities that might arise. Which of these is the most appropriate for a given mission? That will depend on the scenarios that the agents are likely to encounter: the less predictable these are, the more suitable the commando-team option. Similarly, the more unpredictable the environment that an organism finds itself in, the more suitable it is to replace innate specification with plasticity. Moreover, there will be variability in the degree of plasticity necessary. For some missions a simple "If x then y; otherwise z" instruction will suffice; this is barely a step-up from a precise innate coding. For other missions the agents will be given much more autonomy and will be told to pursue certain high-level goals (food, orgasms, mild temperatures, etc.) whilst avoiding troublesome outcomes (pain, nausea, etc.) in whichever way they deem best, and to adjust their efforts as they go according to their relative success. These adjustments are the espionage equivalent of learning. This metaphor nicely illustrates the oft-made point that the distinction between learning and evolution is a grey one (e.g. Calvin, 1987; Dennett, 1995; Gallistel, 1999; Tooby & Cosmides, 2005). As discussed in the section on self-organisation above, it is the normal business of natural selection to build devices that are able to act in its stead, and this includes those that allow the organism to learn. If they fail to do this they will be selected against, just as those agents that do not act in accordance with the goals of their mission will lose their stripes. In short, a brain capable of learning is a proximate mechanism like any other: an organ built by natural selection to solve the unpredictable design problems posed by the organism's environment.

All of which makes learning a fundamentally different type of process from natural selection, which is an ultimate level explanation. It explains why a trait exists, while learning is part of the story of how that functionality is implemented. As such, natural selection and learning are, like natural selection and self-organisation, explanatorily symbiotic rather than being competing sources of design. Much the same applies to culture. Reconsider the results outlined above, in which cultural transmission gives rise to linguistic structure. Plainly, this process relies upon the agents' internal architectures in some way. The knowledge or skills required by each agent are often very weak, and hence do not need to be as strong as might otherwise be assumed (by the Chomskyan nativist, for example); in fact, the contribution of the agents' innate capacities can in fact be remarkably limited: small, domain general preferences for regularity are perhaps all that is necessary (Kirby et al., 2007). Nevertheless, the ability to make generalisations based upon commonalities in the meaning space is an important prerequisite. Thus whilst there is no selection in the model itself, there is still selection in the construction of the model – in fact, this is unavoidable. Now, if in a co-evolutionary model natural selection selects for those learners than have these abilities, then it is not the case that cultural transmission offers an alternative source of design, but rather than cultural transmission is a process that is utilised by natural selection in the production of adaptive phenotypes. In fact, this is precisely what happens in one recent model (Smith & Kirby, 2008). Agents were evolved biologically at the same time as their language evolved culturally and, as discussed, the agents were equipped with biases for generalisation. What the model finds is that the strength of the agents' bias is almost irrelevant to the emergence of a structured language. Almost, but not quite: the bias must be non-zero. However, once it is greater than zero its strength makes no difference to the final language. That is, once the effect of cultural transmission is accounted for, all that is required for the emergence of certain linguistic universals is that agents be equipped with a non-zero preference for generalised forms.

This result has significant implications for many issues in linguistics and cultural evolution, not to mention language evolution itself. What is germane for the present discussion, however, is that although an agent's bias for generalised forms can be domain general and very weak, it must still exist. This gives natural selection something to act on: if the cultural process is set in motion, so that agents learn from their cultural environments, but this turns out to be maladaptive, then natural selection could, in principle, select against the social learning that underpins it. Indeed, if there is a cost associated with such a bias, and if that cost is correlated with the strength of the bias, then the smallest possible bias is selected for (ibid.). So what this model finds is that natural selection selects for precisely those circumstances in which cultural transmission delivers design at the cheapest possible cost. If this is right then the difference between cultural evolutionary accounts of linguistic structure (see also Christiansen & Chater, 2008) and the accounts based upon an innate Universal Grammar is not to do with the design algorithm that underpins that structure, but rather the process by which that structure is delivered.

My point here is not to play down the importance of culture to language evolution. On the contrary, it is clearly pivotal, as results like those I have outlined demonstrate. It is essential that we better understand how cultural transmission bridges the gap between the genome and language if we are to properly explain the evolution of language (Kirby, 1998, 1999). However that is precisely what we do when we investigate any biological mechanism: explore how evolutionary logic is phenotypically implemented. In this respect cultural transmission is like self-organisation or passive dynamic walking: a mechanism that delivers adaptive ends without the need to code every detail. "Mother nature is not a 'gene-centrist'… the process of natural selection doesn't favor transmitting information via genes when the same information (roughly) can be just as reliably, and more cheaply, provided by some other regularity in the world" (Dennett, 2003, p.171). This means, amongst other things, that once a species has evolved whatever capacities are necessary to engage in the cultural transmission of a language-like communication system then "the genome may as well store the vocabulary [or indeed the syntax – TSP] in the 'cultural environment'" (Maynard Smith & Szathmáry, 1995, p.219; following Tooby & Cosmides, 1990b). In short, there will be circumstances in which natural selection selects for scenarios in which the process of cultural transmission can be relied upon to deliver the required design, and the conditions under which such a strategy might be pursued can be specified (e.g. Boyd & Richardson, 1985).

One example, of obvious relevance to evolutionary linguistics, where culture may be a particularly appropriate mechanism is in the identification of in-group members. Accents are culturally transmitted, and acquired automatically and easily early in life. Learning an accent in adulthood is far harder, which means that for most of us the only accent we can produce fluently is that of the local group into which we were born – unless, of course, we spend a large period of time with the group in question, in which case we become part of the group. Accent is thus in many ways the ideal marker of group membership (Roberts, in press): it is hard to imitate, but not impossible if enough time is spent with the in-group, and it is highly salient. Correspondingly, outsiders are more-or-less instantly identifiable; and the properties that make that identification possible depend upon the fact that language is culturally transmitted. Thus if it is true that we have evolved to use accent as a cue to group membership then it would imply two things. First, it would be an instance of adaptation to an environment that specifically includes the regularities provided by cultural transmission. Second, and this is the key point, if recognition by listeners of a speaker's group affiliation has fitness consequences for that speaker (as it surely would if the above scenario is correct), and if natural selection reacted to those changes (perhaps by some subtle modification of how accents are acquired and/or advertised) then it would be an instance of natural selection using the fact that languages are culturally transmitted to deliver adaptive design. Importantly, it does not matter for the present discussion whether or not this story is correct; what matters is that it is possible. I speculate not to propose a hypothesis, but rather to give an illustration of how natural selection might use the fact of cultural transmission to deliver design.

Viewing cultural transmission in this way changes how we view language, with significant implications for the argument from design. For Pinker, Bloom and other Chomskyans language is a biological trait possessed by all humans. Within this framework language, by definition, takes the form of innate constraints that help guide acquisition. It is a biological trait in its own right. However the view of cultural transmission espoused above allows us to conceive of it in an alternative way, along the following lines. One biological trait possessed by humans is a propensity to acquire the cultural trappings of our local environment. This includes, most obviously, grammar, the lexicon and all the other aspects of a linguistic system, but also cultural norms, taboos and, in many cases, writing and other symbolic systems (Tomasello, 1999; Vygotsky, 1986). (There may or may not be adaptations specific to the acquisition of language; either way it does not change the argument.) This trait to acquire these systems gives rise to an idiolect – an individual representation of the linguistic system of the local group members. Crucially, this idiolect is most parsimoniously conceived of not as a trait in its own right but rather a consequence of another trait – the ability to acquire a linguistic system. An analogy can be drawn with the process of imprinting. The propensity, or instinct, to imprint is an innate characteristic of many animals, and it leaves the animal with, presumably, some sort of neural representation of the other animal or thing that it has imprinted upon. Yet we would not want to say that this representation is a biological trait of the animal. It is instead a psychological and neuroscientific phenomenon, which consequently requires explanations from those disciplines. That is not to say that biology should remain silent on such matters – such intellectual isolationism would be indefensible in an interdisciplinary thesis such as this. The observation is rather that there is little that the tools of biology, as traditionally conceived, can say about it.

By analogy, an individual idiolect should be seen not as a biological trait in and of itself, but rather as a consequence of the instinct to acquire such a system (a point emphasised, ironically enough, by Pinker himself (1994)). This instinct requires biological explanation (Sperber, 1990; Sperber & Origgi, 2004), but the consequent idiolect that it gives rise to is, by-and-large, the domain of linguistics, psychology, neuroscience – and cultural evolution, which offers a (very promising) candidate explanation for the structure that idiolects clearly exhibit. Languages are then statistical abstractions over a set of individual idiolects. They must be both learnable and expressive, two pressures that pull in what are often opposite directions. We can consequently think of language as an evolving system in its own right (Christiansen & Chater, 2008; Croft, 2000; Deacon, 1997; Kirby, 1999). However, the structured idiolects that result are psychological and linguistic phenomena, more-or-less outside the purview of biology. Of course, not all of psychology is beyond the scope of biology. Indeed, part of this argument is that the instinct to acquire a language is something that biology can help explain (and more generally, it will be come clear, if it is not already, that this thesis is sympathetic to the idea of an evolutionary psychology). Nevertheless, it does not seem unreasonable to draw a distinction between, on the one hand, the psychological consequences of acquiring the cultural trappings of the social environment into which we are born and, on the other, what could be termed biologically 'basic' aspects of our psychology – a category that includes the instinct(s) to acquire cultural traits.

If this is right then language structure is not an instance of design in nature. It is, to be sure, a structured, ordered phenomenon, one that may well be explained by the evolutionary process of cultural transmission. However it is an acquired characteristic rather than a biological trait. The term *epiphenomenal* is often used to obfuscate, but here it may usefully denote the status of linguistic structure *vis-à-vis* biology. This means that the first premise of the argument from design is not threatened: natural selection explains design in nature, but not elsewhere (how could it?). It is for the same reason that the design of the computer on which I type these words is not explained by natural selection: it is the result of many basic psychological traits, but it is not part of any individual's biological make-up (not even their extended phenotype, which requires that variation in the genome correlates with variation in the entity under discussion (Dawkins, 1982, 2004)). Of course, 'in nature' is a somewhat clumsy and vague phrase, and there will likely be shades of grey: my computer is intuitively further removed from 'nature' than my idiolect. Nevertheless the distinction seems *a priori* a reasonable one.

Standard evolutionary theory, which grants to natural selection a privileged status by virtue of the fact that it is the only source of design in nature, is thus not threatened by the process of cultural transmission; or, more generally, by the existence of other forms of heritable variation. I do not intend to discuss at length the relationship between non-genetic inheritance and the natural selection algorithm (see e.g. Dickins & Dickins, 2007b; Mameli, 2004; Odling-Smee et al., 2003), but within the context of the present discussion it is instructive to consider a recent exchange in the journal *Behavioral and Brain Sciences*. The target article (Jablonka & Lamb, 2007), a précis of the book *Evolution in Four Dimension* (E4D) (Jablonka & Lamb, 2005), argued that genes are not the only source of heritable variation in life, the others being epigenetic (e.g. cellular), behavioural (e.g. learning) and symbol-based (e.g. culture). One reply pursued a similar line of argument to that presented above: that since natural selection selects the mechanisms that are involved in non-genetic inheritance then these other inheritance systems do not challenge natural selection as the ultimate source of design. The title of that reply captures the point nicely: Designed calibration: Naturally selected flexibility, not non-genetic inheritance (Dickins & Dickins, 2007a). In reply the authors say:

> "Dickins & Dickins claim that because non-genetic inheritance systems, which they see as mechanisms for calibrating genetic responses, have been designed by natural selection, neo-Darwinism, with its focus on genetic adaptation, can retain its privileged position in evolutionary theory. We certainly agree that there is no alternative to natural selection as an explanation of complex adaptations. However, what we tried to do in E4D was not to challenge the explanation of adaptation by natural selection, but rather, to extend the scope of natural selection, and offer a more inclusive framework for thinking about the origin of selectable heritable variation" (Jablonka & Lamb, 2007, p.379, italics added)

I read the phrase 'complex adaptations' in this quotation as roughly synonymous with 'design in nature', in the sense discussed above. Furthermore, I do not think that the phrase natural selection is being used as the more general sense of a universal Darwinism (Dawkins, 1983; Hull, 1988) that could be applied to cultural variants, since a central thrust of the target article is that Lamarkism (the inheritance of acquired characteristics) is common within the inheritance systems under discussion. It seems, then, that the crucial point that natural selection is the only source of such design is accepted, even by those who advocate that we should pay more attention to other forms of inheritance, including cultural transmission. There is no doubt that a better understanding of such processes will enhance our understanding of evolution, and that is particularly true in the case of language structure. However that does not challenge natural selection's "privileged position in evolutionary theory". Indeed, it is recognised (e.g. Laland & Sterelny, 2006) that the phenomena that are emphasised in frameworks like niche construction (Odling-Smee et al., 2003), evolutionary developmental biology (Carroll, 2005) and non-genetic inheritance can be captured by the standard theory (see e.g. Rice, 2004 and other citations in Laland & Sterelny, 2006). What these approaches seek to do is to expand on how that theory works, and for that they are to be welcomed (but see Dawkins, 2004 for arguments why this pluralism might ultimately muddy the waters).

However the key claim, that natural selection is the only source of design in nature, is not challenged by these frameworks, nor by the process of cultural transmission.

For the Chomskyan, language structure simply is a biological phenomenon. Other views are possible. In particular, we can consider structure to be epiphenomenal, and the various studies of the impact of cultural transmission discussed above point towards this latter view. This difference also has implications for the second premise of the argument from design, that language is well designed for communication. We will now see how.

### 2.2 Premise II: Language is well designed for communication

As was the case with the first premise, Chomsky's hostility to the argument that language is well designed for the purposes of communication is clear:

> "Language is designed as a system that is 'beautiful', but in general unusable. It is designed for elegance, not for use, though with features that enable it to be used sufficiently for the purposes of normal life… the system is elegant, but badly designed for use." (Chomsky, 1991, p.49)

> "[Language] seems to have many properties, including some deeply rooted in its basic design, that make it dysfunctional, unusable, although adequate for actual use." (ibid., p.50)

> "Language design as such appears to be in many respects 'dysfunctional', yielding properties that are not well adapted to the function language is called upon to perform… there is no reason to suppose, *a priori*, that the general design of language is conducive to efficient use." (Chomsky, 1995, p.162)

> "Language is not properly regarded as a system of communication… it can of course be used for communication, as can anything people do – manner of walking or style of clothes or hair, for example. But in any useful sense of the term, communication is not the function of language,

> and may even be of no unique significance for understanding the functions and nature of language." (Chomsky, 2002, p.76, italics in original)

The previous section's analysis allows us to make some sense of Chomsky's claims: language could be so much better designed, if only humans existed on a different peak in the adaptive landscape. The vertebrate eye contains a basic design error: the optic nerve and the wires that form it are in front of rather than behind the light cells themselves, a flaw that produces blind spots (Dawkins, 1986). We earlier saw that Chomsky has made the same point about the vertebrate spine (2004). Language is, according to Chomsky, dysfunctional in much the same way: it is simply not how you would design an object that is called upon to perform the task(s) that are demanded of it. Pinker's response to these views, this time alongside Jackendoff, is incredulous:

> "These claims are, to say the least, surprising. At least since the story of the Tower of Babel, everyone who has reflected on language has noted its vast communicative power and its indispensable role in human life. Humans can use language to convey everything from gossip, recipes, hunting techniques, and reciprocal promises to theories of the origin of the universe and the immortality of the soul… Moreover the design of language – a mapping between propositions and sound – is precisely what one can expect in a system that evolved for the communication of propositions." (Pinker & Jackendoff, 2005, p.224)

Pinker, Bloom and Jackendoff's argument is, in short, that language is well designed to "encode propositional information" (Pinker, 2003, p.27). Pinker and Bloom (1990) offer up a number of specific examples of this, including lexical categories, which allow us to distinguish basic ontological categories; rules of linear order that allow us to distinguish between predicate and argument; and verb affixes that signal the temporal distribution of events. Several of the commentaries on their article took issue with these claims. In particular, Piattelli-Palmarini, a then colleague of Chomsky's at M.I.T., argued that grammar is littered with poor design. He gives two specific examples (the first of which is the same as one of the examples offered by Pinker and Bloom as good design).

Piattelli-Palmarini observes (following Dowty, 1990) that although some simple temporal states (past, present, future) are typically marked syntactically, there are other temporal states for which this cannot be done without recourse to "elaborate, strained, prolix periphrasis, supplemented with lexical pointers" (Piatelli-Palmarini, 1990, p.753). The specific example he gives is the simultaneous occurrence of two events at some point in the future, which is not properly expressed by any of John will discover that Marcia is pregnant; John will discover that Marcia will be pregnant; John will have discovered that Marcia (is) (was) (will be) (? will have been) (has been) pregnant; or any other even moderately elegant construction. Language's inability to express such an idea in a straight-forward way immediately shows, so the argument goes, that it is not as well designed as could be. Piattelli-Palmarini's second example is conditionals and contracts, which "map onto a desperately mixed syntactic bunch" (p.754, italics in original). This would be unexpected if, as Pinker and Bloom contend, one of the main evolutionary reasons to transfer propositional content was to manage social relations. Two further examples come from an earlier article (1989). First, why does no recorded language form questions or passives by inverting the order of phonemes? This would be a simple and elegant, and therefore probably adaptive, solution to the problem of how to turn a declarative into a question, but it is not used. Second, why do humans struggle with sentences with multiple centre-embedded clauses? Again, this is a simple and elegant solution to the respective design problem, as testified by its universal use in computer languages, but humans are poor processors of such structures.

Such idiosyncrasies force us to question whether or not language is well designed for the transfer of propositional content. Piattelli-Palmarini thus issues Pinker and Bloom with the following challenge: "how inadequate (how dysfunctional) must a structure be before an adaptationist admits that it cannot have been shaped by the proposed function?" (1990, p.753, italics in original). Pinker and Bloom's reply is to observe that there are always higher peaks somewhere, just not here, nor anywhere nearby: "how adequate (how functional) must a structure be before an anti-adaptationist admits it was shaped by the proposed function… even the most adapted structure will be unable to do some things. One could just as easily argue that legs are dysfunctional because humans are slower than a speeding bullet, less powerful than a locomotive, and unable to leap tall buildings in a single bound" (1990, p.773, italics in original). They have a point: as noted earlier, we must place some constraints on the size of our adaptive landscape, otherwise the claim that a trait is suboptimal becomes vacuous. It has recently been suggested (Marcus, 2008; following Clark, 1987) that virtually all of human cognition is suboptimal is some way or another, but, as we observed in §2.1.1 above, this is trivially true; what matters is whether a trait is suboptimal given the constraints of evolutionary history. On the other hand, Pinker and Bloom's response does somewhat sidestep the issue (Botha, 2003): we cannot just assume that Piattelli-Palmarini's suggestions for how syntactic structure could be better designed are forever cut-off because of historical contingency. Equally, however, we cannot assume the contrary either. Ultimately we simply don't know whether these are small steps or giant leaps (for mankind). Neither are we likely to know anytime soon, if ever.

All of which could be said to leave us at an impasse. However, we have allowed the participants in the debate to dictate the grounds on which it takes place, and with that has come a set of assumptions about the fundamental nature of language. Both parties take a broadly computational view, in which language is considered to be, at its core, a symbol-manipulation device that translates between internal meaning and phonemic form. Indeed, though Piatelli-Palmarini disagrees, quite vehemently, with Pinker and Bloom with respect to the relationship between natural selection and language, he does "find it refreshing not to have to argue for linguistic [i.e. Chomskyan – TSP] facts" (1990, p.752). This leads one to adopt what is commonly known as the code model of language, and to equate communication with information transmission. Pinker and Bloom's text (1990), for example, is littered with the terminology of codes and algorithms. They characterise language as "the transmission of propositional structures" (p.707) that "map… meanings onto pronounceable and recoverable sounds" (p.776). The following passage is representative: "the vocal-auditory channel has some desirable features as a medium of communication: it has a high bandwidth… however it is essentially a serial interface… the basic tools of a coding scheme employing it are an inventory of distinguishable symbols and their concatenation" (p.713, italics added). The basic idea, whose roots can be traced back to Aristotle (see Sperber & Wilson, 1986) is that communication can be conceived of as a process whereby two information processing devices (human brains) directly map internal meanings into external signals in both production and reception. In this picture to encode or to decode an utterance is to perform an act of machine translation, in which a lexicon is searched for the meaning of each of the utterance's constituents, and these meanings are then combined to form the meaning of the utterance. A similar process, in reverse, accounts for production. The defining formulation of this model is Claude Shannon and Warren Weaver's Mathematical Theory of Communication (1949), captured in figure 2.1, which sowed the seeds of information theory and is still the dominant paradigm of communication in artificial intelligence and associated disciplines. More specifically, there is a wide if implicit assumption amongst many linguists, including the protagonists in the present discussion, that this is a reasonable way in which to conceptualise communication.

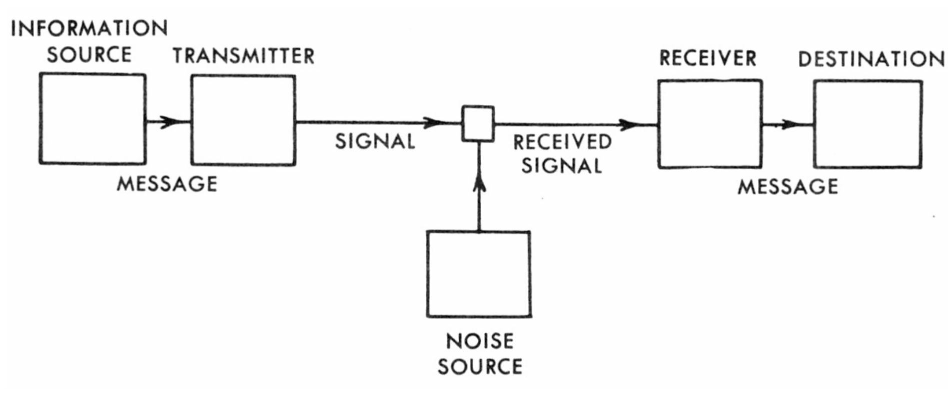


Figure 2.1 (reproduced from Shannon & Weaver, 1949): The code model of communication. Communication is here an act of machine translation. Dominant within artificial intelligence, this model is widely assumed by many, sometimes knowingly, sometimes not, to be a suitable account of linguistic communication. The information source is one human mind which then encodes the internal, semantic message that is to be communicated into phonetic form. That is then transmitted via the medium of speech or gesture and decoded by a corresponding receptive device.

Over the past 40 years or so pragmatics has developed and refined an alternative to this picture (Austin, 1955; Grice, 1971, 1975b; Origgi & Sperber, 2000; Schiffer, 1972; Sperber & Wilson, 1986). Briefly mentioned in §1.3, the ostensive-inferential model of communication posits that communication is achieved through the production and interpretation of evidence for the meaning that is to be communicated. The act of production is called ostension and the act of comprehension inference. The evidence is provided through the physical alteration of the shared environment (i.e. by speech, gestures, or whatever other medium is used), an act that triggers the inference of the intended meaning. For example, if I offer my girlfriend a cup of tea in the morning, she may respond "I've already cleaned my teeth". In doing so she provides evidence that she wishes to decline my offer, yet she does not say as much explicitly. Neither must the evidence necessarily be verbal, or even linguistic: I can gesture towards a friend's newly arrived plate of chips, and in doing so provide evidence of my desire to have one of the chips myself. Different pragmatic theories (see Huang, 2007; Levinson, 1983 for surveys) disagree about precisely how this communication is achieved, but all agree that production is ostensive and that its goal is to induce a particular change in the listener's mind; and that comprehension is inferential and its goal is to discover the speaker's intended meaning.

If this picture is correct, and the next chapter will argue that it is, then the question to ask about language is not how well designed is language for the purposes of encoding and decoding, as Pinker et al. and Piatelli-Palmarini do, but rather how well designed is language to allow the construction and interpretation of evidence for some belief or another. This is a subtle but important difference. For one thing, it leads us to focus not on language but rather on "languaging" (Becker, 1991, p.33); that is, the act of using language itself. The question for the argument from design is thus not 'Is linguistic structure well designed for information transmission?', but rather 'Is the capacity for language well designed for the twin acts of ostension and inference?'. If this is correct then arguments about the adaptive status of various aspects of syntax are at best premature. Note also that this question is, in general, the focus of animal signalling theory; that is, of social evolution theory as it is applied to communication. The defining question is not whether the structure of, say, the Tungara frog's croak is adaptive, but rather why they croak at all, and what is achieved by doing so (Maynard Smith & Harper, 2003). Of course, phenomena like the peacock tail and the red deer roar (both of which will be discussed in chapter 4) show that answers to this question often require that we examine signal form, but these discussions are framed by the matter of function, not the other way around.

Our discussion of the adaptive status of language is thus put on hold. In order to continue we must establish a position on what our model of language is to be: is it to be built upon principles of information transmission and code, or upon principles of ostension, inference and use? There are many arguments from within pragmatics as to why the latter model is to be preferred, and my discussion, which follows in the second half of the next chapter, will touch on several (but by no means all) of these. However the central argument, developed in the first part of the next chapter, is original to me, and is conducted in the spirit of cross-disciplinary synthesis: I will first defend a general definition of biological communication, and then I will ask what that definition entails for human communication in particular. We will see that the pragmatic approach follows quite naturally from a social evolutionary view of communication.

# Three

# What is communication?

The defining goal of this thesis is to approach language, and particularly pragmatics, from the perspective of social evolution theory. To accept that such a project has intellectual merit is to accept the desirability of consilience, the interlocking of causal explanation across disciplines (Wilson, 1998). This perspective is sometimes criticised as excessively reductionistic (e.g. Rose, 1997) or even as an attempt "nihilate" (Berger & Luckmann, 1966, p.133) the higher-order disciplines. Such claims are based on the mistaken idea that reductionists wish to make all disciplines that explore human behaviour, from history to economics, "subservient" (Rose, 2007) to biology. There are many eloquent replies to this (e.g. Dawkins, 1986; Dennett, 1995; Weinburg, 1992; Wilson, 1998), and so I will not regurgitate those arguments here, except to repeat one simple point (following Dennett, 1995): that there are both bland and preposterous readings of the reductionist claim. The bland reading is simply that it is possible to unify different disciplines. This is trivially true: I am a biological (not to mention social) entity, bound by the laws of physics and chemistry. The preposterous reading is that the reductionist wishes to abandon the theories, terminology and laws of the higher-order disciplines and describe the phenomena of interest in purely lower-level terms. (Probably) nobody holds this view. The agenda of consilience is to use the tools of the lower-level disciplines to account for the tools of higher level disciplines, and consequently to converge upon cross-disciplinary explanation. More than any other part of the thesis, this chapter, in which I will construct a working definition of communication that can be applied across the natural and social sciences, is conducted in that spirit. It begins with a definition from evolutionary biology, but that definition will, by the end of the chapter, have been modified so that it is applicable across a range of explicanda that extend beyond evolutionary biology into psychology and linguistics, and perhaps even further.

The definition with which we will begin is taken from Maynard Smith and Harper's book Animal Signals (2003):

> "We define a signal as any act or structure which alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved." (p.3)

This is a definition predicated on the selection pressures behind signals and their corresponding responses. Even within the biological sciences, not all accounts of communication share this emphasis; in particular, the absence of any mention of *information*, which we have already seen is an implicit assumption of many in linguistics and elsewhere, strikes some scholars as surprising (Stegmann, 2005). Indeed, across a range of disciplines that includes biology, "the concepts of *information* and *signal* form integral components of most definitions of communication… [and] we tend to think of biological signals as conveying or carrying information" (Hauser, 1996, p.8, italics in original). The lack of consensus on what form a definition should take has led several authors (e.g. Grafen, 1990; McGregor, 2005; Slater, 1983) to express pessimism that any single definition will be sufficient. On the other hand, others refuse to accept that such a situation is inevitable: "There must be something wrong [with the present ambiguity]. There must be a way of defining 'communication'" (Dawkins, 1995, p.75). I share this view, and this chapter attempts to provide just such a definition. This is surely desirable: agreement on technical terms is in general essential for scientific progress (Brown, 1983; West et al., 2007), and the utility of a rigorous definition of communication is plainly far wider than the uses to which it will be employed in this thesis. To pick just one example, the predictions made by biologists about the behaviour of organisms will depend in part on whether or not they think communication occurs (Diggle, Gardner et al., 2007; Diggle, Griffin et al., 2007). My analysis will also offer precise definitions of concepts associated with communication: *signal*, *response*, *cue* and *coercion*.

The first half of this chapter will show that the definition given above (slightly amended for clarity; see below) works for animal communication not just because it correctly account for various *prima facie* instances of communication and non-communication, but also because it captures within it the evolutionary logic that must underpin communication. Once this definition has been defended, and in particular once it has been shown that this is preferable to an account predicated on information transfer, I will suggest a way in which it can be generalised to include human utterances, which, since they do not themselves evolve, do not appear to be addressed by the definition above as it presently stands. I will then explore whether or not the modified definition appropriately fits our intuitions about human communication, and reflect on how it ties in with pragmatic and psychological accounts of communication. We will find that it fits very tightly with the ostensive-inferential model outlined in the previous chapters. This is partly because it emphasises, like the definition above upon which it will be based, that in communication interlocutors *do* things to one another.

## 3.1 The adaptationist account of biological communication

Although the definition quoted at the outset of this chapter provides the inspiration for the present analysis, it is in fact not quite precise enough. For that reason, the following will be used instead:

> A signal is any act or structure that (i) affects the behaviour of other organisms; (ii) evolved because of those effects; and (iii) which is effective because the effect (the response) has evolved to be affected by the act or structure. Communication is the successful completion of a signalling act.

The only substantial difference between the two definitions is the explicit requirement that the response not simply be adapted but be adapted to fulfil its half of the communicative dynamic. That much is implied by the other definition, but it is best that such criteria be made explicit. Moreover, this particular criterion is necessary to address one of the concerns that I will shortly discuss.

This definition is the logical conclusion of a number of adaptationist definitions that stretch back many years. Such definitions take as their starting point the idea that a signal must influence the behaviour of another organism. That much is not sufficient on its own, as it does not distinguish between, say, a push and a signal that tells another organism to move. This is why several authors (e.g. Hasson, 1994; Krebs & Dawkins, 1984; Shettleworth, 1998) have added a further criterion that the behaviour be selected in order to influence the other organism. It is here that adaptation is introduced into the definition. However, the receiver's role in the interaction is not yet given the prominence required. Communication is "neither the signal by itself, nor the response, it is instead the relationship between the two" (Wilson, 1975, p.176, italics added). We should therefore demand that the roles of both signaller and receiver be necessary for communication to occur (although Wilson's own definition only demanded that the signal be adaptive for either one or both participants). That is, there is no communication if either signaller or receiver is missing from the equation. The natural idea that a signal is something that is 'projected' into the world to be 'consumed' by possible receivers is apt to lead us astray, in that it renders us liable to forget or neglect the fact that for a communicative dynamic to exist in the first place there must first be both a signaller and, less obviously, a receiver. This point will be re-emphasised when we discuss how this definition can be made applicable to human utterances.

This two-pronged nature of the definition of a signal is why one landmark paper (Krebs & Dawkins, 1984) was able to describe animal communication as the result of an evolutionary arms race. In an earlier piece (Dawkins & Krebs, 1978) the authors had focused on the signaller's behaviour and whether it is better described as information or manipulation. In the later article, following the logic of their work on evolutionary arms races (Dawkins & Krebs, 1979), they promote the receiver's role from passive receptacle to active player. Specifically, they characterise the receiver as 'mind-reading' the signaller, as a counter-balance to the signaller's 'manipulation'. Communication thus becomes the result of an arms race between these two behaviours. This conception of animal communication is often said to assume that all communication entails conflict of interest (e.g. Grafen, 1990; Haven-Wiley, 1983; Shettleworth, 1998; Slater, 1983) which is plainly not the case. Nevertheless, its underlying logic is the same as that which is implicit in our definition. Although there have been isolated examples (e.g. Dusenbery, 1992; Grafen, 1990; Lewis & Gower, 1980; Maynard Smith & Harper, 2003) of definitions that demand that the interaction should be adaptive for both parties, such a 'full' definition is yet to be commonly adopted. One exception to this trend is the community of computational modellers of communication (Bullock, 1997; di Paolo, 1997; Noble, 1998; Oliphant, 1997) who have converged upon a definition equivalent to that suggested here, but phrased in terms of causality. I will discuss this in more depth later, when I come to ask how the definition might be applied to humans.

Note that the symmetrical format of the definition allows for a straight-forward distinction between communication, cues and coercion. This can be seen in figure 3.1 (Diggle, Gardner et al., 2007). The second clause is used to distinguish a signal from a cue (Hasson, 1994), an example of which would be, say, height: the size of a predator could be useful information for prey, and may hence cause an effect in the prey, but we would not wish to say that the predator signals its height to the prey. (This would be doubly true if the predator had not even seen the prey, but had only been seen.) We may draw a similar distinction with the third clause: coercion refers to the scenario where the receiver is not adapted to respond to the signal. One example would be the push/tell distinction mentioned above. Another would be camouflage: the camouflage affects the behaviour of the other organism (predators are less likely to eat camouflaged prey) and the camouflage evolved for that reason (Hasson, 1994). However there is no evolved response (Maynard Smith & Harper, 2003). The predator is hence coerced not to pursue the prey. The predator may, of course, have evolved heuristics to determine the nature of, or how much time to spend on, its search for (camouflaged) prey. However, this behaviour is not a response to the camouflage – it (presumably) did not evolve to be affected by the camouflage. This example serves to illustrate why we should demand that the response be not just evolved, but evolved specifically to form one half of the communicative dynamic. The same requirement is also necessary to properly address three concerns that have been or might be raised with the definition.

| | action evolved to trigger reaction? | reaction evolved to be triggered by action? |
|---|---|---|
| communication (signals; responses) | Y | Y |
| cue | N | Y |
| coercion | Y | N |

Figure 3.1 (adapted from Diggle, Gardner et al., 2007): The relationship between communication, cues and coercion.

### 3.1.1 Concern I: The impossibility of an unreceived signal

One possible concern is that the insistence on a dyad does not allow for the possibility of an unreceived signal. Although I have not found an instance of this concern in the literature, it is worthy of brief discussion because it makes clear the distinction between signalling and communication, and brings the interacting notions of signals and responses into even sharper focus. The concern is that if signals have, by definition, a corresponding response then what are we to make of scenarios in which something we would like to call a signal is produced but where the intended receiver pays no attention or does not notice the purported signal? On the definition under consideration it would appear that we cannot call this a signal, however much it may be intuitive to do so. However, we may say that the signal will induce the response on average. That is, although a signal may on occasion fail, perhaps due to noise or poor design, on average it will induce a response and will hence undergo selection. In other words, failure is unexpected but possible.

It is this potential failure that marks the distinction between communication and signalling, which are otherwise synonymous. Both refer to the production of a signal that is designed both to utilise and to be utilised by the receptive capabilities of another organism. Signalling, however, may occasionally fail, despite its design. Communication, in contrast, refers to the successful completion of a signalling act. This implies that there is no such thing as failed communication, only failed, that is, unreceived, signalling. This distinction also allows us to pin down the corresponding definition of a response: an act (or structure) that is caused by a signal and is adapted to be so. More precisely, a response is any act or structure that: (i) is the effect of some act or structure of another organism; (ii) evolved to be affected by that act or structure; and (iii) is affected because the other act or structure (the signal) has evolved to affect this act or structure.

### 3.1.2 Concern II: Substantive effects are too liberal – the case of reciprocity

The second and the most subtle concern with the adaptationist definition of communication is that any definition based on substantive effects rather than information will be too liberal. An example where this liberalism might start to bite is reciprocity (West et al., 2007; following Grafen, 1990): the exchange of otherwise altruistic behaviours (Trivers, 1971). How might such behaviours satisfy our definition? A behaviour that will form one half of a reciprocated pair affects the future behaviour of the other individual. We may also suppose that it evolved to do just that (after all, in and of itself it has negative fitness consequences, so it is unlikely to be the result of drift or other non-selective evolutionary forces). Finally it would seem, at least at first glance, that the 'response' – the reciprocation – evolved as part of the same dynamic. Yet we would not, prima facie, want to term reciprocal behaviours communicative.

However, a more careful examination of this case shows that it does not satisfy the third condition. This is because the selection pressure that acts on the 'response' is not, contra intuition, to have one's behaviour altered by the 'signal', as required by the definition. Rather, it is to reverse the dynamic and alter the other player's behaviour. In other words, reciprocation is not communicative but actually consists of the iterative coercion of each other's behaviour. This argument requires some expansion. In a reciprocal relationship each behaviour carries a cost to the actor and a positive benefit to the receiver (Trivers, 1971; West et al., 2007). The roles of actor and receiver are alternated, so that the series of payoffs for each participant alternates between positive and negative. Now, let us examine in detail the situation after each positive payoff, where one participant has received a benefit from the action of the other, and is now due to pay a cost. What selection pressure(s) induce the individual to perform the costly action that is required to maintain the reciprocal relationship?

The answer to this question is not that the selection pressure is to reward the other participant's cooperation. That is a costly act, and will be selected against. Instead, natural selection must be sensitive to more than the immediate payoffs (which are negative), and somehow account for the medium- or long-term expected future benefits of the relationship (which are positive). (The obvious way in which this could be done would be if individuals are able to remember the past behaviour of others (Silk et al., 2000) and/or learn from others' experiences (Enquist & Leimar, 1993).) This expectation that the act will produce future benefits for the actor is not the same selection pressure as one that rewards the other participant. On the contrary, sensitivity to expected future benefits entails that the individual induce the other participant into a further iteration of the relationship.

To be clear: there are two possible selection pressures. One is to reward the other participant. The other is to induce the other participant into a further act of altruism. The first of these cannot be stable; the second may be (subject to other criteria, specifically that the net payoff to each participant be positive). Yet the criterion that the response be evolved to be affected by the signal corresponds to the first of these selection pressures, and is consequently unstable. Instead, in order that the relationship be stable the 'response' must be under a pressure to induce certain behaviours in the other participant (namely a further iteration of the relationship). Therefore the costly act in each iteration of reciprocation does not satisfy the third condition of the definition of communication. Reciprocity is not communicative.

Note that this case illustrates the need for the third clause in the definition to explicitly demand that the response be adapted to utilise the signal, and not simply be adapted. That is, the question about whether a behaviour is a response to a signal must be answered in terms of selective forces, and hence function, rather than in terms of a proximate, mechanistic cause and effect relationship. This is another point that will be of importance later, when we discuss how the definition might be applied to human communication. Only if the selective forces behind the behaviour are to be affected by the purported signal can we term the interaction communicative. Instead of asking "Under what circumstances has evolution caused the reaction to be produced?" we must ask "Why is the reaction selectively advantageous?". In most cases these two questions will lead to the same answers, but there will be instances, like that of reciprocity, when they do not. As a result, without this criterion reciprocity would indeed satisfy the definition. Note also that this analysis implies that reciprocity may be characterised as the iterative coercion of one another's behaviour, since it satisfies the first two clauses of the definition of communication but not the third.

### 3.1.3 Concern III: Ambiguity in conflict scenarios

The final concern is that the definition's third condition, that the receiver's response should have evolved, introduces a degree of ambiguity in scenarios where the participants' interests conflict (Stegmann, 2005). This is supposedly because whether or not the receiver responds to the signal will depend on whether or not they are likely to emerge from the encounter victorious or defeated. The argument is illustrated with the example of male red deer, who roar at each other as a form of competition for females (Clutton-Brock & Albon, 1979). Those that roar less frequently and with greater formant dispersion eventually give up and leave the females to those deer that roar more frequently and with lower formant dispersion (Reby & McComb, 2003). Uncontroversially we can accept these roars as signals on the adaptationist account. But sometimes roaring (and the associated parallel walking) does not settle the contest. In this situation the two males engage in a fight, with potentially high costs: they push against each other head-on. One male will push the other back and at this point the loser retreats. It has been argued (Stegmann, 2005) that the criterion that the receiver's response has evolved renders the signalling status of examples like this ambiguous. On the one hand, the pushing behaviour of the eventual winner does not appear to be a signal because it does not change the behaviour of the losing deer, which continues to push as hard as it can. Any backwards movement is due not to the result of a change in the loser's pushing behaviour, but instead to coercion by force. On the other hand, if the loser makes an active retreat then the pushing behaviour of the winner has made a difference to the loser's pushing behaviour, since the loser has ceased to push. Hence, whether or not the pushing behaviour satisfies the definition appears to depend upon the eventual outcome of the contest, a plainly unsatisfactory situation.

This reasoning is not quite right, and here is why. Both animals wish to maximise their payoff from the encounter. That payoff includes not just the benefits of access to females or territory, but also the minimisation of the costs required to realise those benefits. Hence their first interaction is not to fight each other, but to display to each other, in a cost-free way, some correlate of their fighting ability. By doing so, the two protagonists may draw conclusions about who would win a fight (and hence who will get the access to the females) without having to engage in the fight itself. This is one reason why dogs display their teeth and not, say, their tails: the former are indices of fighting ability but the latter are not. Similarly, the formant dispersion of a red deer's roar is reliably (negatively) correlated with its size (Clutton-Brock & Albon, 1979), and large deer tend to win fights against smaller deer. It is therefore in the interests of the deer with the greater formant dispersion to retreat. It is only if neither party is willing to back down that the deer begin to push each other. In effect, this scenario emerges only when both parties consider the potential payoff to outweigh the cost and uncertainty of a fight. Hence, the deer lock horns.

This decision may be reversed at any time. Specifically, if it becomes clear that the other deer is a more able fighter, as it does when one deer is pushed back by the other, then the losing deer should choose to retreat from the scene. Crucially, this retreat is *not*, contra the concern in question, an evolved response to losing the fight. Rather, it is an evolved reaction to the new information that the deer has obtained by engaging in the fight: that the other deer is the stronger. The distinction between this information and the defeat itself is subtle but important. The bottom line is that any animal that is adapted to its surroundings will depart any scene if the inputs to its system suggest that it should do, communication or not. The red deer's retreat is a behavioural strategy that is independent of whether or not the scenario is communicative. Moreover, to suggest that the retreat is a response to defeat is to suppose that the defeat *preceded* the retreat. In fact, the retreat is precisely what marks the defeat as being so; the deer's retreat *constitutes* its defeat, and as such it is analogous to a performative speech act. In sum, then, the pushing is not a signal even on the full definition. Instead it is coercive, and the concern that the third condition introduced unnecessary ambiguity is unfounded. On the contrary, the third condition is entirely necessary: as this discussion has shown, it allows us to exclude troubling examples like conflict scenarios, reciprocity and camouflage from our definition. Furthermore, as already noted, the third condition gives a dyadic form to the definition that mirrors the essential nature of evolved communicative dynamics. Nevertheless several authors (e.g. Hauser, 1996; McGregor, 2005; Smith, 1977) have noted that many definitions of communication invoke the transfer of information instead. The next section examines the merit of such an idea.

### 3.2 The informational account of communication, and other alternatives

Here are three examples of the informational definition of communication in the biological literature: "behavior that enables the sharing of information between interacting individuals" (Smith, 1977, p.2); "the transfer of information via signals sent in a channel between a sender and a receiver" (Hailman, 1977, p.52); and "the behaviors by which one member of a species conveys information to another member of the species" (Kimura, 1993, p.3). There is, for some, an intuitive sense in which any adaptationist account of signalling and communication is "incomplete" (Stegmann, 2005, p.1016) since the content of the signal appears to be "missing" (ibid., p.1015) and the definition "has nothing explicit to say about the [signal]'s information itself" (ibid., p.1017). Hence if our intuitions are any guide then perhaps it is desirable to invoke information, even if we do not need to. Indeed, information does appear to offer something that the adaptationist definition does not: specification of a given signal's content at a very fine-grained level (for a wide range of examples and extensive discussion see Bradbury & Vehrencamp, 1998). Yet the logical conclusion of adaptationist accounts of communication is to eschew the notion of information altogether and instead focus only on behaviours and adaptation (Dawkins & Krebs, 1978). This does not imply that information is not part of communication; only that it need not (and, as we shall see, should not) be definitional.

The argument, in short, is that in order to specify exactly what information is transferred we first have to specify the signal's function, which brings us immediately back to the adaptationist definition. Let us flesh this claim out. To begin, we must define information. Appeals to intuition are insufficient here if our goal is to derive a precise definition of communication, so we cannot simply invoke some everyday sense. More formally, information can be considered to be equivalent to a reduction in uncertainty (Shannon & Weaver, 1949). But uncertainty of what? Such a definition is incomplete without a statement about what is relevant. A simple thought experiment illustrates the problem. What uncertainty is reduced if I receive an email from a friend with instructions about where to meet for dinner this evening? Most obviously, there is a reduction in the uncertainty of where I should expect to find my friend this evening, but there is also a reduction in uncertainty about many other things: the correct functioning of my computer, the correct functioning of my email program, the correct functioning of the satellites that carried the message in its digital form, and so on. Hence although a signal most certainly conveys information, exactly what information it conveys is determined at least in part by the receiver: if I wish to know about the correct functioning of my email program (it may have recently played up on me, say) then my friend's email is informative in that regard; if, on the other hand, I have no concerns about my email program then the email is not informative (in that regard). This is all, of course, in addition to the information about dinner that forms the body of the email, which we assume to be relevant.

But why do we assume that the body of the email is relevant? Only because it would be unusual indeed for my friend to email me about where to meet if we had no previous plans to meet at all, or if my friend knew I was busy elsewhere, or if, even, my friend was busy and could not meet me. In short, we assume that the signal is in some way designed to achieve a particular purpose. Other aspects of the signal (like, for example, its appearance in my email program in an apparently normal manner) may carry other sorts of information; specifically, they may carry information that was not part of the signaller's design of the signal (i.e. about the normal functioning of my email program). As I will discuss shortly, these sorts of inferences are the human equivalent of a cue. Therefore in order to argue that communication entails the provision of information we must also specify what is relevant about the signal, otherwise we are left with a definition so broad that it includes cues: traits or behaviours that are informative but are not designed to be so. How do we define what it means to be relevant? Anticipating our discussion of relevance in chapter 5, we can say that relevant signals are those that are of use to the receiver; non-relevant signals those that are not. And of course, if the receiver has nothing to gain from the signal then we should expect the receiver to evolve to ignore the signalling behaviour. Correspondingly, the signaller will cease to perform the behaviour. Hence relevant signals are those that are adaptive for the signaller to perform and for the receiver to respond to. If the communication is not in the interests of one or the other organism then any signalling or responding behaviour will be maladaptive: even if there are no direct costs to being ignored (as a signaller) or drawing false conclusions (as a receiver) then the behaviour will still be maladaptive rather than neutral because of opportunity costs and wasted energy (Maynard Smith & Harper, 1995).

We have now arrived back at the non-informational, adaptationist definition of communication: the signal must be designed to have an effect on the receiver, and the receiver's response should also be designed to utilise the signal appropriately. In other words, when we flesh out a definition of communication based upon the transfer of information then we see that it collapses to the non-informational adaptationist definitions that we already know are sufficient to define communication. Or to put it another way: the adaptationist definition necessarily incorporates the informational definition. If we insist that information is transferred in communication then the best we can say is that it is an emergent feature of communication, and "not an abstraction that can be discussed in the absence of some specific context" (Hauser, 1996, p.6). Consequently an informational account that does not discuss the evolutionary function behind the signal necessarily brings with it a degree of underspecification, since we cannot even specify what the informational content of a signal is until we have specified what the signal is for.

As an example, consider fluctuating asymmetry (FA), the deviations from perfect symmetry exhibited by almost symmetrical structures. It has been suggested that FA is a reliable proxy measure of homeostatis (Mather, 1953), and hence that organisms that exhibit low FA should be more attractive than those with high FA scores (Møller, 1993). Is symmetry a signal? In the absence of any behaviour that brings attention to symmetry, the intuitive answer must be no. But on an information-based view the answer will often be yes: FA carries information, that information is of value to the receiver (the potential mate), and conveying that information would be beneficial for at least some senders; specifically those with low FA scores. In contrast, FA does not qualify as a signal on the definition defended here unless there is a specific display that brings attention to it. (In this case we may even wish to argue that the display is the signal, rather than the FA itself, though whichever we prefer the argument is not changed.) In the absence of a display, FA should instead be seen as a cue (Maynard Smith & Harper, 2003). This example illustrates the general problem with information-only definitions: they do not specify how or why a signal is relevant, and hence they have no consistent means by which to distinguish between signals, cues and coercion.

Moving on, another historically popular view is that signals should be conspicuous, exaggerated or stylised. This idea arises from early ethological work on how signals first emerge (Huxley, 1966; Tinbergen, 1952), which emphasised the fact that signals are likely to be derived from other behaviours and that once they have obtained their new function are likely to increase in salience so as to achieve the desired communicative effect (the roots of this idea can be found in Darwin's principle of antithesis (Darwin, 1889)). However, there are plenty of subtle examples, observable only with great care or detailed video analysis, that will fall foul of this criterion (Dawkins, 1995). Indeed, economy of effort is often noted to be a common feature of animal signals, an observation that comes into direct conflict with the idea that signals will evolve to be highly salient (ibid.). Relative salience is therefore an inappropriate feature with which to define signals and hence communication.

Other authors have suggested refinements to the adaptationist approach, despite its internal logic. Such proposals are likely to lead to a degree of confusion. Indeed, the stated aim of Maynard Smith and Harper's book, was "not to report any new facts [but rather]… to bring order out of chaos" (2003, p.1), and their discussion of the definition of an animal signal suggests that they see these proposed alternatives as a route to "semantic confusion" (p.6). They criticise two specific suggestions. The first is that the notion of a cue be limited to behavioural traits, a move which would allow for a distinction between cues and signs (Galef & Giraldeau, 2001). Whilst there are interesting differences between physical and behavioural traits, a term is nevertheless necessary for those features that affect behaviour but are not functionally designed for this purpose. That role is presently performed by cue, and is widely understood as such. Moreover, this distinction is, as we have seen, utterly crucial to a rigorous definition of communication; without it we fail to represent the nature of communicative dynamics and are thus left, inevitably, with ambiguous examples like FA. A distinction between the physical and the behavioural would be of value, but redefinition of an established term like cue seems the wrong way in which to achieve this, particularly in a field that already has a large degree of semantic inconsistency (Maynard Smith & Harper, 2003). The same problem bedevils the suggestion that the key distinction between a signal and a cue is not whether or not it has evolved for the given effect in question but is instead the combination of two other features: first, that cues are always 'on' while signals are switched between 'off' and 'on'; and second, that once a cue has been produced it costs nothing more to express it while the same is not true of signals (Hauser, 1996). This plays to our intuitive idea that signals are projected into the world in some way, rather than being consistently observable features. Yet it is unsatisfactory not only because of the semantic confusion that it is likely to induce but also because there are several examples that do not fit at all easily into this new framework. For example, under this definition fixed badges of status are cues while coverable badges are signals. Yet some supposedly fixed badges can be more or less prominently displayed, depending to the environmental context (Maynard Smith & Harper, 2003). Another example would be aposematic colouration. In some species, for example monarch butterflies, this is permanently 'on', while in others, for example octopuses, it is switched between 'off' and 'on'. Yet we would surely wish to use the same term for both examples. Neither is there an escape route offered by insistence that the overt display of such traits be considered the signal, for no such display takes place in the case of monarch butterflies. Similar criticisms will apply to most and perhaps all refinements of the adaptationist approach. This is because that approach contains within it the logical form of the communicative act, and hence any refinements will necessarily be deviations from that form. Once that occurs, ambiguous examples are likely to arise.

## 3.3 A consilient definition of communication

We have, then, a definition of biological communication that has stood up to detailed analysis. This is a desirable outcome in its own right: communication is an important biological concept that demands an agreed definition. However, it has been conducted as part of a wider project regarding the functionality of natural language. The next step, therefore, is to evaluate the argument from design in light of the clarity provided by our definition. Yet we are apparently faced with an immediate and obvious problem: individual utterances do not themselves evolve, and hence it is not immediately clear how the definition we have derived might apply.

It turns out, however, that this is not the major problem it first appears. Consider any animal signal; the red deer roar, say. Do the individual roars evolve? Of course not; although we are happy to talk about the evolution of the red deer roar, what we mean is the evolution of the capacity and instinct to roar. Applied to human utterances that same shorthand would lead us to talk about the evolution of human utterances when what we actually mean is the evolution of the capacity to produce utterances and the pragmatics of doing so. Whilst the shorthand is essentially harmless in the case of animal signals, to insist upon using it for humans as well is to invite misunderstanding. I therefore suggest that we replace evolved for in our definition with the more general notion of design. As discussed in the last chapter, the ultimate designer of traits and behaviours is natural selection, and therefore the substitution makes no difference to the suitability of our definition; it is mere convenience, although it is important that we recognise the shorthand, and keep it in mind. What the substitution allows is for us to think about communication in terms of the proximate sources of design, which may be more explanatorily satisfactory than to think only in terms of ultimate function. In the case of linguistic utterances that proximate source will often be human cognition. Indeed, and as will shortly be discussed, the definitions of linguistic communication that have arisen from pragmatics, psycholinguistics and the philosophy of language are predicated on our ability to intentionally design both the fact and the form of our utterances.

We have, then, the following definition of communication, which I will draw on heavily in the remainder of this thesis:

Communication occurs when an act or structure:

(i) produces a reaction in another organism;

(ii) was designed to produce such a reaction; and

(iii) is able to do so because the reaction is designed to be so

Coercion refers to the situation where only conditions (i) and (ii) are satisfied, and cues to when only conditions (i) and (iii) are satisfied. If only condition (i) is satisfied, and (ii) and (iii) are not, then the interaction may be termed accidental. These relationships are captured in figure 3.1 above, but with 'evolved' replaced by 'designed'.

Some examples and non-examples may be of use at this point. If I tell my friend to meet me in George Square at 7pm this evening, and she does indeed go to George Square at 7pm as a result of my instruction, then communication has occurred. I have altered my friend's future behaviour; my utterance was designed to do that (why would I bother to tell her to meet me in George Square if I didn't want her to go there – and note that this applies even if my motives are nefarious); and my friend's arrival in George Square at 7pm is a designed reaction to an utterance that tells her that I expect to meet her at that time and place. In general, we can tell a similar story about all utterances produced under normal circumstances.

Here is a more complex case. Suppose that I am fed up that my flatmate never does his washing up; so much so that one day after dinner I do only my own washing up and leave his on the side, rather than do it all, as is usual in our flat. Let us further suppose that my intention is that my flatmate understands that I am fed up with his laziness. Is this communication? If my flatmate comes to appreciate my frustrations then criterion (i) is satisfied. Criterion (ii) is satisfied ex hypothesi, and criterion (iii) is satisfied because my flatmate's new attitude is a designed reaction to the stimulus I have provided. This scenario is therefore communicative. Note, however, that if my intention was not just to alert my flatmate but also to cause him to do more washing up in future then criterion (ii) is satisfied only if his newfound appreciation is translated into action. The communicative status of this example thus depends on the exact nature of my goals, and whether or not they are achieved. This example thus reinforces the point made earlier, in the discussion of reciprocity, that questions about the communicative status of actions and reactions can only be answered with respect to their ultimate function.

Now suppose that I actually tie my friend's hands to the sink for the evening, against his will. Initially he refuses to do his washing up, but later does so out of boredom, for there is nothing else for him to do. This scenario is not communicative but coercive, because criterion (iii) is not satisfied: although his reaction might be said to be designed (washing up may be the best available way to allay his boredom, say) it is not designed as a reaction to being tied up; it is instead a reaction to having nothing else to do. The difference is between a reaction to the stimulus itself and a reaction to the consequences of the stimulus. Again, this echoes an important point made earlier, and stated again here in more general terms: the correct question to ask about potential signals and responses is not "Under what circumstances is such behaviour produced?" but instead "Why is the behaviour performed?".

To complete the range of possible scenarios, imagine now that I have left my flatmate's washing up out not because I was frustrated with him, but because I was short of time and needed to leave the flat (but that I found time to do my own washing up). My flatmate might draw from such a scenario the sort of inferences discussed above. Now although criteria (i) and (iii) may be satisfied, (ii) is not, since the purported signal (me leaving my flatmate's washing up out) was not designed to make my flatmate come to conclusions about my attitude to his washing up. It should thus be classified as a cue, albeit one that has triggered an incorrect reaction in my flatmate.

To summarise, we have discussed three possible reasons why my flatmate might do more washing up as a reaction to some behaviour of mine. In the first case, I left his washing up out with the intention that he recognise this as a sign of my frustrations, and he did more washing up as a consequence. This is communicative. In the second case I tied my flatmate's hands to the sink, and although he did more washing up this was because he was bored. This is coercive (although note that if I had tied my flatmate's hands so as to make him realise my frustrations, and that he had done more washing up having appreciated my situation, this would count as communicative). Finally, I have left my flatmate's washing up only because I am in a rush, but he has drawn (false) inferences from this that I am annoyed at him. In this case the left washing up acts as a (misleading) cue. I submit that these accounts match our intuitions about the communicative status of each of the scenarios.

One final example is merited, as it illustrates the important point that the proximate source of design need not be the same for both actor and reactor; all that matters is that the designed action and reaction be focused on each other. Vervet monkeys have three different alarm calls, one for each of their main predators: snakes, leopards and eagles (Seyfarth et al., 1980). Yet even when every member of a vervet group is making the eagle alarm call, suggesting that they are all aware of the presense of the eagle, vervets will continue to call. But who to, if every member of the group already knows about the eagle? It seems that the call is an essentially automated response to the presence of the eagle, and that vervet cognition does not filter call production. On the other hand, the reception of the calls does seem to be modulated by vervet cognition: vervets will survey the immediate area the first time the call is produced, but subsequent repetitions are ignored (see Tomasello & Zuberbühler, 2002 for a review). Hence the proximate sources of design are different for production and reception (Seyfarth & Cheney, 2003; Tomasello, 2008), but this does not pose a problem for our definition; all that matters is that the action and the reaction are designed, and are focused on each other. This of course raises the question of how we determine whether or not a (re)action is designed. For many animal signals this will require a historical argument, which we can make with the usual tools of evolutionary biology, in particular experimental manipulation. When human intentionality becomes involved we must appeal in part to our intuitions. This is not entirely satisfactory: appeals to intuition are in general poor guides. However the danger may be limited in this case, precisely because humans are intention-detectors par excellence. We are so well attuned to the differences between deliberate and incidental behaviours that we cannot help but interpret actions in functional terms (Csibra & Gergely, 2007). This should come as no surprise; it is precisely what we should expect from a heavily social species equipped with sophisticated cognitive abilities (Dennett, 1987).

We have, then, a generalised definition of communication applicable across the natural world. In a sense it is actually more than a definition of communication. Technical terms specific to a particular discipline will inevitably be laden with the paradigmatic assumptions of that field, but when taken elsewhere those assumptions may not be shared. Behavioural ecology, from which the Maynard Smith and Harper definition originates, places great emphasis on the organism's relationship with its environment, and this is reflected in the definition we have obtained. However, other disciplines (linguistics is a good example) may not so readily see the need for this emphasis. Thus for researchers in such disciplines the account above may appear to do more than simply state what communication is; it also posits that there are certain requirements that must be satisfied if communication is to occur. As such it could be seen as more of a theory than a definition.

One possible criticism of that theory is that it is circular: signals and responses are explanations for each other. This is true, but it is a strength of the theory rather than a weakness since, as discussed above, this synergy is inherent in communication. There can be no signal without a response, nor a response without a signal (Wilson, 1975), and so it is desirable that they are explanations for each other. An alternative way to capture this mutual dependency is to refer to causality (Oliphant, 1997): an interaction is communicative if it is both 'exploitative', in that the reaction has in part been caused by the expectation that the stimulus would only be produced under certain circumstances, and at the same time 'manipulative', in that the stimulus has been produced under the expectation that it will cause the reaction. In my terminology to exploit is to cue and to manipulate is to coerce, but the idea is the same. However, I have reservations about the terminology employed here. Manipulation and exploitation echo the notions of mind-reading and manipulation that were briefly discussed in the discussion of animal signals (Krebs & Dawkins, 1984). Whilst the internal logic of that account is entirely correct, the language employed carries with it a suggestion that all communication involves a conflict of interest, which is plainly not the case. Indeed, it seems inherently unsatisfactory to capture linguistic communication as, simultaneously, manipulation on the part of the speaker and exploitation on the part of the receiver. It is better to see these twin acts as a pair of designed behaviours each of which presupposes the existence of the other. This does of course raise something of a chicken-and-egg problem. This will be discussed and explored at length in chapter 6.

This chapter began with a definition of communication, one that is in active and fruitful use within evolutionary biology (e.g. Diggle, Gardner et al., 2007; Diggle, Griffin et al., 2007), and defended it against a number of specific concerns. We have now generalised it so as to include human utterances, and we have seen that it now accounts well for various *prima facie* instances of human communication and non-communication, just as the original definition did for animal communication (Stegmann, 2005). We can now return to the unresolved matters from the last chapter: what does this definition mean for our conception of language? Once we have an answer to this we will be able to properly assess the argument from design.

## 3.4 Models of language

At the end of the last chapter it was noted that in order to address the question of whether language is well designed for communication we had to first ask what precisely we meant by communication, and what that means for our models of language. Two broad possibilities were outlined: the code model and the ostensive-inferential model. To recap: the code model considers production and reception to be twin acts of machine translation, whereby internal meanings directly map onto external forms, while the ostensive-inferential model considers communication to involve the production and interpretation of evidence for the intended meaning. The definition of communication we have derived in this chapter points towards the latter characterisation with remarkable precision. This section will discuss why that is the case, and thus pave the way for the first section of the next chapter, which will revisit the issues left hanging at the end of the last chapter.

The ostensive-inferential model is a construct of pragmatic approaches to human communication. That is, the ostensive-inferential model follows quite naturally from the Gricean view that meaning is an intention on the part of the speaker to achieve a particular effect in the mind of the listener, via the production of evidence for that meaning. This notion of meaning is often termed speaker meaning to distinguish it from linguistic meaning, which, roughly speaking, refers to the literal meaning of a word: for a noun, the thing it refers to in the world. (In everyday use meaning is often used for both speaker meaning and linguistic meaning, and is hence frustratingly ambiguous, but separate terms are used in French (intention vs. signification), German (Gemeintes vs. Bedeutung) and Dutch (bedoeling vs. betekenis), to pick just three (Clark, 1996).) More specifically, Grice posited (1975; see also Searle, 1969; Sperber & Wilson, 1986; Strawson, 1964) that for a speaker to mean something by a stimulus they must intend: (i) that their production of that stimulus induce a reaction in their audience; (ii) that their audience recognise that this is their intention; and (iii) that this recognition at least in part motivate the reaction. A linguistic signal is then the use of a stimulus to achieve a particular speaker meaning (Clark, 1996). In other words, linguistic signals produce reactions in listeners (the first criterion of our definition of communication; see §3.3); these reactions are intended by the speaker (the second criterion); and they occur at least in part because the listener allowed it to be so (the third criterion). Moreover, the Gricean intention at the heart of this account of communication is a "curious" (ibid., p.130) intention in precisely the same way that the type of adaptation necessary to define animal communication is curious: their dependence on interactivity means that these notions cannot be played out without both the signaller and the receiver's participation. Indeed, psychologists (e.g. Clark, 1996; Tomasello, 2003a) often emphasise that communication is a participatory act: whilst it is obvious that a signal cannot be received without a signaller to produce that signal in the first place, we should also remember that we cannot signal without a receiver.

How should we interpret this convergence? The definition of communication defended in the first section of this chapter is one predicated on adaptation. As such, it sits at the ultimate level of explanation, and represents a top-down approach to the study of human communication. Pragmatics and psycholinguistics are, in contrast, very bottom-up disciplines, in that the general (but by no means exclusive, of course) modus operandi is to analyse utterances on a case-by-case basis and build our theories based on those observations. What we have found is that the top-down approach predicts, with remarkable precision, the foundational assumption of pragmatics, namely the Gricean perspective on meaning. As such it provides social evolutionary support for that enterprise, but more than that the convergence is evidence that both approaches are on the same track as each other. In chapter 5 we will explore what else our definition of communication predicts for pragmatics. However before then there is unfinished business from the previous chapter: what does this result mean for our model of language, and thus for the argument from design?

The Darwinian agenda has thus led us, quite naturally, to an ostensive-inferential perspective on language. Is such a model a sufficient characterisation? That is, does it capture everything we wish it to, or are additional tools necessary? Whilst it is plain that the recognition of intentions is necessary in communication, Grice's "greatest originality" (Sperber & Wilson, 1986, p.25) was to suggest that this is indeed a sufficient characterisation; that so long as there is some way to recognise a speaker's intention then communication is possible. If this is so, then should we abandon the code model altogether? To do so we would have to show that the phenomena that we typically think of as code can be described solely in terms of ostension and inference. Here is how that can be done. Consider the linguistic meaning of a word to be common ground: knowledge shared by the interlocutors, and known to be shared. Ostension then becomes a matter of using a stimulus in conjunction with the contextual environment to invoke that common ground. Linguistic meaning becomes, on this account, a phenomenon that emerges from ostensive-inferential communication in the limit, when context provides only a tiny (but importantly non-zero; see below) fraction of the stimuli necessary to infer speaker meaning. In sum, the ostensive-inferential model subsumes the code model (Hoefler, 2008).

This perspective requires some expansion. Communication is a problem of coordination – to converge upon a shared understanding of the speaker's intended meaning. It is solved through the use of conventions: regular, arbitrary common ground that is used as a coordination device for a recurrent coordination problem (Lewis, 1969). Chapter 6 reports an experimental investigation into how humans converge upon such conventions, but the key point for now is that they do, and that they use them in communication (linguistic or otherwise). Whenever we produce an utterance we invoke these conventions and hence the common ground that underpins them. The listener must then establish how these conventions contribute to speaker meaning. As such, the listener's task is at least in part necessarily inferential, and this is the case even for the most conventional of linguistic acts (greetings, for example). Yet once we allow this much it is hard to see where the line should be drawn between code on the one side, and ostension and inference on the other. Communicative acts form a spectrum with respect to the degree to which they are conventionalised, from not at all (e.g. the early stages of ontogenetic ritualisation (Tomasello & Call, 1997), a process I will discuss in chapter 6) to almost entirely, for example greetings. The same is true of the individual components of these acts, and words in particular: spontaneously produced novel words are not conventionalised at all, while proper nouns are highly so. But even here inference occurs: if I am in the company of two friends both named Tom, then when one of them says /tom/ code will be insufficient for me to deduce whom they wish to address. The task may be a trivial one (if they are looking directly at me, for example) but it is still inferential in nature. Context is always necessary, even if it is just to make clear that the situation is 'normal'. The intuitive idea that at some level the linguistic meaning of a word or other linguistic stimulus can be isolated from the contextual environment lies at the heart of the code model, but it is a mistake: "a ready answer to the empirical question [of] whether [or not] the meanings that, in general, a word or linguistic device may serve to convey form a small finite set… [is] a resounding no" (Origgi & Sperber, 2000, p.152). It is for this reason that linguistic meaning can only be roughly characterised as the 'literal' meaning of a word, since that literal meaning is inevitably conventional, and hence just as subject to contextual variation as any other piece of shared knowledge. One implication of this is that the traditional assumption that context simply provides an adjustment or embellishment of the decoded forms has the picture somewhat backwards. Instead "interpretation should be viewed from start to finish as a heavily inferential process that is merely constrained by encoded material" (Wedgwood, 2005, p.52; see also Carston, 2002).

None of this is to deny the existence of reliable associations between linguistic constructions and possible referents in the world. Neither do I question that these associations need to be explained, in particular from an evolutionary perspective. Rather, the claim is that it is unnecessary to consider these associations as different in any qualitative way to other sorts of common ground that we routinely make use of in communication. The only difference is quantitative. If this is correct then there should exist 'semi-words': items that lie on a continuum between linguistic and non-linguistic common ground. Examples could include vocalisations like "meh!", which could be given a number of different glosses, not just in different contexts but even within the same context: "I don't know", "I don't care", "I'm not interested", "I have no idea how to respond to that", and so on (and which, ironically enough, was announced as a new entry into the Collins English Dictionary just as this thesis was to be submitted). The general picture is nicely summarised as follows:

> "The context-dependent nature of linguistic production and understanding entails, among other things, the inevitable underspecification of linguistic forms. Language does not hold or 'convey' [linguistic] meaning per se, but simply provides cues for [speaker] meaning production in context. A conceptualization occurring in a specific instance of language use is evoked by the linguistic forms used, but is necessarily far richer than any information specifically

> associated with those forms; such information… is merely an abstraction from experience or use of the forms" (Kemmer & Barlow, 2000, p.xxi)

(Note that the term *cues* is not being used here in the same sense as I have defined it above. Here it is being used roughly synonymously with *stimuli*; a conclusion that, incidently, I can only draw inferentially). Given the convergence of adaptationist and pragmatic approaches to human communication it will come as no surprise to note that that this account of linguistic communication shares an important parallel with our analysis of animal communication: models and definitions predicated on information transfer and code are ultimately dependent on design and functionality. This is because they exclude the local environment and hence have no logical way to exclude the large (and arguably infinite (Quine, 1960)) number of obvious misinterpretations. The issue runs deep: how to interpret information in a contextually-sensible way is a (*the*?) major question in the philosophy of artificial intelligence, where it goes by the name of the *frame problem* (McCarthy & Hayes, 1969). Our definition of communication and the ostensive-inferential model of communication both overcome this problem by reference to the design that underpins each communicative act.

We have, then, a working definition of communication and a sense of what that definition means for our model of language. We are now ready to return to the matters left hanging at the end of the last chapter. What does our account of communication mean for the argument for design and hence for the adaptive status of language? It is to these matters that we now return.

# Four

# Function and evolutionary stability

To recap (briefly): there are two premises to the argument from design with respect to language. First, that natural selection is the only (ultimate) source of design in nature; and second, that language is well designed for the purposes of communication. In chapter two we concurred with Pinker and Bloom on the first point. On the second point we observed that a proper answer requires that we first establish exactly what is meant by communication. In the last chapter we concluded that communication involves organisms doing things to one another, and hence that the correct model of language is an ostensive-inferential one. This means that with respect to the argument from design, the crucial question is not whether language form is well designed for the purposes of information transfer (which was the issue at the heart of the Pinker/Bloom vs. Piattelli-Palmarini disagreement) but rather whether our productive and receptive abilities are well designed to provide and interpret evidence in favour of the meanings that are to be communicated, and whether we do so at appropriate times. In short, are humans pragmatically fluent?

The answer to this is an (almost trivial) yes: the very existence of language is dependent upon it, and every successful use of language testimony to our abilities (Origgi & Sperber, 2000). We might even argue that the very existence of a discipline of pragmatics is evidence in favour of pragmatic fluency; otherwise what would there be to study (ibid.)? Indeed, it is hard to understate the degree of pragmatic homogeneity across human populations. To be sure, some people speak more or less than others, and there is some individual variance in the degree to which speakers are attentive to their listeners, but that variance is dwarfed by the space of possible pragmatic behaviour, most of which never occurs. When it does, for example if we produce genuinely irrelevant utterances in everyday conversation, it is so deeply unusual that it is actually quite baffling to audiences (Dessalles, 2007). We are thus led to the conclusion that our pragmatic abilities are the product of a perfectly ordinary instance of natural selection, and that, correspondingly, they are an adaptation for the purposes of communication, appropriately defined. (I will not continue to repeat this final clause. Hereafter the term communication is used in the sense described in the last chapter.)

This does not mean, it should be noted, that communication is or ever was the only selection pressure acting on our ostensive and inferential abilities. As was discussed in §2.1.1 on exaptation, it is in fact trivially true that at some point the selection pressures acting on what was to become these abilities would have been different – it is just a matter of going back far enough. In particular, there is nothing in the picture presented here that excludes the idea that language also helps and helped shape our mental life (Fodor, 1975; Hurford, 2007). All that is suggested is that at some point natural selection acted in a positive way on our capacity to do things to one another with words. Whilst it is likely that certain aspects of language perform other functions as well (for example internal representation), this does not change the fact that our pragmatic behaviour is well equipped to engage in acts of ostension and inference. Beyond this, however, there is little we can say with any confidence (Fitch et al., 2005), a point that will be expanded upon in §4.1.1 below.

It is worthwhile to note that this view of 'language', as a tool with which we do things to others and hence to the world, is more common amongst psychologists and biologists than it is amongst (many, but certainly not all) linguists who, following Chomsky, treat language as "an abstract core of computational operations" (ibid., p.180), essentially isolated from the outside world (note how this isolation is perfectly in keeping with Chomsky's refusal, discussed in §2.1.2, to allow a role for the external environment in his explanations of language). If we were to accept such a definition it would indeed be quite natural to conclude that communication is not a major function of language. However this approach to language and syntax is a theoretical commitment more than it is an empirical observation (Lakoff, 1991). Indeed, the latest incarnation of the Chomskyan paradigm is evolutionarily implausible (Kinsella, in press). The theory-driven approach employed here suggests that those who take the evolutionary perspective seriously should converge upon a definition of 'language' that is more in line with the traditional biological/psychological view than the Chomskyan one. Syntax should obviously not be excluded from such an account, but our models of how syntax might work should be built upon this pragmatic foundation, a view very much in line with the functionalist and cognitive alternatives to Chomskyan generativism.

The view of language developed in the last chapter also has obvious echos of one major area of pragmatic study: speech act theory. Indeed, one of that field's defining texts is entitled *How to Do Things with Words* (Austin, 1955), with the emphasis very much on the *do*. The central point is that we should conceive of utterances not as things that we say but rather as things that we do. This is most clearly illustrated with what are called *performatives* – utterances that utilise verbs like *surrender* or *declare*, in which the very act of production fulfils the idea they express. Such utterances clearly *do* things (with words). On first glance performatives appear to differ from what are called *constatives*, in which statements are made, but the distinction collapses upon detailed analysis (Austin, 1955; Searle, 1969). To illustrate, consider the utterances "I find you guilty" and "You did it" (examples taken from Levinson, 1983). Although we might argue that the former is a performative and the latter a constative, both have the same effect if uttered by someone with the authority to judge; that is, they both *do* the same thing. Using such examples what speech act theory shows is that once we open the door to the most obvious performatives there is no principled way in which we can later close it to other utterances. We must therefore accept that *all* utterances do things, and also, following the arguments presented in the previous two chapters, that this is the biological function of language.

The story does not end there, however; our goal is not simply to examine whether or not this is the case, but to ask what this might mean for our theories of language, and of pragmatics in particular. The next chapter will take up that task. Before then there are a number of issues related to the matter of function that must be addressed. First, what is the relationship between the view developed thus far and other proposed adaptive scenarios for the emergence of language. Second, what is its relationship with the school of linguistic thought called functionalism, which argues that linguistic form can only be explained with reference to linguistic function? These two questions will be explored in the first section. The remainder of the chapter is then dedicated to the problem of evolutionary stability, which arises from a simple functional analysis of communication: if dishonesty or unreliability pays, then what keeps speakers from pursuing such behaviours? The second section explores the problem in some detail, and in particular highlights the distinction between honesty and reliability. These two terms have previously been used synonymously, but the case of human communication, informed by insights from pragmatics, highlights an important difference between them. The third section discusses the various solutions to the problem, and the fourth asks which most likely applies to humans. We will find that the answer is in fact a straight-forward one: we are honest and reliable because a reputation for being otherwise would be socially disadvantageous. This may be a banal conclusion, perhaps even a truism, but despite this more elaborate hypotheses have been proposed instead (e.g. Knight, 1998; Power, 2000; van Rooij, 2003). This chapter will spell out how the stability of human communication can be explained in the ordinary terms of animal signaling theory, and in doing so shows that post-hoc hypotheses are probably unnecessary and in some suggested cases are actually incoherent.

## 4.1 Biological and linguistic function

### 4.1.1 Biological function

A number of scholars have suggested various purported adaptive scenarios for language (see Számadó & Szathmáry, 2006 for a review). Amongst those that have received the most attention are those of social grooming (Dunbar, 1993, 1997), and sex (Burling, 2005; Miller, 2000). I will discuss these ideas in a little detail, but the key point that should be made is a general one – that any particular application of language is too specific to be the function of language. Social grooming and sexual courtship are clearly not all we use language for. Rather, language is better seen as a general-purpose device that can be used to achieve each of these goals, and many others. This is true of many human traits:

"hands may be very good for throwing rocks, shooting arrows, and pressing triggers, but we do not infer that manipulability therefore arose as an adaptation for success in aggression" (Gould, 1991, p.53). Indeed; instead we simply conclude that such manipulability is an adaptive end in its own right; a versatile tool with a wide range of applications. Similarly, we use a range of linguistic devices to fulfill a diverse range of social evolutionary functions. Language's great expressivity reflects its diverse selection pressures. Furthermore, the grooming and sexual selection hypotheses, and others like them, are limited in their capacity to explain precisely what is interesting about language. They do not, for example, give us any real insight into why language, and in particular linguistic structure, need be as complex as it is (Pinker, 2003). After all, there are myriad communication systems in the natural world that perform these and similar tasks, but none is as complex as natural language. It is instead as a synchronic explanation that adaptive hypotheses are most likely to be useful. This point is not lost on advocates of the gossip and sexual selection hypotheses, who actually make a more precise claim than their critics (e.g. Bickerton, 1996) appreciate; namely that the purported function is simply phylogenetically primary, in some sense: "I suggest that language evolved to facilitate social interaction, and the ability to refer to objects in the environment [and other specifically linguistic features]… arose as an evolutionary by-product of this capacity" (Dunbar, 1993, p.727). Consequently, when we evaluate claims about the evolutionary function of language we "need to be... careful about unpacking the layers in the selection cake" (ibid.). The data will determine which hypothesis is the most likely, and consequently I support calls (e.g. Dessalles, 2007) for more ethology in language evolution.

As discussed, social grooming and sexual selection are amongst the most prominent adaptive hypotheses. The grooming idea starts from the observation that non-human primates groom each other not only for hygienic reasons but also as a way to bond socially. Why do humans not do this? We live in large groups, far larger than that of other primates. The question of our ecologically natural group size is of course complicated by the fact that the layer of familiarity in human relationships are multiple and multi-faceted. Nevertheless a good guide can be found if we extrapolate from the group sizes of other primates, which correlate according to neo-cortical size (Dunbar, 1992). Doing so produces a figure of 150 (ibid.). If, furthermore, we extrapolate from the time it takes a primate to groom another individual it becomes clear that humans would not be able to maintain grooming as a tool of social bonding in such large groups (Dunbar, 1993). The suggestion, then, is that language, and specifically gossip, allows us to 'groom' multiple individuals at a time, and thus to overcome the problems of time management associated with a large group size.

Language certainly has a significant role to play in human sociality; as one comprehensive study of modern language use concludes, "play with language form and function leads to a range of social and cultural purposes" (Carter, 2004, p.112). Moreover, in the final chapter of this thesis I will suggest that the social brain hypothesis (Byrne & Whiten, 1989; Dunbar, 1998a; Dunbar & Shultz, 2007; Humphrey, 1976) has a major role to play in the phylogenetic story of language. Furthermore, a three-way correlation between grooming, group size and the number of calls used by a species appears to hold over a very wide range of primate species (McComb & Semple, 2005), and it does seem that gossip takes up a large proportion of our conversational time (Dunbar et al., 1997; Haviland, 1977).

However, the gossip-as-grooming hypothesis can be argued to take the idea too far. First, the mechanisms of grooming and gossip appear to be different: while primate grooming releases a large number of the chemical compounds known to help facilitate social bonding (Panksepp, 1998), gossip, and indeed language use in general, does not do the same. The response to this objection (Dunbar, 2004) is to point to laughter, which does release the endogenous opioids in question. So now the argument is that gossip induces laughter which, in turn, helps to bond groups, but this only produces two new questions: how much of gossip actually induces laughter, and how much laughter is produced in gossip-free contexts? The parallelism between language and grooming is now another step removed, and so, therefore, are claims about their equivalent functionality. Second, we do not seem to gossip for gossip's sake. If gossip were simply social currency, then we would expect the topics discussed during gossip to be essentially random. However, when we look more closely at the content of gossip we find clear, evolutionarily unsurprising preferences. For example, we are more interested in gossip about those of the same age and sex, and are more likely to pass on gossip that casts allies in a favourable light and rivals in an unfavourable light (Buss & Dedden, 1990; Greengross & Miller, 2008; McAndrew et al., 2007; McAndrew & Milenkovic, 2002). We are also better able to evaluate the veracity of gossip where that gossip is of social importance to us (de Backer et al., 2007; Hess & Hagen, 2006), and evidence that the sources of gossip are independent increases the likelihood that it will be believed (Hess & Hagen, 2006; Sommerfeld et al., 2008; Sommerfeld et al., 2007). In short, we are interested in gossip not simply because it fulfils one's social responsibilities, but because the content of gossip is of paramount importance for survival and prosperity in a social environment. Finally, there is the issue of evolutionary stability: why should the first primates to be groomed linguistically accept the cheap one-to-many token of gossip when the more personal one-to-one token of grooming has been previously available (Power, 1998)? Under the pressures of time-management that are crucial to the gossip hypothesis (Dunbar et al., 1995), individual grooming would have represented an even greater sacrifice than previously, and would thus have been even more valued by recipients. The question of how one stable economy (the one-to-one currency of grooming) morphed into another (the one-to-many currency of gossip) is not presently addressed. In sum, sociality, and in particular increased group size, will surely have been critical for almost all aspects of human cognitive evolution, but its influence may not be quite as specific as the grooming hypothesis supposes.

Much the same is true of the sexual selection hypothesis. Whilst a number of anthropological observations (see Locke, 2001; Miller, 2000) and evolutionary psychology experiments (see §4.1.2 below) make it clear that humans use various aspects of language in sexual pursuit, competition and choice, the only conclusion this warrants is that sex is just one context in which language can be usefully applied. Indeed, the stronger hypothesis, that the capacity for language itself evolved under the pressures of sexual selection, faces at least two significant problems (Fitch, 2004). First, sexually selected traits are usually sexually dimorphic, with the displaying sex (usually the male) expressing such traits exclusively, or at least to a far greater degree, than the choosing sex. Second, sexually selected traits are not usually expressed until puberty. There are arguments that can be advanced in response to these criticisms. With regard to the first, the degree to which we should expect to see sexually dimorphic display is in large part determined by the relative levels of parental investment in child rearing: the more investment a parent has to make, the less likely they are to display sexually selected traits (Andersson, 1994), and in humans both parents contribute. On the second point, it could be argued that full linguistic competence, in the widest sense, does not develop until puberty: although we normally think of fluency in terms of syntax, semantics, phonetics and other language-internal phenomena, humans are not pragmatically adept until post-adolescence – which is precisely what the sexual selection hypothesis would predict. These responses may or may not float, but even the most vociferous advocate of the sexual selection hypothesis (Miller, 2000) acknowledges that the story cannot be quite so straight-forward. In general, all claims regarding 'the' function of language will be problematic precisely because, as discussed above, to argue exclusively in favour of one or the other is to mistake present use for ancestral function: "it is as if the evolution of organs of locomotion such as wings or legs were discussed only in terms of the effects of locomotion such as fleeing predators, finding food, or finding mates, without considering the proximate function of these organs, namely locomotion itself" (Origgi & Sperber, 2000, p.141). It may be the case that specific aspects of language have been modified by sexual selection, and in that case we may find it useful to separate out the numerous functions for which language is used, but just as the only general statement we can make about the evolution of legs is that they are organs for locomotion, the only general statement we can make about language is that it is a tool for communication – the designed alteration of others' behaviour.

### 4.1.2 Linguistic function

What is the relationship between this conception of the biological function of language and linguistic functionalism? Within linguistics, functionalists are those for whom "[linguistic] form is so beholden to meaning, discourse, and processing that it is wrong-headed to specify the distribution of the formal elements of language by means of an independent set of rules or principles" (Newmeyer, 1998, p.10). This sets functionalism apart from the other basic approach to linguistic form, formalism, for which a central focus of linguistic enquiry is an exposition of "the formal relationships among grammatical elements independently of any characterization of the semantic and pragmatic properties of those elements" (ibid., p.7, italics added). The (often acrimonious) debates over which of these two basic orientations is the more appropriate have a long history, but the crucial point is as follows: the formalist is committed to a description of linguistic form that is independent of use, discourse, pragmatics and other 'external' phenomena, while the functionalist insists that linguistic form cannot be coherently studied without reference to such matters. Formalists are thus quite naturally drawn towards a code model of language, in which, as was noted in §2.1.2, syntax is taken to be a formal, computational system of symbol manipulation, whereby phonological input is translated into propositional form. In contrast, the theoretical commitments of functionalism mean that, whilst there are many variants of the basic functionalist approach (Croft, 1995; Newmeyer, 1998), there is a greater emphasis placed on the importance of use in the determination of linguistic form. Consequently functionalists typically reject the autonomy of various aspects of language, and although they differ in specification of exactly which aspects cannot coherently be considered autonomous (Croft, 1995), syntax is typically among them (as we have seen, the proper Chomskyan will insist on syntactic autonomy). This has on occasion (e.g. Givón, 1979; Halliday, 1985) led to the rejection of the term syntax because, following its use in the philosophy of language, it is often placed in contrast to semantics, a dichotomy that the functionalist is unwilling to accept. Grammar, a more inclusive term that emphasises the interplay between the lexicon and the act of symbol manipulation, is often used instead.

Some examples of functionalist explanation would perhaps be useful here. In nearly all languages there are three basic sentence types: imperative, interrogative and declarative, and that these are used paradigmatically for ordering, questioning and asserting, respectively. It could thus be suggested (see Levinson, 1983) that this is because humans are in general more concerned with these three functions than others. No further explanation is considered necessary. Second, in the English word *computations* the derivational affix *–ation* comes before the inflectional affix *–s*. This, it has been proposed (Bybee, 1985), iconically reflects the conceptual closeness of the affixes: *-ation* has more influence on the linguistic meaning of the word than *–s*, and so is closer to the stem (*compute*). Such accounts are, of course, "suspiciously *post-hoc*" (Levinson, 1983, p.40); and to overcome that the theoretical constraints that help us determine 'the' function of a particular instance of linguistic form must be specified before we elaborate on what that function might be. In many respects this is the defining goal of linguistic functionalism (Newmeyer, 1998). (A more satisfying account of word form is diachronic: the stem reflects the oldest usage of the word, with each affix added only to the beginning or end of what was available at the time; *nation* is older than *national*; which is in turn older than *nationality*; and so on to recent words like *denationalisation* (Hurford, in press).)

It should be no surprise, given the mission statement of this thesis, that I think neo-Darwinism can provide such a foundation (see also Tooby & Cosmides, 1990b). After all, adaptationist approaches are in general functionalist, whether they are concerned with language or any other form of behaviour: they explain physical or behavioural form by reference to the environment and thus to the function that the trait is required to perform (see Newmeyer, 1991 for an attempt to reconcile formalism with Darwinism). It is for this reason that one of the replies to Pinker and Bloom's article was entitled *Welcome to Functionalism* (Bates & MacWhinney, 1990, p.727). One idea of how evolution might inform linguistic functionalism comes from a series of computational simulations that examine how the iterated cycle of use and acquisition might address the basic functionalist problem of how to explain the match between form and function (Kirby, 1999). Another comes from the many evolutionary psychology studies of various aspects of language that have been published in recent years. This body of work obviously does not use the term functionalist itself, but its defining concern with how function can help determine form makes it inherently functionalist. For example, the finding that men will adjust their pitch inversely to their perception of relative social status (Puts et al., 2006) is precisely an instance of how one aspect of linguistic form (pitch, in this case) is affected by a function to which it is applied (the signalling of self-perceived social status). Formalists are typically content to grant such explanations, since, by their own definition, they consider phenomena like pitch to be marginal to the linguistic system. Semantic creativity (Franks & Rigby, 2005; Griskevicius et al., 2006) and vocabulary size (Rosenburg & Tunney, 2008) have also been shown to vary along adaptive lines (in these cases by sexual context), and perhaps the formalist would grant these too. The extension of this type of work into the domain of syntax would thus be a worthy target for future evolutionary psychological research.

It should be stressed that the results mentioned above should not be taken as support for an account entirely based upon sexual selection, but rather as an illustration of the fact that language form can vary according to function and also, moreover, that the nature of such variance can be predicted on basic evolutionary grounds. Evolutionary psychologists have thus far ploughed the sexual selection furrow more than any other, perhaps because it is of more immediate interest, perhaps because it is more tractable, but there is no reason why, at least in principle, the same approach could not be widened to include, say, how form changes according to the relative importance (measured in adaptive terms) of the topic of immediate conversation. Note also that work on infant- and child-directed speech (e.g. Burnham et al., 2002; Kitamura et al., 2002) and on alignment in conversation (Pickering & Garrod, 2004) could also be interpreted in a similar way, i.e. as the partial determination of form by function.

It would seem, then, that the functionalist agenda could be cast in the ordinary language of Darwinism (and, indeed, vice versa), and of social evolution theory in particular. That is an exercise beyond the scope of the present thesis, which attempts to explain how pragmatics can be understood in Darwinian terms, but it does raise the question of the relationship between pragmatics and functionalism. The two have clear commonalities, in particular an emphasis on usage. Their interaction can be seen in, for example, grammaticalisation (Hopper & Traugott, 1993), in which lexical morphemes take on functional roles, often precisely because of their pragmatic usage. However there appears to be surprisingly little literature that explicitly addresses the nature of the interplay between the two. Perhaps this is partially because one, pragmatics, is a topic of enquiry, while the other, functionalism, is a methodological approach, and hence the two are not directly comparable. Moreover, the topic with which functionalism is most concerned is not pragmatics but grammar. This fact hints at the broader scope of pragmatics and the different intellectual context in which it exists: functionalism is often set up as an antidote to formalist approaches, and is consequently focused on the question of exactly how function can determine form; while pragmatics is concerned not only with the influence of function, but also with the broader question of how context helps determine function in the first place and an elucidation of the cognitive abilities that allow this to happen (Levinson, 1983).

On the surface we might expect functionalism to share a closer relationship to social evolution theory than pragmatics. After all, functionalism's primary concern with how function determines form obviously mirrors the adaptationist agenda that underpins social evolution theory. However, perhaps because of its wider breadth, it is in fact pragmatics that addresses the same sorts of concerns that social evolution does – hence the focus of this thesis. An obvious example is the problem to which we will now turn, and which will occupy the remainder of the chapter: that of honesty, which does not fall within the domain of functionalism as usually understood, but which is central to pragmatics. Indeed, it is the principle concern of the Gricean maxim of Quality. However, in keeping with the tenor of the thesis, I will approach the problem not from pragmatics, but rather from social evolution theory; and at the end of the next chapter (in §5.4.2) we will see how the two approaches converge. However pragmatics is far from excluded here – in fact, more than anywhere else in the thesis this chapter will see pragmatics shed light on issues of social evolutionary concern.

## 4.2 The problems of evolutionary stability

The problem (note the singular – the inconsistency with the title of this section will be explained shortly) of the evolutionary stability of communication is typically cast in the following terms. If one can gain through the use of an unreliable signal then we should expect natural selection to favour such behaviour. Consequently, signals will cease to be of value, since receivers have no guarantee of their reliability. This will, in turn, produce listeners who do not attend to signals, and the system will thus collapse. This is the defining problem of animal signalling theory (Maynard Smith & Harper, 2003), yet with regard to language it has received only limited attention (exceptions are Hurford, 2007; Knight, 1998; Lachmann et al., 2001). Moreover, where it has been raised the tools of animal signalling theory have too often been misapplied (the most clear example of this is Knight, 1998, which I discuss below). This chapter attempts to put that right. So: what general processes keep communication systems stable, and which might apply to human communication?

That makes the problem sound like a conceptually straight-forward one and in many ways, at least with respect to animal communication, it is (the answers are not so easy to follow though, since they are often mathematically quite involved e.g. Grafen, 1990) However for human communication the matter is more complex, since there are two (analogous) problems rather than one. At one level communication is an inherently cooperative act: there must be some agreement about what signals refer to what phenomena in the world – there must be a shared agreement on the mappings between 'meaning' and form. (I put 'meaning' in scare quotes only because it is not clear what it might mean for an animal to have meanings in any recognisable sense of the term, a point that is expanded on below.) At another level that signal must be something that the audience can place their trust in, so that they are not misinformed in any way. And then, of the course, at a third level the goals to which communication is applied may be more or less cooperative: two individuals with a shared goal will use communication for cooperative ends, but two individuals with mutually incompatible goals will use it antagonistically. Importantly, however, for it to be even used antagonistically it must already be cooperative in the first two senses. An explanatory analogy between the first and third levels is with a game of tennis (or indeed any competitive sport) (Hurford, 2007). To even be able to play tennis with each other we must both recognise the rules of the game and play within them; refusal to do so means that we cannot even play a meaningful game at all. In the context of communication we can call this communicative cooperation: interlocutors must agree upon the meaning of a signal. However, once we have agreed to play by the rules of tennis we will, if we are intent on winning the game, play as uncooperatively as possible, pushing the ball to the corners of the opponent's court and generally trying to force errors in their play. This is material cooperation (or rather: non-cooperation), and within communication it is entirely optional. This distinction between the first and third levels has been previously outlined (ibid.). My suggestion is that we also recognise another type of cooperation involved in communication, nestled between these two: the honest use of signals. Interestingly there is, for the pragmatician, an obvious term for this type of cooperation: informative cooperation. The reason it is obvious is that it recognises the distinction, briefly discussed in chapter 1 (see figure 1.2) and well recognised within pragmatics, between an individual's informative intent and their communicative intent. To recap: the former refers to the speaker's intention to inform the listener of something, and the latter to their intention that the listener recognise that they have an informative intention. Pragmatics thus recognises that when a speaker produces an utterance they do not just intend that the listener understand whatever it is they are talking about, but also that they intend that the listener understand that the utterance is an act of communication designed to achieve an informative intention. We can thus distinguish between the communicative layer, which is about the fact that there is a coherent communicative act, and which requires a reliable mapping between meaning and form; and the informative layer, which is about the fact that the content of the utterance is a reliable guide to the world, and which requires honesty on the part of the speaker.

Correspondingly, we have two types of cooperation necessary for communication: communicative cooperation and informative cooperation. We also have a third, entirely optional type: material cooperation. For example, when I lie to my supervisor about the work I have done I am reliable but dishonest; and when I shout at him because he has not read the work I have done I am reliable and honest but materially uncooperative. As necessary conditions, the first two layers demand social evolutionary explanation. Indeed, in many respects the evolutionary stability of cooperative enterprises is the defining problem of social evolutionary theory (Axelrod, 1995; Axelrod & Hamilton, 1981; Dawkins, 1976; Maynard Smith, 1982; Wilson, 1975). There are two problems to be addressed, then: one regarding how signaller and receiver can agree upon a shared 'meaning' for a given signal (communicative cooperation); and another about whether the signaller uses that meaning in an honest way (informative cooperation). To distinguish between the two problems, and to disambiguate between two terms that have previously been used synonymously, I suggest that the former problem be termed the problem of signal reliability, and the latter the problem of signal honesty. The difference is depicted in figure 4.1. These two problems are formally equivalent; that is, they have the same logical structure. As a result the possible solutions are identical too. Of course, this does not mean that the two problems need actually have the same solution – it is perfectly possible that the problem of reliability will be solved differently to the problem of honesty in any particular case.

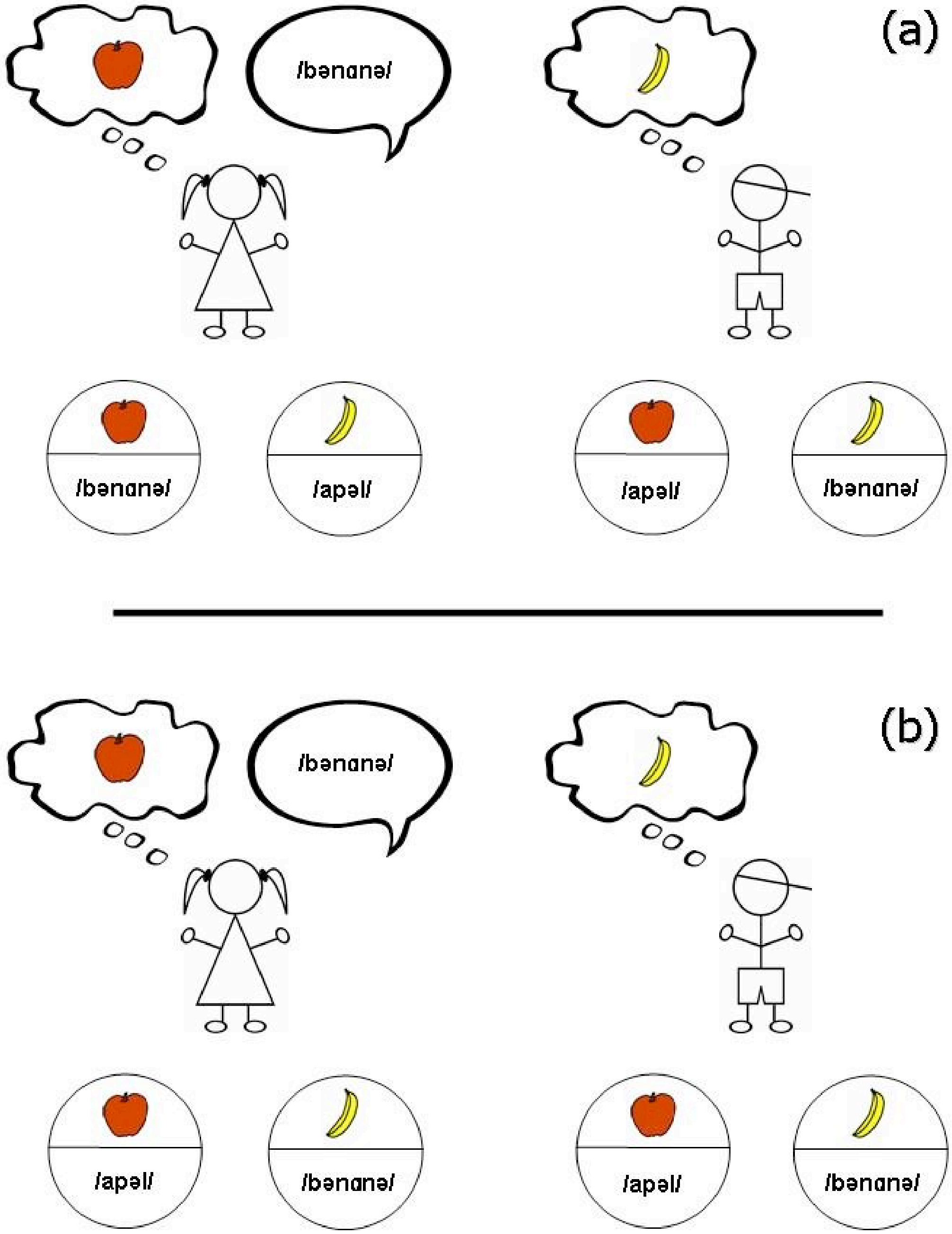


Figure 4.1: The twin problems of (a) reliability; and (b) honesty. In both cases the girl has said /bənɑnə/ having thought of an apple, and this fails to correspond to the boy's mapping of the sound (which is as per the convention in English). However the reasons for this failure are different in each case. In (a) the girl has a different (in fact, the precise opposite) mapping from sounds to meaning than the boy, and this makes her *unreliable*. In (b) she has the same mappings as the boy but has chosen to communicate a different meaning than the one she has thought of, and this makes her *dishonest*.

Before we ask about the possible solutions to these problems, I want to comment briefly on why these distinctions have not previously been recognised by animal signalling theorists. This will also allow us to clarify what is and (more importantly) what is not implied when we talk about the 'meanings' of animal signals. The key point is that (linguistic) meaning is an invention of human cognition – it is a label we attach to reliable associations between stimuli and their consequences. There is therefore no way to separate an animal signal from its 'meaning', since there is no such thing – there are only anthropomorphic glosses, which we use simply for convenience, with no necessary implication that animals have 'meanings' in any sense that we would recognise them. There is, it is true, now ample evidence that some non-human primates, at the very least, possess sufficiently rich internal representations that they could be said to have meanings of some sort (Hurford, 2007), and in that case, then, the twin problems of honesty and reliability might arise. (Having said that, animal signalling theory is in general reluctant to engage in cognitive speculation, and is happy to take behaviour as its principle explanandum.) However more generally the notion of animal signals having meaning in the same way that our utterances do seems at best unlikely, and is probably incoherent. Yet the problem of reliability is precisely one about a disjoint between, on the one hand, the linguistic meaning of the signal and, on the other, the convention that maps that signal onto a particular meaning. So if there is no linguistic meaning to the signal, as is the case with animal signals, then there can be no disjoint. Thus the only problem for most animal signals is one of honesty – but even here there is a terminological problem: the very notion of honesty seems to presuppose that a signal is a proposition with some truth-value that can be assessed, but for a signal to have a truth-value it must have a linguistic meaning. However, I will not suggest an alternative term, for at least two reasons. First, a suitable alternative that nicely captures the basic idea is not forthcoming (except perhaps reliability, but to suggest that would be to willfully invite misunderstanding); and second, the term honesty is in such widespread use in the animal signalling literature, with little if any apparent confusion, that redefinition seems both unwise and unlikely to succeed. On the contrary, the use of anthropomorphic gloss is a common strategy in behavioural ecology and social evolution (Grafen, 1999).

This section has introduced and discussed a number of other terms, and so it seems appropriate to summarise them and their relationships to each other. Figure 5.2 does this. The nature of the problems of evolutionarily stable communication should now be clear, and we can thus ask about possible solutions.

| layer of communication | gloss | type of cooperation necessary | corresponding evolutionary problem |
|---|---|---|---|
| communicative | Do interlocutors have the same meaning-form mappings as each other? | communicative | reliability |
| informative | Is the signal an accurate representation of those mappings? | informative | honesty |
| material | Is the communication used to achieve mutually beneficial goals? | none | none |

Figure 5.2: The different types of cooperation involved in communication.

## 4.3 Solutions to the problems of evolutionary stability

As was mentioned above, in game-theoretic terms the problems of honesty and of reliability are logically identical: in both cases speakers may be tempted to defect (i.e. be unreliable/dishonest) but if all speakers do then the system will collapse. Their possible solutions must therefore also be the same. The most famous is the handicap principle. Although introduced to animal signalling theory in the 1970s (Zahavi, 1975), and almost simultaneously to economics (Spence, 1973), the idea goes back much further, at least to 19th-century sociological discussions of the conspicuous consumption of the leisure class (Veblen, 1899). Indeed, such expenditure serves as a nice illustration of the basic idea: the purchase of expensive, conspicuous objects (Ferraris, Tiffany jewellery, etc.) advertises to onlookers that the purchaser can afford to make such purchases, and therefore must be well off; the cost of the objects is a handicap that only the most affluent can afford. As a further example, consider the following passage from a recent newspaper article about stock market trading in the City of London: "Certain clients even expected such behaviour [drinking and drug taking, often to excess, during working hours] from their brokers, viewing their antics as proof that they were so good at their job, they were given free rein to behave as they pleased" ('Tricks of the traders', 2008). By the same logic, large tails make peacocks less dexterous and slower than they would otherwise be. Only the highest quality peacocks can afford such a handicap, and hence the peacock tail acts as a reliable indicator of quality. Consequently the peacock tail has become the exemplar par excellence of the handicap principle. Recent research suggests that this story may not be complete: larger tails do not seem to translate into more sexual opportunities (Takahashi et al., 2008; but see Loyau et al., 2008 for a rejoinder). The basic idea is nevertheless clear. The handicap principle was originally met within evolutionary biology with some skepticism, with a number of models and arguments produced that purported to show that it was unlikely to work (e.g. Maynard Smith, 1976), but that changed once a formal proof of its stability (Grafen, 1990; see also Maynard Smith, 1991) was published. There was no such similar skepticism within economics; on the contrary, its proponent (Spence, 1973) was awarded the Nobel prize in part for his articulation of the idea.

Although initially paradoxical, the logic of the handicap principle is an ingenious solution to the problem of evolutionary stability. It is perhaps for this reason that it is sometimes assumed to be the only process by which we might stabilise communication (e.g. Knight, 1998), and hence that if communication is to remain stable then signals must incur costs over-and-above those necessary to actually produce the signal in the first place. This distinction between the costs that are necessary to produce the signal and any additional costs that are paid as a handicap is captured by the terms *efficacy costs* and *strategic costs* respectively (Guilford & Dawkins, 1991; Maynard Smith & Harper, 1995), and the handicap principle is in essence a statement that communication can be stabilised by the payment of strategic costs. However, although they are sufficient, it is not the case that strategic costs are necessary for stability. On the contrary, such a claim is false both theoretically and empirically: several alternative processes have been identified by animal signalling theorists (see Maynard Smith & Harper, 2003 for a review), and there are many instances of signalling in nature in which no strategic costs are paid: the status badges of male passerines (Rohwer, 1975; Whitfield, 1987); the display of an already fertilised ovipositor by female fruit flies (Maynard Smith, 1956); the hunting calls of killer whales (Guinet, 1992); and, of course, human language are just a few of the many examples (for more, and more details on the listed examples, see Maynard Smith & Harper, 2003). One further point should be emphasised, as it will be of critical importance later: it is the signal itself that must incur the strategic costs. The handicap principle is unstable if the costs are transferred onto some other associated behaviour. That is, there must be a causal relationship between signal form and the cost incurred. The need for this association may obvious, but as we will see, much turns on it. Moreover, it is not universally appreciated: previous suggestions about how the handicap principle might be relevant to the evolution of language, discussed in the next section, are fatally undermined by their dissociation of costs and signal form.

What are the alternatives to the handicap principle? Indices are causal associations between signal 'meaning' and signal form. This link precludes even the possibility of unreliability or dishonesty. An example is the roar of red deer, where formant dispersion is reliably (negatively) correlated with the deer's size (Reby & McComb, 2003) as an inevitable consequence of the acoustics of the deer's vocal apparatus. The deer's larynx descends upon vocalisation, and the comparative evidence suggests that this is the result of a selection pressure to exaggerate one's size (Fitch & Reby, 2001). However that process seems to have gone as far as it can without compromising other aspects of the deer's anatomy (ibid.). As a result it is actually impossible for the deer's roar not to carry reliable information about its size and hence its social dominance; deer can lower their larynx no further, and hence the formant dispersion of their vocalisations is unfakeable. Other examples (again from Maynard Smith & Harper, 2003) of indices include male jumping spiders, who expose the ventral surface of their abdomen as an indicator of their current condition (Taylor et al., 2000) and snapping shrimps, who advertise their claws to each other as a way to avoid physical conflict (Versluis et al., 2000). To state the idea of an index more formally, signals can be free of strategic costs and evolutionarily stable so long as the efficacy cost of the signal is positively correlated with the trait in question (e.g. a function of size, in the red deer example) (Hurd, 1995; Lachmann et al., 2001; Számadó & Szathmáry, 2006).

What explanations are available when a signal is free of strategic costs and is not indexical of meaning? Several possibilities have been identified, but there is an open question about how best to categorise them. One classification (Maynard Smith & Harper, 2003) suggests a three-way division between coordination games, repeated interactions, and punishment. Coordination games are those in which some common interest overrides any conflicting motivations the participants might have (Silk et al., 2000). The classic example is the 'War of the Sexes': the husband wants to go to the pub for the evening and the wife wants to go to the theatre, but they share an overriding common interest that whatever they do they want to do it together. A real-world example is courtship in fruit flies, who mate only once in their lifetime. If a male attempts to court a female after this mating she will display her ovipositor to him and thus advertise that his efforts are futile. He then ceases courtship immediately (Maynard Smith, 1956), and thus the female's signal saves both parties what would otherwise be wasted time. Formally such games can be settled only if there is an asymmetry in the relationship such that one player or the other backs down, and if this asymmetry is known to both players (Maynard Smith, 1982). In a repeated interaction the longer-term payoffs of honesty may outweigh the short-term payoff of dishonesty (Silk et al., 2000), and hence the problem should not arise. We have already noted the important equivalences in the twin problems of communication and cooperation, and repeated interactions are an answer that is applicable to both. Indeed, it is the basic logic behind reciprocity, in which individuals trade what would otherwise be altruistic and hence evolutionarily unstable behaviours (Trivers, 1971). Finally there is punishment, in which one individual, at a cost to themselves, actively punishes another for unreliable/dishonest signalling (Clutton-Brock & Parker, 1995). This will of course act as an incentive against such behaviour, but this only really moves the problem on to a different locus, since we must now ask why punishing behaviour will evolve if it is itself costly. Indeed, it seems to be a prime candidate for of the tragedy of the commons (Hardin, 1968), since all individuals get an equal share of the payoff (of stable communication) but can let others pay the costs (of punishing others) that are necessary to obtain it.

On the surface this three-way distinction between coordination games, repeated interactions and punishment seems a reasonable one, but in fact it is not at all clear how we should distinguish between the three. If, for example, an individual abuses a system of mutually-beneficial repeated interactions and their partner then refuses to re-enter into the relationship, should we classify that refusal as the collapse of an arrangement of repeated interactions or as punishment? I see no way to consistently distinguish between the three processes described above, and I therefore suggest that they all be grouped together in a higher-order classification as deterrents. In general, deterrents refer to the situation where reliable/honest communication is cost-free, but where dishonesty is costly. It has been shown that not only is such an arrangement stable, but that where it occurs costly signals will be selected against (Gintis et al., 2001; Lachmann et al., 2001). That is, the costs are paid by those who deviate from the Evolutionarily Stable Strategy (ESS), whereas in the handicap principle costs are paid as part of the ESS. An example is the status badges of male passerines, the size and coloration of which correlates with the bird's resource holding potential (a composite measure of all factors that influence fighting ability (Parker, 1974)), but which carries no strategic costs (Rohwer, 1975; Whitfield, 1987). When the badges are experimentally manipulated to deviate from this correlation then the birds either get into fights that they will inevitably lose, or they lose out on territory and mating opportunities (Rohwer & Rohwer, 1978). It is these costly outcomes that act as the deterrent that keeps the system stable, yet they are not costly for other passerines to enforce.

We thus have a three-way classification of the basic processes by which communication may be stabilised. These possibilities are mutually exclusive and, in the absence of other suggestions, are taken to be exhaustive:

- indices, in which signal form is tied to signal meaning
- handicaps, in which signal cost is tied to signal form, and hence acts as a guarantee that is incurred by reliable/honest signallers
- deterrents, in which costs are incurred by unreliable/dishonest signallers

The question that must be answered is which of these most likely applies to human communication, and natural language in particular.

## 4.4 Evolutionarily stable communication in humans

Linguistic form is famously unrelated to meaning (de Saussure, 1959): there is no formal relationship between the words cow, pig and sheep, even though all refer to farmyard animals. We may therefore immediately discount indices as an explanation of why linguistic communication is evolutionarily stable. Handicaps may also be excluded for similar reasons: there appear to be no strategic costs associated with utterance production. Despite this "crippling problem" (Miller, 2000, p.348), there have been at least two attempts to apply the handicap principle to language. The first suggestion is that performance rituals act as a costly signal of one's commitment to the group, and hence that performers are trusted as in-group members (Knight, 1998; Power, 1998). However there is nothing in this model to stop an individual paying the costs to enter and remain in the in-group and then once inside behaving dishonestly or even unreliably. This is because the costs of the performance are not causally tied to the individual's subsequent utterances. The same problem undermines the suggestion that politeness phenomena act as a handicap (van Rooij, 2003), in that they reduce the speaker's social standing relative to the listener, place them in the listener's debt or otherwise incur socially relevant costs. For example, the utterance "I don't suppose there'd be any possibility of you…" can be read as an announcement that the speaker is prepared to incur some social cost in order to ensure that the desire which follows the ellipsis is satisfied. Let us accept, for the sake of argument, that this is correct. It still does not follow that politeness is a handicap, since the costs incurred are not paid as part of the signal. Any costs that do arise do so as a result of the signal's meaning rather than its form. If politeness does place us in social debt then this would be an example of a self-imposed deterrent rather than a handicap: it imposes a social obligation on us to return the favour in some way, and we do not renege on this because the threat of social exclusion deters us from doing so. In fact, the inapplicability of handicaps to language was recognised by the inventor of the handicap principle: "Language does not contain any component that ensures reliability. It is easy to lie with words" (Zahavi & Zahavi, 1997, p.223).

If handicaps and indices are not appropriate then we are drawn by deduction towards deterrents as a solution to the two evolutionary problems of reliability and honesty. There is an intuitive appeal to both ideas: unreliable communication, for example if one says /`kat`/ to refer to dogs, is deterred because it means that one will not be understood, and hence cannot achieve one's communicative goals; and dishonest communication will result in a loss of trust and the consequent social costs. In fact, deterrents are what we should logically expect to find in humans. In general, when indices are not available, and when the expected gains from dishonesty or unreliability outweigh the costs, then costly signals must be employed to ensure stability (Lachmann et al., 2001). Deterrents will be used only if verification is both possible and cheap. Applied specifically to humans, this means that costly signals will be used either in one-time interactions, or when communicating about otherwise unverifiable information (ibid.). This is an important result, yet it has received bafflingly little attention from evolutionary linguists. Despite being published in a major journal (Proceedings of the National Academy of Sciences) and being cited extensively elsewhere (more than 60 times as of November, 2008) it has, for example, received only one citation in any of the collections of work that have arisen from the Evolang conferences that have taken place since the article's publication (Cangelosi et al., 2006; Smith et al., 2008; Tallerman, 2005; Wray, 2002), and that was in the most recent proceedings, in my paper on this topic, in which I made the same point as now. A similar finding is that signalling can be stable if unreliability is costly (Gintis et al., 2001). Although that paper does not mention language, its applicability is clear, and yet it too has yet to be recognised by the language evolution community. It should also be noted that deterrents allow signals to take an arbitrary form (Lachmann et al., 2001). The fact that utterances are cheap yet arbitrary is too often taken to be paradoxical: "resistance to deception has always selected against conventional [arbitrary – TSP] signals – with the one puzzling exception of humans" (Knight, 1998, p.72, italics added). This is, as the examples discussed above show, simply not true. Instead, once we remove the requirement that costs be causally associated with signal form, as we do if we place the onus of payment on the dishonest individual, then the signal is free to take whatever form the signaller wishes. This paves the way for an explosion of symbol use.

What deters humans from dishonesty and unreliability? There is an obvious candidate: reputation. For this to work it must be possible for individuals to modify their future behaviour in the light of others' behaviour, and to keep in mind who is and is not to be trusted. These are tasks that the human brain performs with relative ease, often subconsciously (Pentland, 2008), but we should nevertheless recognise them as crucial prerequisites. Moreover, emotions like anger ensure that we do not repeatedly trust those that have cheated us (Ekman, 1992; Tooby & Cosmides, 1990a), and what can be termed epistemic vigilance (Sperber & Wilson, 2008) limits our exposure to the consequences of acting on potentially dishonest communication. Indeed, the experimental evidence suggests that we are very sensitive to untrustworthy behaviour. For example, we may be more likely to recall the identities of cheaters than cooperators (Chiappe et al., 2004; Mealey et al., 1996; Oda, 1997; but see Barclay, 2008; Barclay & Lalumière, 2006). We are well attuned to the detection of unfakeable physical cues of dishonest behaviour, for example a lack of eye contact and a large number of unfilled pauses in speech (Anolli & Ciceri, 1997; Scherer et al., 1985), and these appear to be cross-cultural (Bond Jr. et al., 1990). In fact such cues may even be seen not only when we are deceptive but also in our everyday appearance: when presented with a number of faces and asked to recall them later, experimental participants are more likely to recall the identities of individuals who later defected in a game of prisoner's dilemma, even when they do not have access to this information (Yamagishi et al., 2003).

We are also very sensitive to our own reputational status within the social group in general, and are keen to maintain our standing: cooperation can be maintained in various economic games once reputational effects are added, but not otherwise (Milinski et al., 2002; Piazza & Bering, 2008; Semmann et al., 2004, 2005; Wedekind & Milinski, 2000). This is true even if we experience only subtle cues of a potential loss of reputation, such as stylised eyespots on a computer (Haley & Fessler, 2005). Such effects have also been found in more ecologically-valid conditions: an honesty box for tea, coffee and milk in a University common room received greater contributions when the small picture above it was a pair of human eyes rather than a flower (Bateson et al., 2007). This attentiveness to one's own reputation and to cues that it may be affected by current behaviour should not be a surprise, since a loss of reputation will mean exclusion from the local group, a heavy penalty for a social species like ourselves.

Indeed, the emerging consensus from the burgeoning literature on the evolution of cooperation is that reputational effects are crucial to stability (Fehr, 2004). A similar story seems to hold in primate societies (Gouzoules & Gouzoules, 2002).

There are some further points that should also be made, which together highlight the plausibility and coherence of this picture. The first is that if the deterrent of reputation is the main tool by which stability is maintained, then the second-order problem of how the deterrents are enforced does not arise. In general, if stability depends upon the fact that defectors receive some sort of punishment, then a question arises over who will carry out that punishment. To punish is itself costly, and so the problem of stability rears up again. However, with reputation no individual is asked to bear the brunt of the costs of punishing others, because social exclusion is not itself costly to enforce. On the contrary, it is the most adaptive response to individuals with a reputation for unreliability or dishonesty. That is, if all individuals act in their own interest then that is enough to ensure that the dishonest or unreliable individual will be deterred. No further enforcement is necessary. This underlines the point that the cognitive capacity to keep in mind who is and is not trustworthy is crucial to this explanation. The second point is that the effects of reputation are likely to snowball once language in some form or another is off the ground, since individuals then become able to exchange information about the honesty and reliability of others (Enquist & Leimar, 1993). This may explain why so much of our conversational time is dedicated to gossip (Dunbar, 1997): by discovering more about others' experiences of third parties we are better able to inoculate ourselves against anti-social behaviour in the future. Finally, the third point is that whilst linguistic communication may be maintained in the way described, other forms of human communication may be kept stable by other forces. Indeed, in our discussion of the handicap principle, passing reference was made to the human practice of conspicuous consumption. There are good reasons to think that this is more than an analogue: the advertisement of one's wealth by such means may indeed be a sexually selected honest signal (Miller, 2000). If so, it would accord with the general prediction that costly signals will be used when verifiability is not easily available

(Lachmann et al., 2001): other than displays of wealth, the female has no real way to know whether or not the male has the requisite resources to raise offspring.

The conclusion that we are honest and reliable simply because if we were not then people would not believe or listen to us, respectively, seems so banal as to hardly be worth stating. Yet it has not yet been properly recognised by those interested in the origins and evolution of language, who have, where they have addressed the question at all, tended to generate additional post-hoc hypotheses in which language is treated as a special case (e.g. Knight, 1998; Power, 2000), rather than seek to explain the stability of human communication with the existing tools of animal signalling theory. What we have seen in this chapter is that no such claims are necessary to explain the stability of linguistic communication. On the contrary, the traditional tools of animal signalling theory offer a candidate explanation of stability, one that fits tightly with both our intuitive ideas of how social contracts work and many experimental data. Direct empirical testing of this idea would be welcome, but even until then, the burden of proof lies with those who wish to claim, contrary to the analysis presented here, that the stability of human communication is not maintained to a large degree by the mechanism of reputation.

# Five

# Mechanism and Relevance Theory

The position we have reached is that language is, at the ultimate level of explantion, a multi-purpose tool, one that can be used to achieve a wide range of conversational goals. We must turn now to the question of proximate mechanisms. How, exactly, are these objectives achieved? That is, in many respects, the defining question of most linguistics – which, broadly speaking, seeks to describe how language works. Correspondingly, a complete answer to our question will necessarily involve a detailed description of how all the various aspects of language (syntax, semantics, phonology, phonetics) operate and how they relate to each other, not to mention their neurological foundations. However, if we are to speak about how this machinery combines to do things to others then we must talk about how the language faculty interacts with the outside world – which is the domain of pragmatics. In fact, this question should in some respects be primary. As was briefly discussed in chapter 3, pragmatic concerns cannot be treated as an adjunct that is to be wrapped around language itself, for this denies the fundamentally inferential nature of comprehension; and, indeed, the ostensive nature of production (Carston, 2002). Of course, we can make some general statements about the internal workings of language, but any attempt to isolate syntax, or indeed any one of the traditional subdomains of linguistics, from its interaction with the environment is to have a wholly inside-out view of the relationship between language and the world (ibid.).

This chapter is thus principally concerned with the question of what our social evolutionary approach can tell us about that relationship. In the last chapter we mentioned, briefly, how one major strand of pragmatic thought, speech act theory, has approached this question – as an exercise in taxonomy. However, while useful, this approach faces an arguably insurmountable problem: even if we account for the more than one thousand English verbs that define different speech acts (as claimed in Austin, 1955), there will still be a vast number of other expressions that will not have been classified. Moreover, it is not at all clear how we might define categories for the speech acts in any systematic and unbiased way. We are thus compelled to grant to each act its own category, but this then means that "the theory of speech acts becomes no more than a paraphrase without predictive power" (Dessalles, 2007, p.274). This chapter will thus pursue a different strategy: the articulation of exactly what criteria our theories of pragmatics should satisfy. Those criteria will arise from a general model of communication applicable across the biological world, which will itself be derived from the definition of communication that was defended in chapter 3. We will see that they correspond, with remarkable precision, to one particular approach to pragmatics – that of Relevance Theory (henceforth RT) (Sperber & Wilson, 1986).

The chapter will begin with a description of RT, so that in the second section, which describes the general theory of communication, you will be able to anticipate in advance its application to RT. In the third section these two strands will be brought together to make clear the central claim of this chapter: that RT is the Tinbergenian mechanism by which the evolutionary logic of communication is enforced in the human animal. The final section will respond to some of the common criticisms levelled at RT.

## 5.1 A beginner's guide to Relevance Theory

RT can be seen as "an attempt to work out in detail one of Grice's central claims: that an essential feature of most human communication, both verbal and non-verbal, is the expression and recognition of intentions" (Wilson & Sperber, 2002a, p.249). This section offers only a brief sketch of the key ideas; interested readers who wish to seek out more detail are encouraged to turn to the book Relevance itself (in particular the 2nd edition, Sperber & Wilson, 1995), or one of the numerous précis (e.g. Sperber & Wilson, 1987; Wilson & Sperber, 2002a). Nevertheless, whilst inevitably incomplete, this overview is intended to contain within it enough detail for the purposes of this chapter. (Throughout this thesis I have and will cite the 1st edition of Relevance, simply because as the original publication it allows the reader to grasp the chronology of ideas and the reactions to them. Where I explicitly refer to the 2nd edition (1995), it is to the postface, which contains a number of important clarifications and refinements, and as such represents the more definitive statement of the theory. There have been no major changes to it since then; recent work has instead focused on the application and testing of RT's core ideas.)

RT is a cognitive approach to communication, and as such it is, in the terminology of Tinbergen's four whys, mechanistic rather than functional. Indeed, RT is explicit that a dissatisfaction with the cognitive aspects of the Gricean approach to communication was a major motivation behind the development of RT: "the elaboration of this idea [that communication involves the recognition of intentions] in the work of Grice himself, Strawson, Searle, Schiffer and others has often taken the form of a move away from common sense, away from psychological plausibility" (1986, p.24). We have already discussed the relative merits of the code model and the ostensive-inferential model of language, and concluded in favour of the latter, which, following Grice, considers communication to involve the provision and recognition of evidence that points the audience towards the speaker meaning that is to be communicated. The basic idea of RT is, broadly speaking, that the very production of an utterance raises in listeners an expectation of relevance, and that this expectation, coupled with the evidence provided by the linguistic meaning of the utterance, is sufficient to guide them to the correct speaker meaning.

So, what is relevance? Stimuli are said to be relevant if they achieve positive cognitive effects in the recipient: a worthwhile difference in the individual's representation of the world. Differences that are not worthwhile, for example banal observations (about the colour of a passing car, say) are negative cognitive effects, and as such are not relevant. Here is a toy example (from Sperber & Wilson, 1986). Consider a context defined by the following three statements: (i) people who are getting married should consult a doctor about possible hereditary risks to their children; (ii) any two people both of whom have thalassemia should be warned against having children; and (iii) Susan has thalassemia. Now consider the following two possible utterances:

(a) Susan, who has thalassemia, is getting married to Bill

(b) Bill, who has thalassemia, is getting married to Susan

Both of these are relevant; that is, they have positive contextual effects, namely that Susan and Bill should consult a doctor about possible hereditary risks to their children. However (b) (and not (a)) contains the new information, not present in the context, that Bill has thalassemia. It thus has a positive contextual effect that is absent from (a): that Susan and Bill should be warned against having children. It is thus the more relevant of the two utterances.

Weighed against positive cognitive effects is the processing effort required to achieve them: the greater the effort required, the lower the relevance. That is, if the same positive contextual effect can be achieved with two different utterances then the one that requires the least processing effort is considered the more relevant. To illustrate (again from Sperber & Wilson, 1986), re-consider utterance (b) above, and compare it to (c), below:

(c) Bill, who has thalassemia, is getting married to Susan, and 1967 was a great year for French wines

Being entirely unrelated to the context (as defined by (i)-(iii) above), the information contained in (c) but not in (b) has no additional positive contextual effects. However, it does require additional processing effort: it introduces more conceptual material, and hence requires more inferential work. It is therefore less relevant than (b). In general, we can make the following two general statements about relevance:

Relevance

(i) All other things being equal, the greater the positive cognitive effects achieved by processing a stimulus, the greater the relevance to the individual at that time

(ii) All other things being equal, the lower the processing effort expanded by processing a stimulus, the greater the relevance to the individual at that time

Note that relevance is a property not just of utterances but of any stimulus, ostensive or otherwise, and it is specific to a given individual at a given time. That is, the relevance of a stimulus is utterly dependent upon the context. Relevance is thus not a binary quality but instead one of degree, and consequently it is measured comparatively rather than in absolute terms. The possibility of a quantitative measure of relevance is admitted, but it is argued that this is "not the kind of notion psychologists should be trying to develop" (Sperber & Wilson, 1986, p.129) because the reality is that any such measure will be psychologically implausible. Moreover, such problems are argued to be inherent and inevitable in more-or-less all cognitive enquiry. This frankness is somewhat disarming, but it is hard to deny: the reality is that we can only measure cognitive processes by proxy. Sperber and Wilson's unwillingness to provide a formal metric is for many an unsatisfactory aspect of RT: "if all it can say is 'a relevant message is one which provides a cognitive effect', and if the idea of cognitive effect remains intuitive [rather than absolute], then it has no significant contribution to make" (Dessalles, 2007, p.278). There is some substance to this criticism: it seems inherently unsatisfactory to ground the notion of relevance in the material world but to simultaneously deny that it can be measured in any absolute way. Later in this chapter (§5.4.3) I will suggest how, at least in principle (although not, it should be admitted, in practice), this problem could be overcome.

As I briefly mentioned above, one of the two core claims of RT is that humans are able to and do maximise relevance. This is formalised in the first, or cognitive, principle of relevance:

> First (cognitive) principle of relevance
>
> Human cognition tends to be geared towards the maximisation of relevance

That is, when they process a stimulus humans will converge upon the interpretation that grants the stimulus the maximum degree of relevance. Put another way, given two competing interpretations of a stimulus, the one that will be accepted is that which achieves the greatest cognitive effect for the lowest processing effort. This tendency is, according to RT, simply a fact about human cognition, and it is argued for in part on basic Darwinian grounds: "we start from the assumption that cognition is a biological function, and that cognitive mechanisms are, in general, adaptations… we assume, then, that cognitive mechanisms have evolved in small incremental steps, mostly consisting in the selection of a variant that performed better at the time than other variants that were around" (Sperber & Wilson, 1995, p.262). This is, incidentally, the only mention of evolutionary considerations in either edition of Relevance.

Recall now that RT is a theory of ostensive-inferential communication, where ostensive refers to the fact that the stimuli are designed to attract the audience's attention and provide evidence that enables them to construct the correct speaker meaning. From this definition and the cognitive principle of relevance it follows that audiences will necessarily have expectations of relevance. In other words, audiences search for interpretations of stimuli that carry positive cognitive effects – that is part of what it means to maximise relevance. Speakers will thus only produce stimuli that they believe will have such effects. Why else would they bother? If they do not have such effects then the speaker is wasting their time and energy. The very production of a stimulus thus becomes a statement that the stimulus is relevant and that it is worth the audience's while to process it. As such it constitutes a "request for attention" (Sperber & Wilson, 1986, p.155). Moreover, of the possible stimuli that the speaker can produce, they will opt for the one that is, in effect, best for them. Thus the introduction of the speaker's interests gives rise to two facts that together define the notion of optimal relevance, which is crucial to RT. As I have described, optimal relevance is defined as the criterion that an ostensive stimulus be both (a) relevant enough to make it worth the audience's while to process it; and (b) the most relevant one compatible with the speaker's abilities and preferences.

We thus arrive at the second core claim of RT:

Second (communicative) principle of relevance

Every ostensive stimulus carries a presumption of its own optimal relevance

In other words, an ostensive stimulus will only be produced if it is believed that it will be relevant, and if one is produced it will be the stimulus most relevant one that is compatible with the speaker's abilities and preferences. It is this fact that allows an audience to address the problem of radical underdeterminacy outlined in §1.3. We noted there that Grice's co-operative principle is not prescriptive of how speakers should act, but descriptive of how they actually behave, and it is because they are armed with this knowledge that listeners are able to converge upon the correct speaker meaning. The second principle of relevance is similar in this regard – it is a fact about ostensive stimuli that allows listeners to make correct inferences about meaning and as such it is not something that interlocutors can somehow 'choose' to 'follow'. It is rather something that they do as a matter of course; "communicators do not 'follow' the principle of relevance; and they could not violate it even if they wanted to" (Sperber & Wilson, 1986, p.162). It is this fact that makes us pragmatically 'fluent', a skill that, as already noted, is so automatic that it is hard for us to recognise it for what it is – which is nothing less than the glue that makes conversation possible at all. This much is of course recognised by the Gricean paradigm, which argues that the cooperative principle is simply a fact about how humans behave. The difference between that approach and RT is that in the latter the speaker's goals are used to derive the central claims. In particular, the second principle of relevance is entirely dependent upon this. Grice, in contrast, simply assumes that speakers are willing and able to provide information. For example, the maxim of quantity simply states: 'make your contribution as informative as is required'. The cooperative principle does reference 'the accepted purposes' of the conversation, but this is an inadequate response to the question 'required for what?'. To answer this we must make explicit reference is made to the speaker's own goals. Grice himself recognised this shortcoming and wished to place his idea on a more secure foundation: "I am… enough of a rationalist to want to find a basis that underlies these facts [of how people behave in communication]… I would like to think of the standard type of conversational practice not merely as something that all or most do in fact follow but as something that it is *reasonable* for us to follow, that we *should not* abandon" (Grice, 1975a, p.48, italics in original).

In many respects the goal of this chapter is to use Darwinism to provide just such a foundation. Now that we have a working knowledge of RT, we can take the next step in pursuit of that goal: to return to our basic definition of communication and use it to derive some general principles of evolved communication.

## 5.2 Towards a general theory of pertinence in communication

In the last chapter we defined communication in terms of design: both the action and the reaction must be designed, and directed at each other. Another way to think about this is to map the set of possible interactions on a graph, as per figure 5.1, with the payoff to the actor along the x-axis and the payoff to the reactor along the y-axis. Interactions that carry a positive payoff to the actor will be termed *apposite*, and those with a positive payoff to the reactor *pertinent* (*relevant* would be the best everyday term here, but obviously that cannot be used as it already has a technical meaning within RT). By definition any interaction that is both apposite and pertinent, i.e. has positive x and y values, is called communication, and when this occurs we can rename the action and reaction as signal and response respectively. Interactions that are apposite but not pertinent (i.e. have a positive x value and a zero or negative y value) are instances of coercion, and those that are pertinent but not apposite (i.e. have a zero or negative x value and a positive y value) are cues. What I want to do in this section is expand on this characterisation of communication (which is only an extension of the definition defended in chapter 3) to derive some basic principles that all (evolutionarily stable) biological communication systems must satisfy. Once that is done we will return to RT, and find that the statements that arise from this general model of communication exactly predict the two principles of relevance. I will then discuss, in the third section of this chapter, how we should interpret that fact.

I will develop the basic principles of communication both formally and, initially, informally. Consider, then, the set of points on the graph in figure 5.1. These points represent action-reaction pairs; that is, interactions between the participants. The question of which of these is most likely to evolve is a game-theoretic question: given a set of possible actions and a set of possible reactions, and a set of payoffs for each party that correspond to each possible action-reaction pair, which interaction will maximise the payoffs of the actor and reactor? (These payoffs should be read as the outcome relative to doing nothing.) Whilst the original temptation will be to assume that the answer to this question will be the interaction closest to the top-right of the graph, the matter is not quite so simple. The tragedy of the commons (Hardin, 1968), in which the group will benefit if everybody cooperates but where each member will do individually better if they defect, shows us that outcomes that maximise the payoffs for all participants are not at all inevitable (in game-theoretic terms, Pareto optima are not necessarily Nash equilibria), and we will see the same lesson play out here. In order to understand why, we need to know more about how a change in the behaviour of one or other of the players will affect the payoffs of each. Figure 5.2 considers a (grossly simplified) model in which there are two possible actions, A and B, and two possible reactions, X and Y. These behaviours collectively define four possible interactions, which are marked with a clear circle, a filled circle, a clear square and a filled square. The circles correspond to the possible outcomes that follow action A, and the squares action B; and the clear shapes correspond to the outcomes where the reactor has performed X and the filled shapes reaction Y. Figure 5.3 then marks these interactions on the graph defined by figure 5.1.

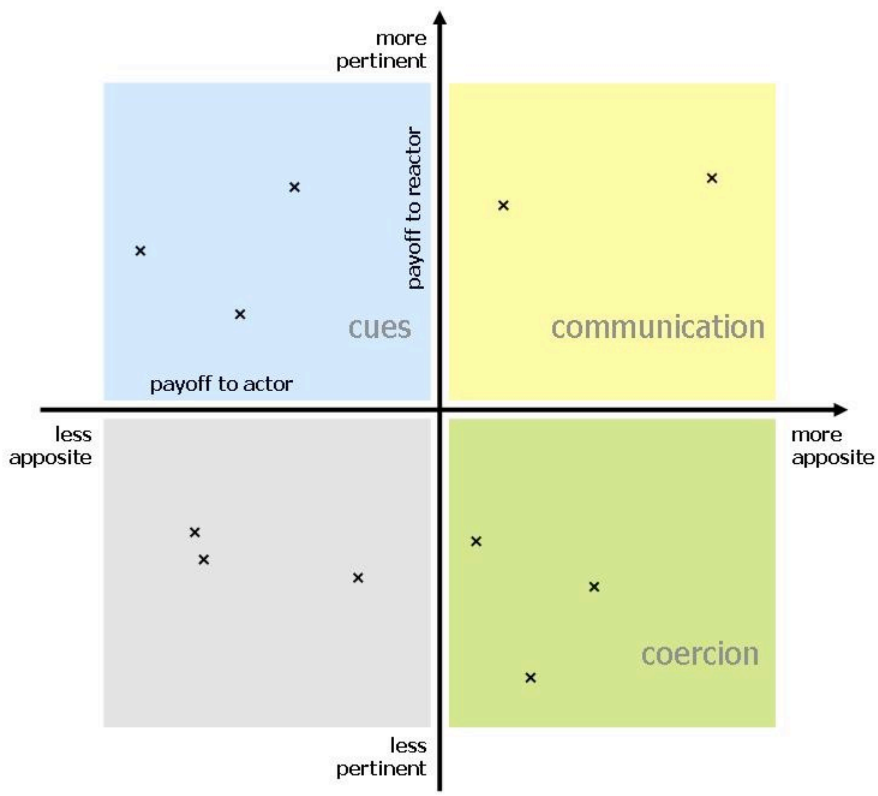


Figure 5.1: The basic model of communication. The x-axis measures the payoff to the actor, and the y-axis the payoff to the reactor. These determine the degree to which the interaction is apposite and pertinent respectively. We saw in the last chapter that communication refers to those interactions that fall in the top-right segment of the graph, cues those in the top left, and coercion those in the bottom-right. A number of possible interactions are marked on this graph with crosses.

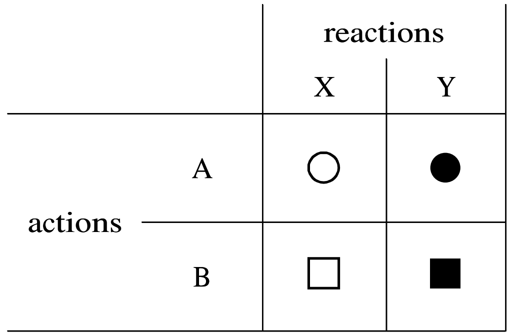


Figure 5.2: Four possible communicative interactions. The actor can perform either action A, which will result in one of the two circles, or action B, which will result in one of the two squares. The reactor can perform either reaction X, which will result in one of the clear shapes, or reaction Y, which will result in one of the filled shapes.

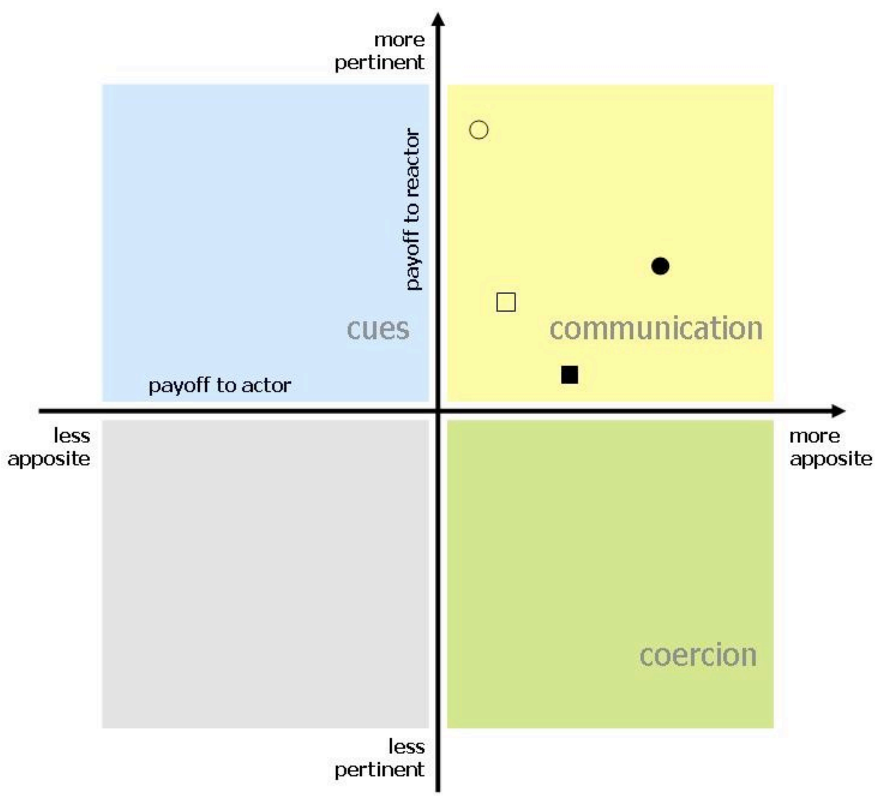


Figure 5.3: Four specific communicative interactions, generated by two possible actions and two possible reactions.

So which will occur? Which actions and reactions maximise payoffs for actors and reactors respectively? Modelling communication in this way defines it as a simple game. The first observation to make about that game is that it is different to many others in the following important respect: the moves are sequential rather than simultaneous (in the terminology, this game is dynamic rather than static). In many games (e.g. prisoner's dilemma, ultimatum game, stag hunt, etc.), the players move at the same time as each other, or at least without the knowledge of the other players' moves, which is effectively the same thing. In our game the actor moves first and the reactor, knowing what choice the actor has made, moves second. This difference is crucial: it means that the reactor is guaranteed to achieve the optimal payoff given the actor's behaviour, and also that the actor's payoff is contingent upon this. This is not the case in static games, and it is a fact that the actor must take into account if she is to maximise her payoffs. In short, she should calculate what the reaction will be for each possible action and then choose her action accordingly. There is, then, a feedback loop, in the sense that the expected behaviour of the reactor is an input into the behaviour of the actor; and once that behaviour is performed then the reactor simply chooses the reaction that is most pertinent (pertinence being defined as the payoff to the reactor). The action-reaction pair that results is known as the Stackelberg equilibrium (named after the German economist Heinrich Freiherr von Stackelberg, who studied duopolies in which one firm (the leader) would act first and the other (the follower) would then act knowing what the leader has done (von Stackelberg, 1934)).

The Stackelberg equilibrium is not, it should be noted, the Bayesian equilibrium of classic signalling games, which are, by definition, games of incomplete information: the signaller has some knowledge about the state of the world, and in light of this they must choose what signal to send to the receiver, who does not have that knowledge, and hence cannot know the exact payoff that they will receive given the signal and their own action. The game under discussion is, in contrast, one of complete information; the receiver knows what payoff they will receive given the signaller's behaviour. This is appropriate, as this is a game about the fact of communication rather than a game of communication, which is what the signalling games of economics and game-theoretic pragmatics (Benz et al., 2006) are. Returning now to figure 5.3, we can observe that the actor should take note of which instance of each shape is most apposite and then choose from this reduced set of possible actions, as depicted in figure 5.4, where the two possible outcomes for each action are grouped together.

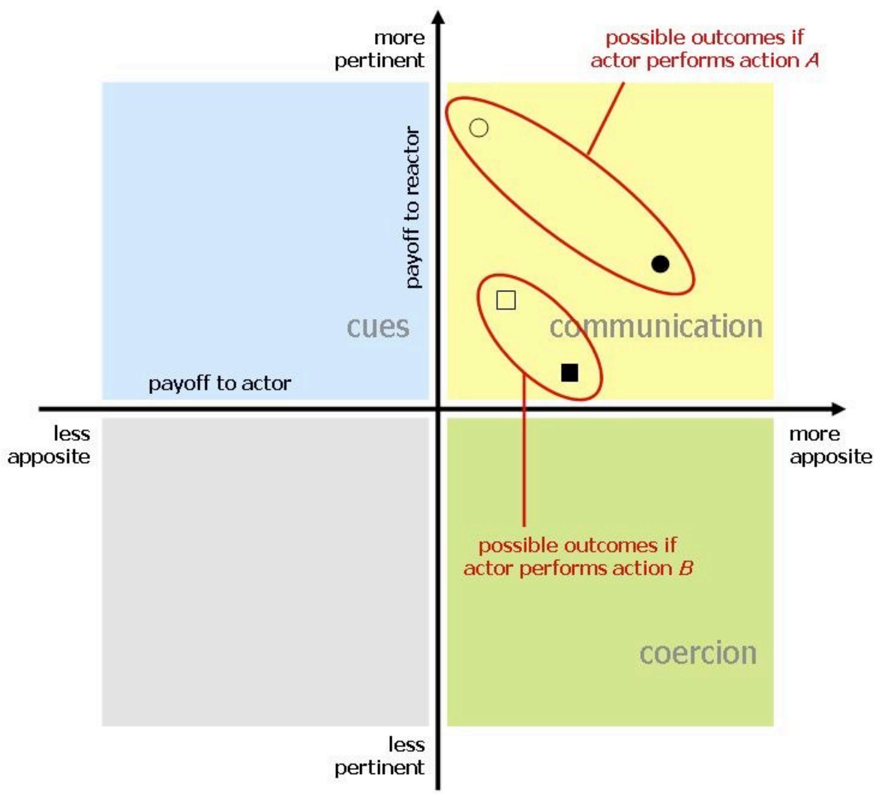


Figure 5.4: The actor's choice. Following figure 5.2, either the actor performs A, which will result in one of the interactions marked by a circle, or they perform B, which will result in one of the interactions marked by a square.

Now, it should be apparent that if the actor performs action A (and so chooses the circles) then the reactor will opt for the clear circle, as that offers a greater payoff; and if the actor performs B then the reactor will choose the clear square. The actor thus has a choice not of all four outcomes but of the two clear shapes, and of these it is the square that offers the greater payoff. This is thus the Stackelberg equilibrium that we should expect to evolve, even though one of the circles would represent a greater payoff to both participants (i.e. would be Pareto optimal); see figure 5.5.

Missing from this model is what should happen if the Stackelberg equilibrium is not in the top-right quadrant of this graph; that is, if the interaction results in a negative payoff for either participant, and is hence neither apposite nor pertinent, to at least some degree. Under such circumstances we should expect the participants who incur the negative payoff not to partake in the interaction at all. To capture this we should add to the model the possibility of doing nothing. To do this we simply include an additional pair of behaviours which we can term the null action and null reaction. These appear in figures 5.6 and 5.7, alongside four other interactions which share the same relative status to each other as the interactions in figures 5.3-5.5. The triangle represents the outcome if the actor does nothing: neither participant receives a payoff at all, either positive or negative. The hashed circle and square just to the left of the origin represent the outcome if the reactor does nothing (i.e. if they ignore the actor) – the negative payoff to the actor reflects the small but non-zero energy and opportunity costs of the act (Maynard Smith & Harper, 1995). We should, if there were no null behaviour, expect the equivalent interaction to be the Stackelberg equilibrium. However the possibility that the participants can opt out of the interaction altogether has changed the equation: it is plain to see that in this new model the Stackelberg equilibrium is the null behaviour represented by the clear triangle. As such, the system has now collapsed, since there are no interactions that are worth the participants' while: now that there are no interactions in the top-right quadrant there can be no communication.

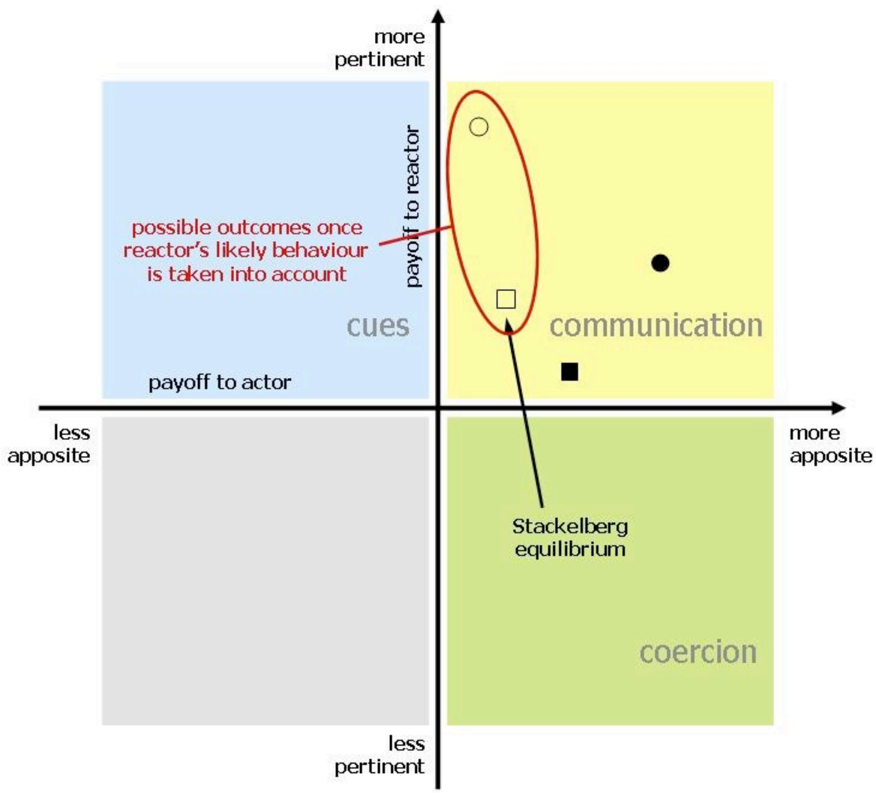


Figure 5.5: **Stackelberg equilibrium.** To maximise her own payoffs the actor should account for their interlocutor's likely reaction. Here the reactions that carry the greatest payoff to the reactor (i.e. are most pertinent) for each action are depicted by clear shapes. The actor thus chooses between these clear shapes. The one with the greatest payoff to the actor (i.e. is most apposite) is the Stackelberg equilibrium.

| | | reactions | | |
|---|---|---|---|---|
| | | X | Y | Ø |
| | A | ○ | ● | ⊕ |
| actions | B | □ | ■ | ▦ |
| | Ø | △ | | |

Figure 5.6: The addition of null actions and reactions. To our earlier figure (5.2, above) we have no added the possibility that either player could play the null behaviour of doing nothing, represented by ø. If the actor does this then there is no reaction necessary, which is why it defines the same outcome (the triangle) in all three cases.

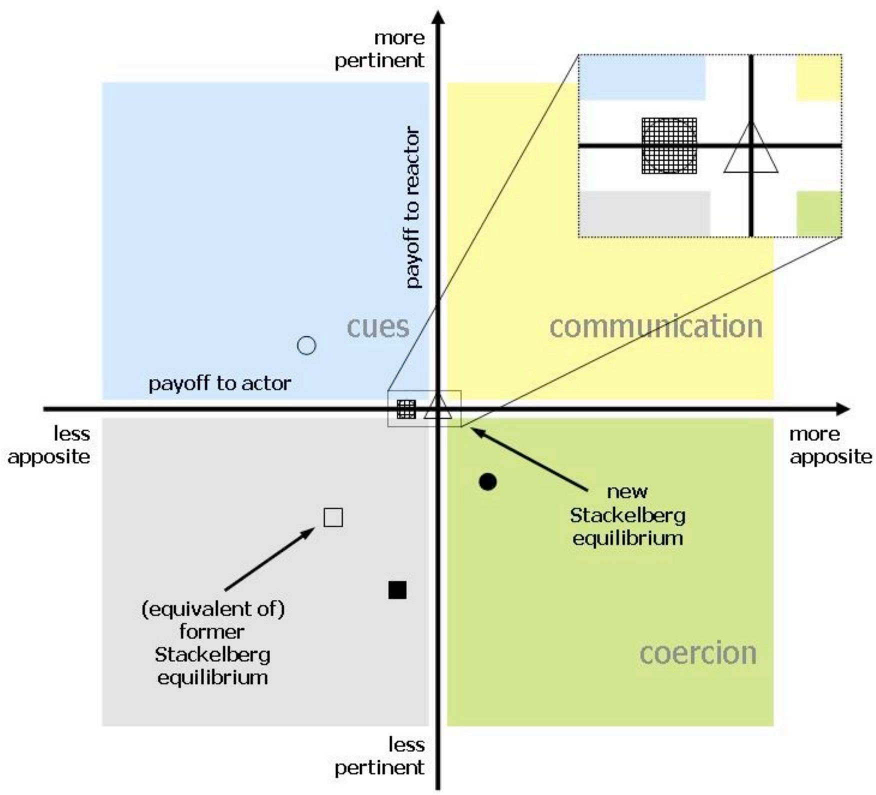


Figure 5.7: The inclusion of null actions. The triangle represents the actor choosing to do nothing, and the circle and square just to the left of the origin represent interactions where the reactor does nothing. Again, the clear square and clear circle represent the best reaction that the audience can give to the possible actions that correspond to those shapes. The actor also has the option of the clear triangle i.e. doing nothing, and in this case that is indeed his best option and hence the Stackelberg equilibrium.

A comparison of figures 5.7 and 5.5 (i.e. the scenarios in which doing nothing is and is not the Stackelberg equilibrium) shows that if the Stackelberg equilibrium includes neither the null action nor the null reaction then the interaction is necessarily both pertinent and apposite. Moreover, since participants maximise their payoffs then the reaction will be the most pertinent reaction possible. What reactions (and actions) are possible? Any and all. Although the illustrations above use only a simple model in which there are just two or three possible actions and reactions, there is no reason that this will not scale up. We might think that the sets of possible actions or reactions are constrained in some way by convention or other phenomena that seem to impact on what behaviours we are likely to perform. However, such considerations are not constraints on what is possible, but rather variables that determine the payoffs associated with each behaviour. A strong cultural convention to produce a particular signal at a particular time (saying "Fine, thank you" after someone has asked you how you are, for example) does not prevent or enable the production of that signal, but it does mean that deviations from that signal are unlikely to record positive payoffs.

I want to highlight some basic facts about this model. They may appear trivial, but they are important nevertheless, as will become clear when we ask what they mean for human communication. Moreover, if the above discussion of RT has primed you in the way I intended, then as I bring these particular aspects of the game to your attention you will be able to anticipate why they are important for the matters at hand. The first basic fact is one of the game's axioms: that reactors will maximise pertinence. Given two possible reactions to a given behaviour, reactors will choose the one that gives them the greater payoff. The second fact is barely more complex, and has already been summarised above: at equilibrium every action that is not the null action anticipates a pertinent reaction; and moreover, the anticipated reaction will be the most pertinent reaction possible. This is an important result, as when it is applied to human communication (see below) it leads us directly to the second (communicative) principle of relevance. I therefore want to state it more formally, so that there can be no doubt that it follows from the model as it is set up. We have, then, a set of possible actions, A, and a set of possible reactions, R, and these sets may be infinite in size. For each pair of actions $a_i \in A$ and $r_j \in R$, there will be an associated pair of payoffs $\Pi_A(a_i,r_j)$ for the actor and $\Pi_R(a_i,r_j)$ for the reactor (relative to doing nothing). We define the null action $a_0$ and the null reaction $r_0$ such that $\forall$ i, j, $\Pi_A(a_0,r_j) = \Pi_R(a_0,r_j) = \Pi_R(a_i,r_0) = 0$ and $\Pi_A(a_i,r_0) = -\varepsilon$, where $\varepsilon$ reflects the opportunity and/or energy cost of the action. Call the actually played actions and reactions a* and r* respectively. It is axiomatic that participants are rational maximisers, so the actual payoffs achieved will be:

$$\Pi_R(a^*,r^*) = \max_j \Pi_R(a_i,r_j) \qquad (*)$$

$$\Pi_A(a^*,r^*) = \max_i \Pi_A(a_i, r^*) \qquad (**)$$

We want to show that if $a^* \neq a_0$ then $\Pi_R(a^*,r^*) > 0$. This shows that a* is pertinent, and once that is true then the fact that a* is optimally pertinent follows immediately from (**). To see that a* is pertinent, suppose that $a^* \neq a_0$ but that, contrary to what we wish to show, $\Pi_R(a^*,r^*) \leq 0$. Then $\Pi_R(a^*,r^*)$ must = 0, since $\Pi_R(a_i,r_0) = 0$ by definition. But if $\Pi_R(a^*,r^*) = 0$ then a* can = $a_0$, since $\Pi_R(a_0,r_j) = 0$. This contradicts our supposition, which therefore must be false, and hence our assertion that if $a^* \neq a_0$ then $\Pi_R(a^*,r^*) > 0$ must be true.

Translated into less formal language, this means that at equilibrium any act that is not the null act will be performed in anticipation of a pertinent reaction. Moreover, when coupled with the fact that reactors maximise pertinence, that reaction will also be the most pertinent one possible. In case the relevance (no pun intended) of all this to RT is not yet clear, let me grant to these statements the status of principles of pertinence:

First principle of pertinence

At equilibrium, reactors will perform those reactions that are most pertinent to them

Second principle of pertinence

At equilibrium, every (non null) action anticipates a pertinent reaction; and moreover, the anticipated reaction will be the most pertinent reaction possible

These are Darwinian imperatives; basic qualities that all biologically evolved communication systems will necessarily possess. Equally fundamental is the other axiom of the model, that actors will perform the most apposite behaviours available to them – the second principle of pertinence depends upon this. Perhaps this axiom should, then, be called the principle of appositeness.

It should be borne in mind that the model has been derived not as a statistical and therefore noisy abstraction away from a set of observations, but rather as a consequence of a definition of communication; one that, as was discussed at length in chapter 3, works because it captures within it the dynamic nature of communication itself. The axioms of the model are thus, when taken together, little more than statements about what it means to communicate (with the additional criterion that the actions and reactions be directed at each other), albeit ones that exist at an abstract level. They get at the very nature of communication itself. The question now is what these principles imply for human communication.

## 5.3 Relevance as the mechanism of communication

Since the account of communication I have described above exists at the ultimate, functional level of explanation, the two principles of pertinence will apply to all evolutionarily stable communication. In each instance this functionality will be manifested in a particular way, and our concern here is with how this occurs in our own species. I have already telegraphed that the answer to this question is RT. To make that explicit, compare the two principles of pertinence derived in the last section with the two principles of relevance (note that RT assumes that human communication is at or at least is very near to equilibrium (cf. '..tends…'):

First principle of pertinence

At equilibrium, reactors will perform those reactions that are most pertinent to them

First principle of relevance

Human cognition tends to be geared towards the maximisation of relevance

Second principle of pertinence

At equilibrium, every (non null) action anticipates a pertinent reaction; and moreover, the anticipated reaction will be the most pertinent reaction possible

Second principle of relevance

Every ostensive stimulus carries a presumption of its own optimal relevance

The equivalence is striking. It represents a convergence of approaches that, like the one between the adaptationist and ostensive-inferential perspectives of communication discussed in the previous chapter, is pleasing because the two models have by-and-large been developed independently. Indeed, RT mentions evolution only once, as part of its defence of the first principle of relevance (as discussed above), and even then it is only in the postscript to the second edition (p.260-3). It is instead, like all theories of pragmatics, a theory built from the bottom-up, while the approach pursued in this thesis is very much top-down. The conclusion to draw is that RT describes the cognitive, Tinbergenian mechanism by which the Darwinian imperatives of communication are implemented in the human animal.

To drive the point home, let us return to a theme that was touched upon in chapter 2 – the idea that rather than hard-code all aspects of an organism's behaviour in the genome, natural selection will build calibration devices that will, in effect, act in its stead. Also mentioned, in passing, was the fact that it is possible to imagine a hierarchy of such devices, and I will now take a brief look at one specific suggestion as to what such a hierarchy might look like. Doing so will enable us to understand the relationship between natural selection and RT in more depth. The hierarchy I have in mind is the

Tower-of-Generate-and-Test (Dennett, 1995, 1996). At the bottom of the tower are Darwinian creatures, those organisms with hard-wired behaviours; if-then machines, if you will, for whom the mapping between genome and behaviour is relatively direct. At the next level sit Skinnerian creatures. As the name suggests, these entities have evolved the ability to adjust their future behaviour in the light of the consequences of previous behaviour. Simple neural networks belong at this level (Dennett, 1996). At the third level are Popperian creatures, who can rehearse the effect of their behaviours in their own mind before they act, and hence can be said to possess foresight: they can evaluate, to some degree or another, whether certain behaviours will result in desirable or undesirable outcomes. This is a commonplace strategy, employed by most animals more sophisticated than simple invertebrates (*ibid.*); it is a "good trick" (Dennett, 1995, p.77) regularly discovered by natural selection. There is, however, a more cunning trick reserved for the occupants of the next level: to recognise and be informed by "the *designed* portions of the outer environment" (*ibid.*, p.377, italics in original). That is, the inhabitants of this level, Gregorian creatures (named after Richard Gregory, a theorist in the role of information in intelligence (e.g. , 1981)), are able to make use of the fact that certain artifacts (be they tools, behaviours, physical traits or whatever) possess *functionality*. Arrival at this level of the tower is thus not just a matter of increased intelligence, but rather a new type of intelligence; Gregorian creatures can not just think, but they can think *about thinking*, including their own. As such, the artifacts in question become not just the products of intelligence but also endowers of intelligence: the pen and paper you might use to solve a difficult arithmetic task actually enhance your intelligence, in a quite literal way. Indeed, a popular idea in contemporary philosophy of mind is that such artifacts are quite literally part of the human mind (Clark, 1997; Clark & Chalmers, 1997). The final level in the tower is the scientific method: *directed* generate-and-test, where we choose which subquestions to ask (i.e. which experiments to do) precisely so as to lead us to answers to our main enquiries. This is not just thinking about thinking, but doing so with a very specific agenda, and it is so powerful a methodology that “all other varieties of generate-and-test are willy-nilly” (Dennett, 1995, p.380) in comparison.

The different levels of this tower are not intended as a taxonomy of intelligence – other hierarchies could be imagined, and we could dispute whether the tower would be improved with the addition or removal of particular levels. On the contrary, the point is instead to highlight the continuity between these different types of intelligence; whilst each level of the tower can be seen as representative of a different adaptive system, they are all part of the same Darwinian structure. For example, which of the Skinnerian creature’s behaviours are reinforced positively and which negatively will itself be shaped by Darwinian logic – organisms that pursue adaptive goals will be selected over organisms that pursue maladaptive goals. (As discussed in §2.1.3, this is one reason why learning and evolution are not just analogous but in fact contiguous with each other, a fact not lost on the father of operant conditioning: “where inherited behavior leaves off, the inherited modifiability of the process of conditioning takes over” (Skinner, 1953, p.83).) Thus, all organisms belong somewhere in this tower. Their home may not be exactly on any of the levels described: they could occupy hard to find half-levels in some far-flung annex, or the basement, perhaps – are there creatures whose every behaviour is entirely pre-specified, who are not even if-then machines? Humans, it should be clear, are Gregorian creatures: we are intentional beings (Dennett, 1987), not just able but compelled to detect and be informed by functionality (Csibra & Gergely, 2007). The question I want to ask is: what does this quality mean for communication?

The defining difference between the different levels of the Tower-of-Generate-and-Test is, as the name suggests, in the domain in which candidate behaviours are generated and tested. In Darwinian creatures, the domain is pure natural selection: progress is made when random mutations produce more adaptive phenotypes. In Skinnerian creatures the domain is associative learning – organisms come to avoid scenarios that produce negative outcomes and pursue those that produce positive ones. As such the testing takes place external to the organism. In Popperian creatures that testing is internalised – behaviours that are anticipated to produce positive outcomes will be preferred over the alternatives. Gregorian creatures do the same, but they can also reflect upon the very process of generate-and-test itself; "Where the Skinnerian creature is able to learn new behaviours, and the Popperian creature is able in addition to try out possible behaviours in mental simulation, the Gregorian creature becomes [additionally] able to actively think about its own thinking" (Clark, 2002, p.189). What this means is that the timescales over which the game of communication is played change from the ancestral to the instantaneous. Darwinian creatures converge upon equilibria phylogenetically, Skinnerian creatures ontogenetically and Popperian creatures by what we could call a 'psychogenetic' process. However, since Popperian creatures cannot model how their interlocutors will behave, the feedback that we have seen is necessary for the achievement of stability (recall that the actor's behaviour is in part informed by the expected reaction) cannot come exclusively from a psychogenetic source. The communicative behaviour of the Popperian creature must therefore be aided by a heavy dose of Darwinian- and/or Skinnerian-ism. A nice example is the vervet monkey calls discussed in the §3.3, whose receptive behaviour appears to have a Popperian aspect to it, in that it is modulated by some sort of assessment risk – vervets already aware of the danger will not automatically flee when an alarm call is heard. In contrast, call production will continue even when all other vervets are also calling. This demonstrates a lack of the insight inherent to Popperianism, and is instead Skinnerian or possibly even Darwinian in nature. Moreover, the reason the insight is missing is that it requires that the vervet recognise that the other vervets are themselves designed organisms. This specific type of insight is reserved for the Gregorian creature, who is able to recognise and interpret the goal-directed behaviour of others. This allows the necessary feedback to occur, and hence the entire adaptive process by which equilibria are found can now occur within the mind of a single individual more or less instantaneously.

We should predict, therefore, that theories about the mechanism of communication in Gregorian creatures place at their core a cognitive implementation of the Darwinian imperatives discussed in the previous section. This is precisely what RT does. It basic tenets, the two principles of relevance, are precisely that. This suggests the following conclusion: that RT, or at least something with the structure of RT, is the inevitable outcome of communication between Gregorian creatures. This is an endorsement of RT not just on evolutionary grounds, but also on its own terms: an articulation of exactly how Gricean intentions are recognised, which is the very skill that is unique to the Gregorian creature, is the explicit mission statement of RT. Given this outcome, we should ask about the main criticisms of RT. Are they substantial, and do they threaten to undermine the enterprise?

## 5.4 Criticism of Relevance Theory

There is a large degree of support for RT, as witnessed by the exponentially increasing literature on the subject (see e.g. Yus Ramos, 1998 for a review), but it is certainly not universal. In fact, as one (negative) review (Levinson, 1989) noted, it tends to divide opinion: readers either find it a powerful source of explanation or vague and unsatisfying (as an example of a positive review see Leslie, 1989; for a particularly negative one see Seuren, 1986; and in general, see the responses to Sperber & Wilson, 1982; Sperber & Wilson, 1987 and the multiple reviews in Mind and language 4(1-2), 1989). Interestingly, given that I have argued that RT follows from an adaptationist approach to communication, some of the most common criticisms (in particular circularity, excessive reductionism and non-falsifiability) have analogues in common criticisms of the adaptationist agenda. This should be no surprise if, as argued in this chapter, they share the same essential character. This section will briefly address these criticisms of RT, and argue that each rests either on a misunderstanding or, in the most problematic case, is not as damaging as it first appears, because the evolutionary perspective offers a possible response.

### 5.4.1 Circularity

Just as natural selection is sometimes mistakenly said to be circular (viz. the caricature that 'the most fit organisms are the ones that reproduce, and the ones that reproduce are the most fit'), so is RT (viz. 'what is relevant is that which has cognitive effects, and that which has cognitive effects is relevant'). With respect to natural selection the key observation to make is that fitness is dispositional rather than deterministic: possessors of an adaptation will not necessarily reproduce at the expense of those without the trait; they are just more likely to. So the correct characterisation of natural selection is not as above, but rather that 'the most fit are more likely to reproduce (by virtue of their adaptations) than others in their population, and those that do in fact reproduce are those that are most fit'. Put another way, natural selection's explanatory power lies in that fact it links together, on the one hand, the observation that traits often carry the appearance of design with, on the other, the fact that relative survival rates will determine what traits are passed on to the subsequent generations. This is not tautological; or if it is then so is, say, F=ma, the basic law of Newtonian mechanics (Sober, 1984).

A similar point applies to RT. Some (e.g. Seuren, 1986; Wilks, 1986) have argued that RT's notion of relevance is circular. The argument runs as follows: "if the determination of the context C for an assumption A is part of the processing of A, then the relevance of A may involve a search for a suitable C, so that A is relevant, but it makes no sense to speak of the relevance of A in C. The objection seems valid" (Seuren, 1986, p.141; following Wilks, 1986). If that seems too abstract a formulation then you are in good company; here is Wilson and Sperber's immediate reaction: "[the objection] may seem valid to Seuren: to us it merely seems hard to parse" (1986, p.154). The more substantial (and barely more complex) response is to point to the reductio ad absurdum of the argument. By the same rationale we can say that "the definition of solubility [is] circular, for if the determination of the liquid L for a substance S is part of the processing of S, then the solubility of S may involve a search for a suitable L, so that S is soluble, but it makes no sense to speak of the solubility of S in L" (ibid.). This is clearly nonsense. (Or even "silly" (ibid.) if you wish to express real distain; that is not Wilson and Sperber's usual style, but in this case they have been incensed by what they see as an unrepresentative and unprofessional review, for which they think the author "ought to be embarrassed" (ibid., p.161).) The reason the argument is nonsense is that relevance, too, is dispositional: individuals with tendencies to produce relevance-theoretic utterances will not necessarily induce more cognitive effects in their listeners than those without such relevance-theoretic instincts; they are just more likely to. Just as natural selection links two deceptively simple ideas, so RT links the Gricean observation that utterances carry an intention to communicate with the Austinian observation that they do things to the world (by way of positive cognitive effects). It is not at all circular.

### 5.4.2 Excessive reductionism

This is the claim that RT reduces the four Gricean maxims to one, a 'reduction' (I use scare quotes because it is not actually a reduction at all; see below) that is argued to be unrealistic (e.g. Kopytko, 1995). This claim demonstrates a fundamental misunderstanding of RT, albeit quite a common one (e.g. Blutner, 1998; Clark, 1996; Roberts, 1996): that RT collapses the four maxims into one (the obvious echoes that relevance has of the Gricean maxim 'be relevant' will plainly not have helped here). The reason that this is a misunderstanding is that Grice's maxims perform a quite different task to the principles of relevance – the theoretical status of the two ideas is quite different (see Wedgwood, 2005, p.48-50 for more on this misunderstanding). Whereas the former is a description of behaviours that speakers are simply thought to aim for, the latter are argued, on the basis of a number of simple observations, to be fundamental facets of human communication. Indeed, this chapter has sought to show that the functional role that is performed by the principles of relevance is in fact an intrinsic property of all evolved stimuli. The cooperative principle, in contrast, has no such grounding, and consequently the maxims "are only operable on the back of considerable amounts of prior context-dependent inference" (Wedgwood, 2005, p.49). Logic and Conversation was, after all, only intended to be tentative and programmatic, and as mentioned in §5.1 above, Grice himself recognised that a more secure foundation was necessary.

One interpretation of the charge of excessive reductionism is that it is symptomatic of a failure to appreciate just how much work the notion of relevance actually performs. The Gricean maxims, and indeed all the various neo-Gricean refinements that have followed, are descriptive accounts that approximate the realities of communication. As such, they are idealised abstractions away from something more fundamental – and it is this more fundamental feature that RT attempts to describe. As such, then, it should be possible to generate the Gricean and neo-Gricean accounts from basic considerations of relevance, and it is not difficult to see how this could be done. For example, it is not difficult to see how three of the four maxims can be generated by the atomised currency of RT: an optimally relevant utterance will also satisfy, first, the maxim of Quantity, since saying too little does not maximise cognitive effects, and saying too much demands increased processing effort for no additional cognitive effects; second, the maxim of Manner, since clarity and ease of comprehension speak to the minimisation of processing effort; and third, the maxim of Relation, practically by definition. The maxim of Quality is a slightly more complex case, and requires reference to the previous chapter, but is subsumed nevertheless (Wilson & Sperber, 1981, 2002b): speakers that wish to maximise cognitive effects could be dishonest, but if they are, then they will lose the opportunity to induce more cognitive effects at a later date, since the listener will lose trust in them. An open empirical question is how these future effects are discounted: what sort of payoff is necessary to induce a speaker into the production of a false utterance that will place at risk future opportunities to induce cognitive effects in the listener? The various neo-Gricean approaches, for example the reduction of the four maxims to two (Horn, 1984); a Q-principle (make your contribution sufficient; say as much as you can) and a R-principle (make your contribution necessary; say no more than you must) can also be generated from the basic consideration of relevance in a similar way. As such, then, RT does not reduce such accounts. On the contrary, the notion of optimal relevance is the fundamental feature of human communication that these imperfect generalisations approaches can only approximate.

### 5.4.3 Lack of a meaningful metric

As we have seen, RT defines relevance to be a trade-off between positive cognitive effects and processing effort; as the ratio of the former to the latter increases then so does the degree of relevance. This raises two related questions (Bach & Harnish, 1987; Gazdar & Good, 1982; Levinson, 1989). First, how can these two qualities be measured? And second, are they to be measured with the same metric? If they are not, then how can we trade them off against each other? This is a serious criticism (which, unlike the previous two, does not rest on a misunderstanding). Even those sympathetic to RT consider it "fair" (Wedgwood, 2005, p.50) and fear that it may be "damning" (ibid.). As we observed in §5.1, Sperber and Wilson's response is to argue that relevance should be measured in comparative rather than absolute terms, for this is more likely to provide a psychologically plausible starting point. This means that judgement of what is more or less relevant will, at least for the time being, rely at least in part on intuition. Although this is unsatisfactory is some regards, in others it is to a large extent inevitable, since issues of measurement are a habitual and almost inevitable problem of practically all psychological study; indeed, all mental processes involve some processing effort, and basically all psychological theories assume that if this increases then associated tasks will be performed less well (Sperber & Wilson, 1986). A related response (Wedgwood, 2005) is to ask whether this is really so unsatisfactory, given the alternatives. The situation we are faced with is that context-dependent reasoning relies on a complex set of factors over which it is very hard to generalise. There are, broadly speaking, two ways in which we can go: either we reason our way from certain minimal assumptions about what is involved in the production of meaning (which is the strategy RT employs, for better or worse), or we idealise away from the messiness of inferred meaning (as linguistics has traditionally done). The former is surely the correct choice if we are to take the existence of inference and the importance of context in communication seriously, and if that leaves us with a notion that is hard to measure in practical terms then so be it: that may be a price worth paying for a theory that captures the realities of ostensive-inferential communication. Indeed, if correct then such a theory could reasonably be taken to be a "fundamental component of any broader theory of language" (ibid., p.52).

Even better, the analysis presented in this chapter suggests a way in which the problem may be overcome; that is, a way in which both cognitive effects and processing effort (and hence relevance) may be meaningfully measured: in terms of contribution to inclusive fitness. The basic idea is that if two individuals, identical in all ways until a particular moment, produce at that moment two different utterances, then the individual who, on average, achieves the greater inclusive fitness is the one that produced the more relevant utterance (this is a similar point to that made about dispositionality in §5.4.1). Extrapolated across all utterances, this means that individuals who are consistently irrelevant will on average achieve a lower inclusive fitness than those that are not. As we already noted, most people are already pragmatically fluent; after all, irrelevance will lead either to misunderstanding or to social exclusion, whichever arrives sooner; and in a social species like ours, neither is desirable. However fluency is (like, say, attractiveness) a relative term, and there will still exist variation in the population, and that variation is an empirical matter. Actual experiments are hard to envisage (although there may be scope for the inventive use of proxy measures of fitness), and the metric is a distal and diluted one, somewhat removed from the usual loci of pragmatic explanation. The suggestion may thus be frustratingly inapplicable, but that is neither here nor there when it comes to the fact of the matter, which is that this is nothing more than the use of (well established, successful) social evolutionary principles in the analysis of pragmatic behaviour.

### 5.4.4 Non-predictiveness and non-falsifiability

Following the criticism that relevance cannot be measured (see above) comes the claim that it is non-predictive and therefore not falsifiable (e.g. Levinson, 1989): if there is no way to measure relevance then, so the argument goes, there is no way to predict how different utterances will be processed. Even if we leave aside the above responses to the issue of measurement, this is still misplaced. While it will remain true that in borderline cases firm predictions cannot be made, there will also be cases (like the thalassemia example in §4.1 above) where the matter is clear and precise measurement correspondingly unnecessary. Under such circumstances predictions can be made. Thus contrary to the claim that "it is not clear how it could be made to have clear empirical application" (ibid., p.456), RT does make predictions that are precise enough to be testable. Moreover, when the corresponding experiments have been carried out they have broadly confirmed the main predictions of RT (see van der Henst & Sperber, 2004 for a review).

To illustrate I will briefly describe what is perhaps the most well known experimental test of RT; or more precisely, of the communicative principle of relevance (that every utterance carries a presumption of its own relevance). The Wason selection task (Wason, 1966, 1968) begins with a narrative of some sort which introduces a rule of the form if P then Q. Four cards are then shown to the participants. On one side of each card is information about P (either P or not-P) and on the other information about Q (Q or not-Q). However participants are only shown one side or the other of each card, so that when a card displays information about P then the corresponding information about Q is not shown, and vice versa. Of the four sides they see one has P satisfied, one has P not satisfied, one has Q satisfied and one has Q not satisfied (termed the P, not-P, Q and not-Q cards respectively). Participants are then asked to select all and only those cards where the hidden information must be made visible in order to judge whether the cards satisfy the rule. The whole task can be presented in abstract terms, as it is here, or in the context of some narrative framework, for example a bouncer who has to ensure that no under 18s (P or not-P) are drinking alcohol (Q or not-Q) (e.g. Griggs & Cox, 1982). The correct answer is the two cards P and not-Q. Success rates vary: when the task is presented in abstract form very few participants give the correct answer, but in more concrete versions success rises quite dramatically (see Evans, 1982 for a review). RT offers a possible explanation of these differences. More precisely, the communicative principle of relevance predicts that this difference can be explained in terms of how the context changes expectations of relevance, and that this will in turn influence responses to the task. To test this the instructional narrative was manipulated such that the addition or removal of certain contextual (but not logically necessary) information would vary the (comparatively measured) cognitive effects induced by the task and the processing effort required to comprehend it. This defines four conditions (effect+/effort-, effect-/effort-, effect+/effort+ and effect-/effort+) and, as predicted by RT, success rates are at their greatest in the first of these, where relevance is maximised, and at their least in the last, where relevance is minimised (Sperber et al., 1995). This result is presented here not as conclusive evidence in favour of RT (if there could even be such a thing) but rather as an illustration of how the core ideas of RT can be empirically examined.

Given that this thesis is an exercise in the application of Darwinism to human behaviour and cognition, I should offer a few summary comments on the relationship between this RT work and the high-profile evolutionary psychology (hereafter: EP) studies with the Wason selection task (Cosmides, 1989). The greatest differences between these two accounts lie in issues of modularity and domain-specificity (see Fiddick et al., 2000; Sperber et al., 1995; Sperber & Girotto, 2002). Some versions of EP (what could be called Evolutionary Psychology, or 'narrow' EP, to be distinguished from evolutionary psychology (Buller, 2005; Gray et al., 2003), which I interpret roughly synonymously with human behavioural ecology) include a theoretical commitment to massive modularity (Samuels, 1998). This, coupled with experimental results from the selection task, is used to argue in favour of a cheater detection module (Cosmides, 1989). RT is also taken to be a domain-specific adaptation (Sperber & Wilson, 1986), but is more general than the very specific cheater-detection module postulated by the narrow EP approach. Since results with the selection task can be explained exclusively through broad considerations of relevance (Sperber et al., 1995; Sperber & Girotto, 2002), then we can only conclude that the task does not allow us to adjudicate on the existence or otherwise of a cheater-detection module, or indeed any other specific form of reasoning. This is not to deny the existence of a cheater-detection module; it is just argued that the Wason selection task cannot provide evidence one way or another, since its results are more parsimoniously explained by reference to general considerations of relevance.

### 5.4.5 The need to be 'applied relevance theorists'

The four criticisms discussed thus far are the most common ones levelled at RT. There are others, many of which shed more insight on the commentator's own theoretical biases and/or misunderstandings of RT than on RT itself. I bring attention to just one (quite quirky) example, which I have chosen because it parallels a common misunderstanding of kin selection (Hamilton, 1964), a notion that is utterly central to evolutionary biology (Griffin & West, 2002). It is the claim that RT demands that individuals be "applied relevance theorists" (Bach, 2006, p.8), which echoes the assertion that kin selection cannot apply to simple organisms because such creatures cannot mentally represent their family tree (e.g. Sahlins, 1977). Both are wrong, and for similar reasons. Humans do not calculate the relevance of each possible utterance. Rather, relevance is a general quality of ostensive-inferential communication itself, one that follows from very basic assumptions about how communication works. Consequently, "communicators and audience need no more know about the principle of relevance to communicate than they need to know about the principles of genetics to reproduce" (Sperber & Wilson, 1986, p.162), just as ants need no more know about Hamilton's rule (1964) to maximise their inclusive fitness and snails need no more know about logarithms to build their shells (Dawkins, 1979). Consequently, the processing effort and cognitive effects that are at the centre of RT are non-representational: "they exist whether or not the individual is consciously assessing them, [and] whether or not they are conceptually represented" (Sperber & Wilson, 1986, p.131). A similar statement could not, incidentally, be made about the Gricean perspective, which posits that speakers and audiences must know the maxims of conversation and recognise when they have been flouted. It is thus this approach, rather than RT, that attributes to interlocutors what are psychologically unrealistic capabilities.

In general, then, the main criticisms of RT as a theory of pragmatics either rest on misunderstandings or are not as fatal as is often claimed. Moreover, the evolutionary approach strengthens the RT response to the most substantial criticism – that of measurement. Where does this leave us? At the beginning of this chapter I noted that a complete answer to the question of the mechanism of communication would involve, virtually by definition, more-or-less the entirety of linguistics. The goal of this chapter was to describe the social evolutionary constraints under which that machinery must operate. This gives us a framework, RT, with which to understand and think about other aspects of linguistics, and so one natural next step would be to explore how this would work in practice. However that project is ongoing. We are thus already able to point to many instances of the application of Darwinian imperatives to matters of linguistic concern, for that is what any and all applications of RT are (even if this is not recognised by its practitioners). This consequently affords us the luxury of being able to focus instead on matters of particular concern to those interested in the evolution of language, for example the emergence of learnt symbolism – which is focus the of the next chapter.

# Six

# The emergence of learnt, symbolic communication

It is time to take stock. We have argued that the biological function of language is communication, appropriately defined, and used that fact to argue that if our definitions and theories of language are to be evolutionarily plausible then we must take the existence of ostension and inference seriously. We then considered the evolutionary problem of stability and saw that it is most likely solved by the human capacity to keep track of other individuals' past behaviour, and to modulate our future interactions accordingly. We then turned to the question of mechanism, and defined some very basic constraints that our theories of pragmatics must satisfy – and observed that these constraints are embedded into the core of one particular approach to pragmatics. The naïve social evolutionist may be content to stop there, since we have coherent answers to the twin evolutionary questions of function and mechanism: language is a tool for doing things to one another, and a large part of how this is achieved involves our capacity for relevance-theoretic communication and comprehension. However, there is a sense in which it would be unsatisfactory to declare our work complete at this point. To help get a better grasp on the reasons for that unease, reconsider the following, which was originally quoted in chapter 2:

> "What kind of biological evolution [is language the result of]? Well, here you have to look at the little bit we know. We can make up lots of stories. It is quite easy: for example, take language as it is, break it up into fifty different things (syllable, word, putting things together, phrases and so on) and say: 'OK, I have the story: there was a mutation that gave syllables, there was another mutation that gave words, another one that gave phrases… another that (miraculously) yields the recursive property (actually, all the mutations are left as miracles).' OK, maybe, or maybe something totally different; stories are free and, interestingly, they are for the most part independent of what language is… the story you choose is independent of the facts, pretty much." (Chomsky, 2002, p.146)

Now, there is clearly a lot wrong with this characterisation of how adaptationism works. I do not intend to get involved in a fine-grained dissection of how and why this is incorrect. Instead, I bring attention to it for a more positive reason: that despite its wrongness, Chomsky's quote does hint at a concern that many linguists will experience when presented with the social evolutionary perspective on language that I have articulated thus far. The concern is this one: that this approach has very little to say about precisely those aspects of language that are most remarkable about language and which, consequently, are those that are of most interest to most linguists. Language is a biologically noteworthy phenomenon for many reasons (Hockett, 1960), of which this thesis has discussed very few so far, and none (or perhaps one, if we include our capacity for relevance) in depth. At a minimum it is not immediately apparent how the present discussion might inform research into phenomena like the critical period (Lenneberg, 1967), syntactic structure (Chomsky, 1957, 1965) or the duality of patterning (Hockett, 1960), to pick just three interesting aspects of language. This means that the approach as it presently stands is of limited appeal.

It is for this reason, perhaps, that evolutionary linguistics has shown more interest in questions of history and phylogeny than those of putative adaptive function. For example, an introductory paper entitled Language Evolution: A Brief Guide for Linguists (Bickerton, 2007) lists the emergence of syntax and symbolism as the primary problems for the field, with the matter of function relegated to a second tier of importance; and many of the landmark books published in the last 15 years (e.g. Calvin & Bickerton, 2000; Deacon, 1997; Hurford, 2007; Tomasello, 2008) focus on historical and/or phylogenetic matters rather than functional ones (an exception is Dunbar, 1997). This contrasts with the situation in what one might assume would be the neighbouring discipline of evolutionary psychology, where function and mechanism are given prime explanatory status (Barkow et al., 1992). However for the reasons discussed, the two have quite different foci. The challenge, then, is to articulate in detail how social evolution (and indeed pragmatics) can inform the traditional concerns of linguistics. In its entirety that is a huge research agenda that is well outside the scope of a single PhD. thesis, but that should not prevent tentative enquiry into one or two individual aspects. Correspondingly, the goal of this chapter is to examine how the theoretical apparatus constructed thus far might help explain the existence of one particular feature of language. The one I have chosen to focus on is learnt symbolism, and more precisely the fact that language depends on the existence of a vast number of symbols (meaning-form pairs in which the relationship is arbitrary; de Saussure, 1959) that must be both learnt anew by each generation and shared by the entire speech community. These twin features, learning and symbolism, give rise to an emergence problem: if there is no relationship between form and meaning, and if meanings are not innately specified, then how can individuals agree on what forms should refer to what meanings (Oliphant, 2002)? There is no particular epistemological reason why I have focused on this problem in preference to the many other suitable candidates, except that it seems a tractable one, and it has attracted a fair degree of previous research interest (e.g. Bickerton, 1990; Hurford, 1989; Tomasello, 2003b); and as noted above, it is considered by many to be a central problem for evolutionary linguistics (Bickerton, 2007).

How will it be investigated? The next section uses the definition of communication articulated in chapter 3 to review some previous work in the emergence of communication, and finds that in a range of instances of such emergence cues or coerced behaviours appear first and are then co-opted by the other participant for communicative purposes. This gives us a working hypothesis for how the Saussurean sign might have emerged. The second section introduces the Embodied Communication Game (hereafter: ECG) that was specifically designed with the problem of emergence in mind. (Although I identified and articulated the theoretical constraints that the game must satisfy (see §6.2), the game's final form was developed in a brainstorming session with Simon Kirby and Graham Ritchie. Graham also programmed the ECG.) The third section reports the results of my experiments with the ECG and the fourth discusses the implications and conclusions of this work, and also draws attention to the importance of earlier parts of this thesis on the design, implementation and interpretation of these results. The final section discusses the role of intentionality in the ECG, and its relationship with the evolution of social intelligence.

## 6.1 Three case studies of the emergence of communication

In chapter 4 we identified three types of cooperation relevant to communication: communicative, informative and material. The first two are necessary for evolutionarily stable communication; the final one not so. The fact that there are two necessary levels makes communication an inherently social act (Clark, 1996; Grice, 1975a; Tomasello, 2003a; Wilson, 1975), even though it can be used deceptively. Moreover, signals and responses depend, by definition, upon the existence of each other: one without the other is neither a signal nor a response. There is, then, something of a chicken-and-egg problem with regard to the origins of communication in any form, be it learnt or innate, symbolic or iconic (or even indexical). This section will review three case studies of the emergence of communication from non-communicative beginnings, and use them to draw general conclusions about how such a process must necessarily occur. The case studies are: (i) an artificial life study into the evolution of communicative behaviour between pairs of simulated robots; (ii) ethological accounts of the biological evolution of animal signals; and (iii) ontogenetic ritualisation, in which pairs of organisms are observed to create shared communicative conventions unique to that dyad. These case studies have been chosen because in all of them the question of how receivers even know that signals are signals is not satisfied *a priori*. This is, as is discussed in §6.2, an important constraint that is not satisfied by previous experimental approaches to the emergence of symbolic communication between humans. It is for this reason that such work (which is also discussed in §6.2) has not been included in this brief survey.

It will be shown that in all three case studies communication emerges not spontaneously but instead via the creation of either a cue or some form of coercion. These behaviours are then co-opted by the other half of the dyad and come to take on a communicative function. In effect, either the productive or receptive half of the eventual communicative scenario is put in place first and then co-opted by the other half. Since much of this section will depend upon the terminology discussed in chapter 3, figure 6.1 provides a reminder of the key terms and their relationships to each other.

| | action designed to trigger reaction? | reaction designed to be triggered by action? |
|---|---|---|
| communication (signals; responses) | Y | Y |
| cue | N | Y |
| coercion | Y | N |

Figure 6.1 (reproduction of figuree 3.1): The relationship between communication, cues and coercion.

### 6.1.1 Simulated Khepera robots

In this first case study (Quinn, 2001), pairs of simulated robots are evolved according to their ability to solve a coordination task, but are not given a dedicated communication channel with which to achieve this goal. The robots are equipped only with two motor-driven wheels and eight infra-red sensors, and their behaviours are controlled by an 'innate' (that is, fixed and not able to learn) neural network. Pairs of robots are placed in the centre of an obstacle-free environment, and their task is to travel as far as possible from that point but in the same direction as each other. If they travel in opposite directions the distances covered cancel each other out, and their final fitness score is zero. Hence success at this task demands coordinated behaviour; robots that do not travel in an at least approximately similar direction to each other will score poorly. The robots are evolved using a genetic algorithm; the best-performing robots are selected for the next generation, with crossover and random mutation on the neural networks.

The communication that results from this process typically takes the following form. First, the robots rotate counter-clockwise. Once one of the robots detects the presence of the other it oscillates back and forth, waiting for the other to rotate to face it.

Once that occurs then the robots move off together in the same direction. The oscillatory behaviour of the first robot to align with the other is communicative, in that it signals to the still rotating robot that it has aligned with the other and is ready to travel (Quinn, 2004). This arrangement can be observed in figure 6.2.

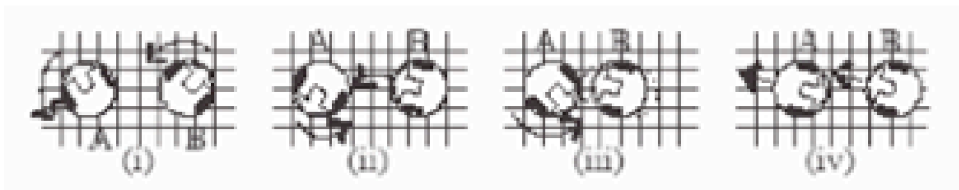


Figure 6.2 (reproduced from Quinn, 2001): Communication between two Khepera robots. In (i) the two robots rotate in an anticlockwise direction. This continues until (ii) one of the robots (in this case B) is aligned with the other. Then in (iii) the aligned robot oscillates back and forth while the other robot continues to rotate until aligned. Finally, in (iv), the robots travel off together.

How did this solution emerge? Initially the robots simply proceeded in straight lines from their starting positions. This was sometimes successful (if the robots travel in approximately the same direction as each other), but equally often it resulted in collisions, which carry a fitness penalty. Hence the robots quickly evolved collision-avoidance behaviours; when near another robot they would either halt or rotate. If this occurred when the robots were heading directly towards each other then a form of deadlock would arise, in which the two robots would variously halt or rotate in response to the close proximity of the other robot, but then travel towards each other when no longer in each other's immediate vicinity. This idiosyncrasy is the result of two almost but not quite conflicting evolved behaviours: remain close to the other robot; and do not collide with the other robot. More precisely, both the presence and the non-presence of the other robot acts as a cue. Both alter the other robot's behaviour and the other robot's reaction is designed to be affected by the stimulus: either to move away somehow, so as to avoid collision; or to try and find the other robot, as fitness is maximised when the robots are close together. However, the other robot's presence/non-presence is not designed to elicit such responses, and so the stronger definition of a signal is not satisfied.

The deadlock is broken, many generations later, when one robot evolves, by random mutation, to advance when faced with such a situation. The other robot, faced with imminent collision, must now evolve to reverse away from the advancing robot. The two robots are now travelling in the same direction, and by continuing to do so they thereby succeed at the task. There is then a pressure for the evolutionary process to find more efficient ways to arrive at and resolve the original deadlock situation, with the eventual communicative result as described above and illustrated in figure 6.2. This final state is communicative: the oscillations of robot A indicate to robot B a readiness to travel in the direction defined by the robots' alignment, and is functionally designed to do so; and the response to this (actually travelling in that direction) is a functionally designed behaviour. If we wish to anthropomorphise, we may gloss the oscillations as 'after you' (Kirby, 2002). The key point is that the robots first evolve one half of the communicative interaction and that this then provides the framework for the other half to evolve. Hence we arrive at a communicative set-up, and the important point for the present purposes is that the cued collision-avoidance behaviour, whilst not communicative itself, appears to be a critical step on the route to communication.

### 6.1.2 The evolution of animal signals

On a similar note, ethological accounts of the origins of communication in the natural world have long emphasised ritualisation: the exaptation of a cue for the purposes of communication (Huxley, 1966; Lorenz, 1965; Tinbergen, 1952). Many examples of cues being exapted for use in communication have been identified and classified; they include threat displays, piloerection, protean display and olfaction (for extensive detail and discussion see Bradbury & Vehrencamp, 1998; Hauser, 1996; Maynard Smith & Harper, 2003). Historically (Darwin, 1889; Tinbergen, 1952) it has been thought that this process must lead to evermore conspicuous and exaggerated signals, the result of an arms race between signaller and receiver, but computational models (Noble, 2000) show that this is not necessarily the case. An alternative process is *sensory exploitation*, in which an organism's receptive abilities are exploited by some other organism for its own ends (see Ryan, 1990). For instance, female birds may search preferentially for red when foraging, because they only see red on certain seeds that are good for them. A male that adds red to its plumage may be able to exploit this preference and thus gain more mating opportunities (Bradbury & Vehrencamp, 1998). This would be an instance of coerced behaviour. However this will now feed back into the female's inclusive fitness. If the coercion affects her positively then she may evolve a more enhanced or nuanced preference for red. More generally, if a new stable state arises once the feedback has been incorporated then her response is now a designed one, and we can thus say that the plumage is a signal. (A terminological note: there will be scenarios where *exploitation*, with its negative connotations, is not such a good term. For this reason some authors prefer the term *sensory bias*.) Furthermore, mathematical models that take into account what the (future) reactor is most likely to do in the absence of a signal suggest that ritualisation is likely to be a more common process than sensory exploitation (Bradbury & Vehrencamp, 2000). This is because signals must be reliable (as discussed in chapter 4), and any cue that is systematically used by reactors will already be reliable, by definition. Coerced behaviours have no such requirement: what is reliable here is the reaction, but that does not solve the problem of evolutionary stability.

Are there any other ways in which biological evolution can give rise to communication? It cannot happen directly through selection for reasons that are already familiar: signals require responses and responses require signals. Indeed, the whole point of this section is to ask how that chicken-and-egg problem can be resolved. Ritualisation and sensory exploitation do so by selecting first for either a cue or a coerced behaviour respectively and then later, once that behaviour is established, for the other half of the equation. The only other alternative is drift. Whilst possible in principle, this requires not just one side evolve an action and the other a reaction, but that the two be mutually compatible and emerge, independently of each other, at more or less the same time. Such conditions make this an extremely unlikely scenario.

### 6.1.3 Ontogenetic ritualisation

Ontogenetic ritualisation is a process observed in which pairs of great apes (including humans) shape each other's behaviour through a series of repeated interactions (Tomasello & Call, 1997), and over time these behaviours come to take on a communicative role. There are two classic examples. The first is the chimpanzee 'nursing poke', in which an infant, held by his mother, pokes the mother's arm so as indicate a desire to feed at her breast (Tomasello et al., 1985; Tomasello et al., 1989). The second is the human infant's 'arms up' behaviour, used to indicate a desire to be picked up by an adult (Lock, 1978). What is the exact process of ontogenetic ritualisation? With the nursing poke, the infant initially attempts to move the mother's arm. Once the mother has detected that this is what he wants she raises her arm to allow access to her breast. As she becomes increasingly sensitive to his intentions she moves her arm as soon as he begins to attempt to move it himself. This interaction occurs sufficiently frequently that eventually the infant only need poke the mother's arm for her to react. 'Arms up' tells a similar story. Initially adult humans must force their hands into a child's armpits in order to lift them. Over time the child comes to lift their arms of their own accord whenever an adult goes to lift them. Finally they lift their arms to indicate their desire to be picked up. The general form of ontogenetic ritualisation can be summarised as follows (Tomasello & Zuberbühler, 2002): (i) individual A performs behaviour X; (ii) individual B reacts consistently with behaviour Y; (iii) based on the initial step of X, B anticipates A's performance of X and hence performs Y; and (iv) A anticipates B's anticipation of X and hence produces X in ritualised form so as to elicit Y.

How does this fit into the general picture that cues or coerced behaviours precede communication? Just as a push is an instance of coercion, so behaviour X (in its initial, non-ritualised form) coerces individual B; who (step (ii)), allows that to happen.

However this is not the behaviour that will be co-opted for communication, which is instead the initial stage of X that (in step (iii)) acts as a cue to individual B. That co-option then happens in (iv). Consider first the nursing poke. Initially the infant (individual A) attempts to move the mother's arm (behaviour X), and the mother (individual B) allows her arm to be moved (behaviour Y). These are steps (i) and (ii). Then, in (iii), the infant's initial movements act as a cue to the mother that the infant wishes to feed: it alters the mother's behaviour and the mother's reaction is designed to be altered in just this way. However, the infant's initial movement is not yet designed for this purpose, and hence it does not qualify as a signal. It is only in (iv), when the infant produces behaviour in a ritualised form, that this final criterion is satisfied, and hence that we can label the interaction communicative. The important observation for the present purposes is that to reach this stage we had to first establish a cue, and that this cue was then used to bootstrap the final signal. This example serves to illustrate the sense in the label ontogenetic ritualisation, since it is strictly analogous to the phylogenetic process of ritualisation in animal signals as described above, and which is occasionally called phylogenetic ritualisation (Burling, 2000).

'Arms up' is a slightly more complex case. Steps (i) to (iii) are the same as above, but step (iv) is slightly different, as it is individual B, rather than A, that uses the context provided by the cue to produce a signal. In step (i) the adult (individual A) lifts the child (behaviour X), and in step (ii) the child (individual B) lifts their arms as they are lifted (behaviour Y). Then the child begins to recognise the initial movements that are associated with lifting and so lifts their arms in anticipation. This is step (iii), and the parent's initial movements qualify as a cue: they alter the child's behaviour and the child's response is designed to be altered. However, the parent's behaviour is not presently designed to elicit this reaction. Now, in the general pattern of ontogenetic ritualisation described above, the next step would be for the parent to produce the initial step in some sort of ritualised form in order to elicit the 'arms up' reaction from the child. Intuitively this seems a likely next step. However, what is usually described is that the child – individual B – produces their reaction in a ritualised way. In this case it seems that it is the reactor that uses the context to bootstrap the final communication system. However, it should be born in mind that the original description of this process (Lock, 1978) was part of a wider discussion of how the child discovers that it can refer to objects and events in the world and hence communicate with others; and as such, it was focused on the child's behaviour rather than the adult's. It may well be the case that adults do perform step (iv), but the original report was not focused on this. The child's use of 'arms up' to indicate their desire to be lifted may possibly represent a fifth stage, in which individual B proactively performs behaviour Y so as to elicit behaviour X. Whether or not this is correct, what is important for the present purposes is that at stage (iii), as with the nursing poke, a cue has been established and that this fact provides the necessary context for a signal-response pair to emerge.

At a minimum, these case studies illustrate why the likelihood that communication will emerge would be dramatically increased if it is possible to first develop either a cue or a coerced behaviour. Design is built into one side of the equation and this creates pressures for design to appear on the other side. It may even be the case that cues or coerced behaviours are all but necessary stepping-stones on the path to communication. (They cannot be entirely necessary since, as outlined above, it could be possible for the productive and receptive capabilities to emerge or evolve simultaneously and coincidentally at the same time, although this is obviously extremely unlikely.) Furthermore, it is interesting to note that the case studies provide a range of cases that include both phylogenetic and ontogenetic emergence, as well as naturally-observed and laboratory-based studies. Figure 6.3 illustrates this point, and in doing so highlights the need for a laboratory study of ontogenetic emergence. The work presented in this chapter will fill that void.

| | | type of study | |
|---|---|---|---|
| | | laboratory-based | naturally-occurring |
| timescale of emergence | phylogenetic | simulated Khepera robots | animal signals |
| | ontogenetic | | ontogenetic ritualisation |

Figure 6.3: Classification of the case studies of the emergence of communication. In all three of these case studies cues or coerced behaviours emerge first and these are then co-opted for communication. The work reported in this chapter, a laboratory-based study of ontogenetic emergence, fills the gap in this figure. Other studies do not fill this void, since, for reasons of embodiment, the communication systems they investigate are not truly emergent (see §6.2, below).

(Interestingly, a similar process to that described above can be observed in language change. In reanalysis (Hopper & Traugott, 1993) linguistic forms (be they syntactic, semantic or whatever), designed to achieve one purpose, are reinterpreted along different lines. The original form is a cue: its production was not designed to illicit the reaction that it does, but the reaction is designed. That is, the listener has assigned it a function, but a different one to that intended by the speaker. It is then later used in this new way, and so can then (assuming it is interpreted correctly) be called a signal. It has thus been suggested (Burling, 2000, 2005) that extreme but still ordinary instances of this process may have been crucial to the evolution of language. This is perfectly plausible, but it should be noted that the alternative metaphor-based account (Heine, 1997) is equally so (Hoefler & Smith, 2008); here, language change occurs via coercion rather than cues: the use of, say, spatial motion to express intention (as occurs with ‘going to’; Heine & Kuteva, 2002; Hopper & Traugott, 1993) coerces the listener into a new interpretation of the stimuli. It is only once they have assimilated this usage into their own lexicon that the stimuli can be termed a signal.)

## 6.2 The Embodied Communication Game

Over the past 15 years or so, many computational and mathematical models (Brighton et al., 2005; Griffiths & Kalish, 2007; Kirby, 2007; Nowak & Krakauer, 1999) have been built in an effort to explore issues in the emergence of communication (see Kirby, 2002; Steels, 2003 for reviews). In part informed by those results, the last few years have witnessed the development of a number of empirical approaches (de Ruiter et al., 2007; Fay et al., 2004; Galantucci, 2005; Garrod et al., 2007; Healey et al., 2007; Selten & Warglien, 2007) to the question of symbolic emergence. This development is undeniably welcome: as mentioned in chapter 1, there is a dearth of experimental data in the study of language origins. As a result all of these studies have helped to define and shape the space of theories about how dyads and populations might converge upon shared meanings. They do this not because they replicate the evolutionary history of language (how could they?), but because they allow us to investigate the precise nature of the cognitive and cultural phenomena that underpin the emergence of shared symbolic communication systems. Such experiments thus provide us with key information about the biological preadaptations for language, and this knowledge then allows us to make inferences about the most likely sequence of events in the evolution of language. However, the conclusions that we can draw from previous studies with respect to the emergence of communication are limited by the fact that all of them (with the exception of those discussed in the previous section) circumnavigate the problem of how individuals recognise behaviours as communicative in order to focus on how signallers and receivers can agree on shared meaning-form pairs once the fact of communication is already established. Yet if we wish to investigate the question of emergence we must address the more foundational problem: how do (potential) receivers even know that there is a signal? Put another way: how do signallers signal signalhood? Or, to conceptualise the problem a third way and this time in terms of pragmatics: how do we communicate and recognise communicative intent?

The many studies (experimental or otherwise) that have investigated the emergence of communication have avoided this problem in (at least) one of three ways. First, the communication channel may be pre-defined (e.g. Fay et al., 2004; Galantucci, 2005; Healey et al., 2007). This will plainly evade the issue since participants know that any inputs that come to them via the communication channel are (almost certainly) communicative in nature. Second, the roles of signaller and receiver may be pre-defined (e.g. de Ruiter et al., 2007; Garrod et al., 2007; Selten & Warglien, 2007). Although this does not make communicative behaviour quite so salient as a pre-defined communication channel, it nevertheless primes the receiver to interpret the signaller's behaviour in communicative terms. Finally, the form space may be pre-defined. That is, the possible forms that a signal might take may be pre-specified by the researcher. If this occurs then whenever such a form appears then the receiver is apt to recognise it as a symbol; their challenge is to map that signal onto some meaning. Such an approach is inherent in game-theoretic accounts of communication (e.g. Lewis, 1969) but may also be seen in some computational (e.g. Hurford, 1989; Smith, 2004), mathematical (Nowak & Krakauer, 1999) and experimental (Selten & Warglien, 2007) work. In sum, all previous studies have pre-defined at least a partial solution to the foundational problem discussed above. Consequently they do not study the *emergence* of communication but rather the subsequent question of how to map form onto meaning. Put another way, previous research has studied how informative intentions (or their modelled equivalent) are satisfied, but has not asked the same about communicative intentions. The one previous study that does not avoid this question is the research with the simulated Khepera robots discussed in §6.1.1 above (Quinn, 2001), and that study was a major source of inspiration for the ECG. However, that work studied the emergence of an innate, iconic system, whereas our specific focus here is the emergence of communication that is learnt and symbolic.

We must, then, be careful to ensure that the problem of how individuals recognise that a given behaviour is communicative in nature is not resolved by the investigative set-up. This means, at a minimum, that neither the communication channel, the roles of signaller and receiver, nor the form space is pre-defined, as such. More generally, the problem's solution must not be an artifact of the experimental design, and we must instead allow communicative behaviour either to emerge from non-communicative behaviour or be created *de novo*. Importantly, therefore, the task should not be one that can be solved with a deductive choice of the most suitable channel from a number of candidate possibilities. Instead, we must insist that participants co-opt their behaviours in the world for communicative purposes. In short, we must demand that communicative behaviour be *embodied*; there should be no *a priori* distinction between communicative and non-communicative behaviour.

The ECG was specifically designed with this constraint in mind. It is an interactive, cooperative two-player game, played over a computer network, in which pairs of participants must coordinate their behaviour to solve a simple task where they lack shared information, yet where they have no interaction with each other except for their movements within the game's world. This means that these movements must perform both tasks necessary to succeed: (i) travelling within the world; and (ii) communication. Consequently, participants must not merely agree on what behaviours correspond to what meaning, but when creating these symbols they must find a way to signal that a given behaviour is a signal. For many participants it is not obvious how they can achieve this goal: many of the pairs of participants are unable to find any form of communication whatsoever (see §6.3 below). This is because the ECG uniquely demands not only that the participants agree on what movements will correspond to what meanings, but that the participants realise that they are able to use their movements to signal to each other at all. Then, once they recognise this, they must find some way to signal the fact that some of their movements are communicative in nature.

In the ECG each player is represented as a stick man, each located in his own 2x2 box. Each of the four quadrants is coloured either red, blue, green or yellow, at random. Each player sees both boxes, and the movements within them, but can see only the colours of their own box. The players move from the centre of each quadrant to the centre of the other quadrants, so they are unable to trace out letters or other symbols with their movements. They press the space bar to finish. Once both players have done so the colours of all quadrants are revealed to both players. If they have finished on identically-coloured quadrants they score a point; if not then no point is scored. Both players then press space again and a new round begins. Screenshots of each player's view, both before and after both players have pressed space to finish the round, can be seen in figure 6.4.

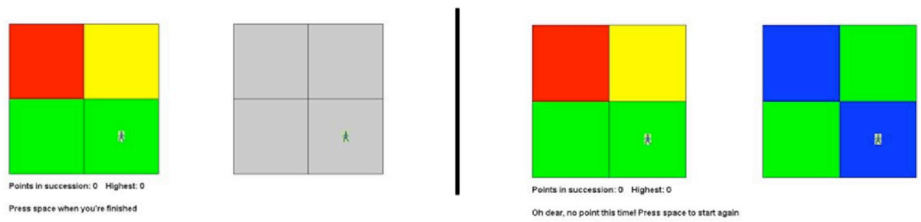


Figure 6.4: Two screen-shots of the ECG. Participants play multiple rounds of the game on networked computers. These screen-shots show the view before (left-hand side) and after (right-hand side) the participants press space. Participants can see their own colours but not the other participant's. Participants move around their boxes at will, and their movements are fully visible to the other participant. At any time the participants may choose to press space, after which they can no longer move during that round. When both participants have done this then all colours are revealed to them. Participants score a point if they finish on the same colour as each other. Here the participants have failed to score a point because they have finished the round on different coloured squares. After each round, the squares are reassigned colours randomly, although there will always be at least one shared colour (in this case, green). Solving the game requires finding some way to communicate the intended destination colour each round.

The colours of all quadrants are randomly assigned in every round, with the proviso that at least one of the four colours will appear in both boxes, so that it is always in principle possible for the players to score a point. The instructions to the participants were explicit that the colours would be randomly distributed, since pilots suggested that otherwise participants would look for patterns rather than attempt to communicate. The pair's final score was their highest number of points scored in succession. This criterion means that the players cannot succeed through the sheer quantity of rounds played;

they must instead find a way to communicate reliably and hence coordinate their behaviour with each other.

Crucially, and unlike previous experimental studies, the set-up of the ECG ensures that the problem of how to signal signalhood must be solved by the participants themselves. The space of possible signals is not defined; any combination of moves could be used. Neither are the roles of signaller and receiver. Finally, no communication channel is pre-defined either. It might be objected that there is only one possible channel and thus that the channel is in some sense pre-defined. However this misses the point that the communicative behaviours must be embodied and thus that the communication channel(s) must be created rather than found. If we define a number of possible candidate channels then the task becomes one in which the participants have to agree on which channel to use; as such, they need not signal signalhood but can instead simply observe which channel is being used by their partner. The task would then be little different to a number of previous studies (in particular Galantucci, 2005) but with additional channels. To properly investigate whether participants can signal signalhood, and if so what that might mean for the emergence of communication, we must do no more and no less than provide them with a world in which they can interact with each other. This means that the immediate questions that participants must address are "how could we communicate?" and "is s/he trying to communicate with me, or not?". These questions contrast with previous studies, in which the question is the subsequent "how is s/he trying to communicate with me?". The a priori knowledge that communication is possible and expected is more or less absent in the ECG, but present in the other studies. It is because of this the ECG is uniquely well suited to the study of how communicative intent is communicated.

As our results show (see below), once this is achieved then the creation of a complete system is straightforward. Therefore the fact that many pairs failed to communicate with each other at all shows that to co-opt one's movement for the purpose of communication is no trivial task. In general it is not at all obvious to participants that there is a communication channel; although anecdotal, it is instructive to note that the most common reaction to the instructions given to participants at the beginning of the experiment was that the task is impossible.

## 6.3 Methods and results

Following basic instructions participants were given a 3-minute familiarisation period in which to play the game. Further, clarifying instructions were then given and any queries addressed. Participants then played the game for 40 minutes uninterrupted. Instructions were given in writing, and at no point mentioned communication; participants were simply told that they had to 'coordinate' their behaviour with the other player (instruction sheets can be found in appendix A). Over the two conditions (described below), pairs played an average of 193.5 rounds in the 43 minutes available to them. At the end of each game subjects were asked about the communication systems they developed or attempted to develop, and these self-reports were checked against the game logs. In addition to a £6 payment for participation, a £20 prize was offered for each member of the top-performing pair. Participants were recruited from a student-employment website. They were randomly assigned into pairs and at no point did they meet their partner.

Successful pairs typically converged upon a system like that described in figure 6.5 (thanks are due to Simon Kirby, who created this and several of the other figures (6.6, 6.7 and 6.8) in this chapter), where there is one default colour that is chosen whenever possible, and when necessary (i.e. when the default colour is not available) particular movements are negotiated to refer to the remaining colours. This strategy is used in dialogue so that the players are able to agree on a destination colour. If, for example, player one has red and green quadrants only while player two has blue and green, then player one would travel directly to a red quadrant and pause. This pause allows player two to either also move to a red quadrant if they have one or, alternatively, to signal one of the other colours. Since player two does not have a red in this example they would signal, say, green. Player one has a green quadrant, and so travels there and finishes their turn. Player two then travels to the green square, finishes, and a point is scored.

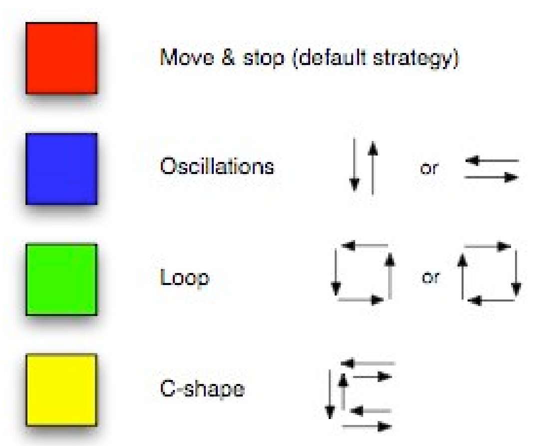


Figure 6.5: A typical emergent system in the ECG. In this communication system red is the default colour. If participants have a red square, they move to it and wait. If they do not have red they will signal one of the other colours by using the movements indicated. If one participant signals a colour that the other participant also has, that participant will move to the relevant square and hit space to end their turn. Otherwise, the participants will signal alternative colours until an agreement is reached.

Of particular interest is the way that such systems emerge. In debrief interviews participants report that such systems are not created fully-formed but in fact bootstrap off an initial strategy of always choosing the default colour if it is available. This strategy is not communicative, but it does allow pairs, once they have converged on the same default colour, to score at above chance levels. However, they are still very limited in the success they can achieve in this way, because sooner or later one or the other player will have a box with no red (or whatever the default colour is) quadrants, at which point the default colour strategy will fail to score. After this has occurred a number of times one of the players will, when faced with a box with no red quadrants, perform some behaviour that is otherwise unexpected of them. This will usually be oscillations along one side of the box, or a loop around the entire box; in short, it is something that differentiates it from the essentially aimless background wandering that typifies the behaviour of participants that have not yet worked out how they might succeed (exactly what this quality is will be discussed in greater length in §6.4 below). Participants report that this behaviour means “No red!”, “Not plan A!” or something similar. Once this signal is recognised the players then choose some other colour on which to finish. The two players may or may not choose the same colour, but once this scenario has arisen sufficiently often the players converge on some agreed colour to choose when the “No red!” signal is given. Then, once the “No red!” behaviour is consistently paired with this second colour, its meaning changes to, simply, “Blue” (or whatever the colour in question is). This entrenchment means that there is now a default colour and a symbol for a second colour in place, and participants consequently report that it was easy to negotiate on symbols for the remaining two colours. They are thus now able to score in every round of the game using dialogue like that described. Figure 6.6 reports the entire process, none of which is a post-hoc analysis; it is what the participants themselves describe in debriefing interviews after the event. Moreover, in all cases participants reported the same story as their partner in terms of which movements corresponded to which colours, and how those associations came about.


Figure 6.6: Stages in the development of successful communication systems. First, in (i), the participants converge upon some shared default colour, usually (in 4 of 5 cases) red. In (ii) one participant performs some movement that would be otherwise unexpected – typically oscillations or circles around the box. This is designed to tell the other participant that this participant does not have the default colour available. As a result different colours are chosen, and soon (iii) the two participants agree on a second-choice colour that they use when one or the other of them does not have the default colour. Then, in (iv), the movement used in (ii) comes to correspond, through repeated use, to the colour chosen in (iii). Finally, (v) now that such a symbol has been established the participants find it straightforward to agree on symbols for the remaining two colours. They consequently develop a system like that in figure 6.5 which enables them to score in every round and hence build a very high points-in-succession score, as shown in figure 6.7 below.

Initially, 24 participants were assigned into 12 pairs. Despite the fact that all participants are (inevitably) already fluent users of a learnt, symbolic communication system, and are immersed in a cultural environment that includes the frequent use of games for entertainment and education, five of the 12 pairs reported that they had failed to achieve any communication at all, while seven did report some success. The accuracy of these self-reports is reflected in the final scores: those that reported success scored 83, 66, 54, 49, 39, 17 and 14 while those that reported failure scored 7, 5, 4, 3 and 3. Pairs played an average of 206.7 rounds ($\sigma$=108.6), and the pairs that reported success all scored significantly above chance (in all cases $p<0.00001$ in a Monte Carlo simulation with 10,000 runs; other p-values obtained in the same way). Of the pairs that reported failure one scored significantly above chance ($p=0.047$). Their score of 7 was achieved with the default colour strategy, which the pair actually employed for a short period of time, but unlike several other pairs they were unable to build on the default colour strategy in the way described above. This can be seen in figure 6.7, which gives a graphical comparison of this pair with one of the successful pairs. None of the other pairs that reported failure scored significantly above chance levels (in all cases $p>0.188$), and correspondingly none reported having converged on any way to coordinate behaviour. Of the seven pairs that succeeded, five (final scores 83, 66, 49, 17, 14) built their system via the process described above. The two pairs that did not do this (final scores 54 and 39) tied the target colour either to a number of movements made from the starting position or to a number of oscillations. In both cases these systems were created as a *de novo* solution by one of the players and then detected by the other.

It seems, then, that the possibility of creating some initial convention (the default colour) is an aid to the emergence of communication. This hypothesis was tested with a second run of the experiment with one single change: whenever a point was scored then the colour on which it was scored would not be available to both players in the following round. This ensured that the default colour strategy would not achieve success even at chance levels, unless combined with a signalling strategy: any attempt to score on the same colour in two successive rounds was guaranteed to fail. The players were not made aware of this restriction. The prediction was that fewer pairs would be able to construct communication systems than did so under the original set-up. This is despite the fact that all of the communication systems observed in the previous condition would be perfectly adequate for this one as well; the change to the game's structure only affects the process of emergence, and not its use once established.

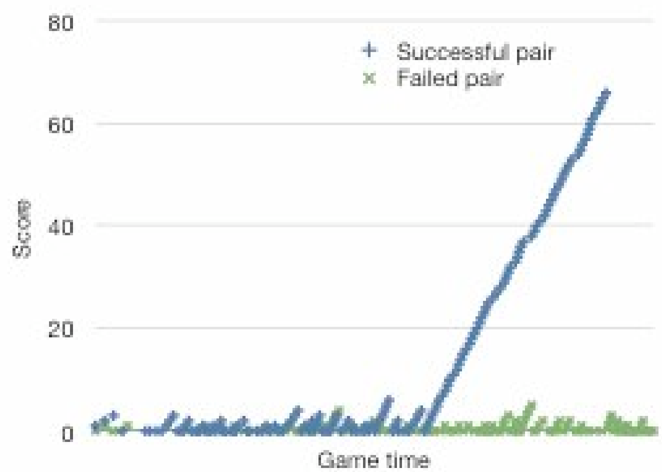


Figure 6.7: A comparison in the performance of two pairs of participants. On the x-axis is real time and on the y-axis the pairs' points-in-succession. One pair (green) failed to achieve any communication and finished with a chance level score of 7, while the other pair (blue) did create a communication system and finished with a highest points-in-succession score of 66. The gaps near the beginning of the time reflect the pause between the initial familiarisation period and the experiment proper.

The players in this condition played an average of 180.3 rounds ($\sigma$=111.3); this is not significantly different from the previous condition ($t_{22}$=0.5883, $p>.05$). Two of twelve pairs reported success (scores: 38 and 14; $p<0.00001$ in both cases), one tying the target colour either to a number of movements, the other using different shapes of movement. Ten pairs reported failure (scores: 6, 6, 6, 5, 5, 4, 3, 3, 3, 2; $p>0.12$ in all cases). These scores are significantly lower than those achieved when players did have access to the default colour strategy, both in terms of the number of pairs that achieved success

($\chi^2_1$=4.44, p<.05) and the average score of pairs in each condition ($t_{22}$=2.39, p<.05). Figure 6.8 compares the two conditions. It seems, then, that the possibility of creating some initial convention onto which communication may be bootstrapped significantly increases the likelihood that a symbolic communication system will emerge.

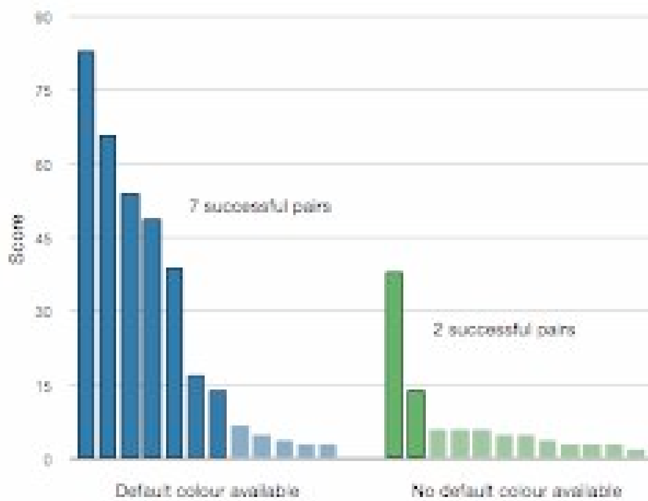


Figure 6.8: Comparison of performance when recourse to the default colour strategy was and was not available. Each bar refers to one pair, with their final score on the y-axis. The darker-coloured bars are those pairs that reported success; the lighter-coloured ones those that reported failure. The difference between the two conditions is significant both in terms of the number of pairs that achieved success ($\chi^2_1$=4.44, p<.05) and the average score achieved in each condition (t(22)=2.39, p<.05).

## 6.4 Discussion

In §6.1 we observed that, at least in the case studies considered, communication emerges via either cues or coerced behaviours. That is, design first appears on one side of the dyad and in doing so enables design to appear on the other. This process is impossible in the ECG, at least within any given round. To see why, try to imagine what an instance of coercion would look like. No such thing is possible, and the same is true of cues. It is true that across rounds cues could occur; one player could pursue the default colour strategy, or indeed any recognisable pattern of behaviour, for their own idiosyncratic reasons, and the other player might notice this and use it to increase success rates. However within rounds there is no possibility of a cue or a coerced behaviour. Yet despite this at least some pairs of participants manage to converge upon symbolic communication. It must therefore be the case that the protosignal is created from scratch and is yet still recognised as such. This suggests the human capacity to create and detect signalhood may enable us to find solutions to the problem of how to evolve communication (symbolic or otherwise) that are in large part denied to other organisms. Interestingly, this process leaves its mark on the final system in the form of the default colour. The idea that communication systems are shaped in important ways by their usage is well recognised; this is, for example, crucial to the process of grammaticalisation (Bybee et al., 1994). The same may be true of emergence: the final systems observed in the ECG, like that described in figure 6.5, do not obviously correspond to any solution that might be designed beforehand. It should be noted that this observation could not be made within previous, non-embodied experimental approaches to the emergence of communication. For this reason, future research in this area is most likely to be productive when the constraint of embodiment is met.

In many respects this works raises more questions than it answers, but that is to some degree inevitable (and perhaps even desirable) in the sort of exploratory work that an interdisciplinary field, lacking a unified approach, must do. The first of these questions is why the establishment of an initial convention should help bootstrap communication? Taking a lead from psycholinguistics, the obvious answer is that the default colour strategy appears to provide the common ground that makes what we can term protosignals (see below) more likely to be used and detected (Clark, 1996). Without this, the task is remarkably difficult, as the number of failures in our experiment testifies. It is perhaps noteworthy that four of the five pairs that bootstrapped communication onto non-communicative beginnings used red as their default colour. This apparent colour preference may be important, as it allows the players to converge upon some common ground without the need for communication. The initial stages of the ECG may be thought of as a Schelling Game (Schelling, 1960), where players must converge upon some shared solution (a Schelling Point) with no pre-existing knowledge of each other's intentions. For example, two players are asked to pick one of three objects – a basketball, a football, and a squash ball – and if they pick the same object then they win a prize. Common ground dramatically increases the likelihood that such problems will be successfully solved (Clark, 1996). If the two players happen to be squash partners, and if they both know that the other player is their squash partner, then the fact that they are squash partners supplies them with some common ground that they can use to converge upon the same solution: the squash ball. In general, the two players must rely on shared assumptions about each other's likely behaviour. In the ECG the human sensory preference for red (Fernandez & Morris, 2007; Teller & Bornstein, 1987) may provide just such common ground and allow them to converge on a default colour. Of course, if all the colours carried equal perceptual salience then the players would have to use pure deduction to converge on one of the colours to be their default, but this is an artefact of the constrained nature of the task. In a more open-ended environment individuals would have to rely on there being some Schelling Point that could be utilised to overcome the problem of a large (infinite?) set of candidate conventions, and ultimately provide the contextual environment necessary to allow the final communication system to be bootstrapped. There is, of course, some limited common ground provided by the very nature of the task: both participants know what their shared goal is, and that the other participant knows as much. Indeed it may be precisely because they have this common ground that some pairs are able to spontaneously create and infer viable systems. However in a more open-ended environment there would be a much larger (infinite?) set of possible meanings, which would likely rule out complete individual creation as a viable approach.

So, with sufficient common ground humans can communicate and detect communicative intent. We can term the behaviours that are intended to perform this task protosignals. They are not yet full signals, of course, since signals, by definition, already have an established and corresponding response; these behaviours, in contrast, are produced in the *hope* that an appropriate reaction will follow. An interesting question is: what qualities must a protosignal possess in order to be detected as such? The answer could conceivably be very little: the emergent behaviour-colour mappings could be nothing more than statistical abstractions, an observation that leads us towards terms like *irrational*: it is behaviour that can have no other purpose. Indeed, this seems quite plausible when the common ground provided by the default colour strategy is present: any deviation from direct travel to a quadrant suggests that something is amiss. However when that context is not available irrationality may be insufficient. Participants who have not yet figured out a way to play the ECG at above chance levels are not motionless, but instead wander around the boxes, for the entire 43 minutes, in what appears to be an aimless way. Certainly, they frequently return to squares they have previously visited in a given round, and so their objectives cannot be classified as travel; and there is no reason (including their own testimony) to think that the participants intend to use this 'background wandering' to communicate. We must therefore call this behaviour irrational, but it is quite different in character to the protosignals we are interested in which, in addition to a degree of redundancy, have a further feature that seems best captured by terms like *designed*, *functional* and *intentional*. Recent work with the ECG (Schillinger, 2008) supports this view: participants who watch videos of others playing the ECG sometimes detect that certain behaviours are communicative yet have no idea what those behaviours might mean; that is, they recognise communicative intent quite independently of informative intent. This means that there must be some important qualities possessed by the protosignal itself.

What might they be? Further research is necessary to answer that question. One possibility would be to manipulate participants' movements within the ECG such that some horizontal movements appear to the other player as vertical, and vice versa. This would not deny any information (in the formal sense that it would not reduce uncertainty) to the other participant, but done systematically it would enable the experimenter to study what qualities behaviours must possess before they are detected as being intentionally or functionally designed. It should be no surprise that such features are important to the way in which humans transcend the usual processes of emergence discussed in §6.1 above. After all, the capacity to detect design in the environment is precisely the quality that separates Gregorian creatures from the rest of the natural world, and in the last chapter we saw that these abilities play a crucial role in the shaping of our pragmatic behaviour. Given its importance there, it is surely no coincidence that they are also important in the detection of communicative intent.

This chapter began with the observation that the social evolutionary perspective on language has so far explained very few of the phenomena that constitute the topics of enquiry in linguistics, but it was promised that the theoretical observations made earlier in the thesis would be applied to that end. In many respects the development of the ECG would not have been possible without that foundation: both social evolutionary and, especially, pragmatic theory inform this research in quite fundamental ways. For example, the ECG's unique quality, the one that marks it out as different from previous approaches to the emergence of communication, is that it investigates the communication not just of informative intent but also of communicative intent. The distinction between communicative and informative intent, first articulated by Grice (1975a), is utterly central to pragmatics (Sperber & Wilson, 1986), and, as we saw in chapter 4, these two dimensions to communication have a close relationship with the social evolutionary problems of, respectively, reliability and honesty. A second example is that the importance of Schelling Points and common ground in the ECG, and indeed in the whole process of emergence, can be straight-forwardly explained in purely relevance-theoretic and hence, as discussed in the last chapter, social evolutionary terms. This is to be expected, or perhaps even demanded, given the arguments in the last chapter with respect to the evolutionary coherence of RT.

How can we interpret our results in relevance-theoretic terms? Common ground is more specifically defined as knowledge that we both have, and which we both know that we both have (Clark, 1996; Lewis, 1969; Schiffer, 1972; Sperber & Wilson, 1982). This account is clearly open to an infinite regress: how do either of us know that we both know that we both know something (Schiffer, 1972)? The development of RT was motivated in significant part by a reaction to this problem and dissatisfaction with the theories proposed to overcome it (Sperber & Wilson, 1982). The solution offered by RT involves a quality, *manifestness*, that is weaker than knowledge in the sense that it does not need to be known to be true, but for which there is sufficient evidence that it is likely and can be assumed to be true. More precisely, "a fact is *manifest* to an individual at a given time if and only if he is capable at that time of representing it mentally and accepting its representation as true or probably true" (Sperber & Wilson, 1986, p.39, italics in original). For example (from Sperber & Wilson, 1982), utterances that sound like English can be assumed to be English, even when it is not known whether the speaker does in fact speak English. The very fact that they have produced an utterance that sounds like English renders their ability to speak English manifest to those who have heard it (and are able to comprehend English), even though it may not be true, and thus cannot be known. The notion of relevance can then be interpreted as anything that makes some fact manifest or more manifest; that is, positive cognitive effects can be understood as those that increase manifestness. Indeed, in RT the idea of common ground is supplanted by the notion of a *mutual cognitive environment*, and this is defined in terms of manifestness.

How does all this apply to the ECG? Individuals perceive the different colours as having different degrees of salience. They can also reasonably expect others to do the same, and thus the mutuality of that experience is manifest to both participants. This makes it a suitable Schelling Point, and does so for the same reasons that, as discussed above, a squash ball is a suitable Schelling Point for a pair of participants that happen to be squash partners: mutual manifestness. A similar story applies to participants' use of the default colour strategy. Once this is in use then it is expected that the other player will, if they have a red quadrant, travel directly to that quadrant. Any failure to do so will thus automatically become manifest to the other player, and it is for this reason that irrationality may be a sufficient quality for signalhood once a context has been established. However protosignals are usually more than just irrational; they are irrationality writ large. That is, they do not just deviate from the expected rational behaviour, but they do so in over-specified ways, most obviously in their use of repetition. This increases their manifestness and thus, if they are consequently detected as protosignals, their relevance.

Now, why should we pay attention to this interpretation? It does not demonstrate any new facts, and seems to replace the intuitive ideas of common ground and salience with what seems to be somewhat baroque terminology: manifestness, cognitive environments, and so on. The answer to this question is that we wish to make clear that all of what is observed in the ECG is interpretable in ordinary relevance-theoretic terms. The social evolutionary perspective has led us towards RT as a theory that places Darwinian imperatives at its core, and the ECG was in part inspired by insights from the same perspecive. A coherent RT-based interpretation of the results is, then, the least we should expect. We may, if we wish, choose to acknowledge this but still stick to other, perhaps more intuitive explanatory frameworks, but this does not deny the coherence of the RT approach.

There are, then, two key points to take away from the work conducted so far on the ECG. The first is that it serves as an illustration of how the theoretical perspective developed in the rest of the thesis can be used to study aspects of language that are not obviously part of pragmatics. The second is the importance of the human capacity to detect design in the behaviour of others. In the last chapter we also saw that this ability, coupled with a social evolutionary analysis of communication, was sufficient to derive the core principles of an already well established theory of pragmatics. We have now seen, empirically, that it is also crucial to the emergence of learnt, symbolic communication, since it enables humans to detect signalhood. The natural next question to ask is how and why such an ability might have evolved.

## 6.5 Intentionality and the social brain

What cognitive capacities are necessary to succeed at the ECG, and to take part in ostensive-inferential communication in general? Such matters are, of course, also the explananda of psycholinguistics, which already makes fruitful use of experimental games to this end (e.g. Anderson et al., 1991). What, then, can the ECG add to this endeavour, and what can it tell us about the evolution of language? Recall that the key difference between the ECG and previous approaches to the emergence of symbolic communication is that in the ECG participants must solve not just the problem of what a signal might mean, but they must also recognise the existence of the signal in the first place. Recall also that this distinction corresponds to the pragmatic distinction between communicative and informative intent. To gloss once more, these are, respectively, my intention that my listener understand, or recognise, that I want to communicate with them at all; and my intention that what I wish to communicate is actually understood. Importantly, the latter is embedded in the former, so we can say that the speaker intends that the listener understands that the speaker intends that the listener understands X, where X refers to the propositional content of the stimulus (see also figure 1.2). The two clauses in italics are, respectively, the communicative and the informative intention.

This characterisation of ostensive-inferential communication contains four intentional verbs: the speaker intends$_1$ that the listener understand$_2$ that the speaker intends$_3$ that the listener understand$_4$. It has thus been suggested that the communication of communicative intent requires that speakers be sufficiently cognitively sophisticated that they can hold each of these intentional states together in some sort of metarepresentational hierarchy (Bennett, 1976; Dennett, 1987; Sperber, 2000). The various levels of this metarepresentational ability are usually referred to as levels of intentionality. 1$^{st}$-order intentionality refers to the awareness that oneself is able to think, 2$^{nd}$-order to the ability to think about what others are thinking, 3$^{rd}$-order to the ability to think about what others are thinking about what others are thinking, and so on and so on. The claim that such skills are necessary in communication is obviously predicated upon the assumption that the pragmatic gloss of the process of ostension must be cognitively realised in some more-or-less explicit way. This claim is immediately challenged by the simple observation that young children are able to communicate, both as speaker and listener, despite not being in possession of such skills. Indeed, even one-year-old infants' can differentiate between behaviours that are accidents and those which are communicatively intended, despite superficial similarities in form (Behne et al., 2005). However it may be the case that infants and children (and indeed adults most of the time) are able to avoid the need for explicit representation of communicative intentions because they instead make use of a great number of biologically basic signals, for example eye contact and body posture, that communicate such intentions without the need for a full cognitive realisation. That is, if somebody is looking directly at us then they to some degree or another prime us to interpret their consequent behaviour in communicative terms (e.g. Kleinke, 1986). Presumably animal communication works in some similar way: particular behaviours trigger receptive devices, and this allows communication to function without any metarepresentation of the process of communication. Simple heuristics like this are often used to implement adaptive functionality (Gigerenzer & Todd, 1999).

The ECG, however, abstracts away from all of this: because participants never meet, they cannot rely on any biologically basic signals for understanding when their partner wishes to communicate. It may consequently be the case that they must instead rely on a cognitive representation of the process of communication. If this is the case, then the communication of communicative intent within the ECG would require 4$^{th}$-order intentionality (the signaller must intend$_1$ that the receiver understand$_2$ that the signaller intends$_3$ that the receiver understands$_4$…) and comprehension of communicative intent 3$^{rd}$-order (the receiver must understand$_1$ that the signaller intends$_2$ that the receiver understands$_3$…). The plausibility of this hypothesis is demonstrated by a study that, using a relevance-theoretic framework, found a close relationship between the ability to comprehend similes, metaphors and irony and the ability to attribute mental states to others (Happé, 1993). Autism Spectrum Disorder (ASD) patients with only first order intentionality were able to comprehend the similes but not metaphors: they could respond appropriately to a sentence like 'he is like a rock' but not 'he is a rock'. It seems that the single word *like* gives the sentence a literal interpretation and therefore does not require the listener to attribute thoughts to another. This contrasts with the metaphor condition that, on an metarepresentational account, requires the listener to attribute a mental state to the speaker. Similarly, irony requires one further level of intentionality (because the speaker comments on some imagined mental state held by another; see Sperber & Wilson, 1986 for discussion). Hence, the listener requires three levels of intentionality in order to comprehend irony, and, using ASD patients with varying deficits in this regard, this is indeed what was found. The present hypothesis could be tested in a similar way or, probably with more control, with young children rather than ASD patients, comparing those who have recently attained the requisite orders of intentionality with those that have not.

The idea that communication is not possible in the absence of biological cues or signals to aid ostensive-inferential comprehension is a tantalising one; if correct it would suggest that the evolution of higher orders of intentionality had a significant role to play in the emergence of a shared system of learnt symbols. Moreover, the ability to represent other minds is not simply the product of increased general intelligence, but rather a form of specifically *social* intelligence: apes are the equal of 30-month old human infants in tests of general intelligence but not in tests of social intelligence (Herrmann et al., 2007). Correspondingly, the social brain hypothesis (Byrne & Whiten, 1989; Dunbar, 1998a; Dunbar & Shultz, 2007; Humphrey, 1976) argues that the evolution of such capacities was an adaptive response to an increase in the group size of *Homo sapiens*. Not only does an increase in the number of individuals that one has to interact with place increased pressure on one's ability to think about others' behaviour, it also increases, at an exponential rate, the number of interpersonal relationships within the group. It is consequently argued to be a major driving force behind the expansion in brain size (relative to body size) witnessed in the human lineage (Dunbar, 1992). Moreover, even if the increase in group size is not quite so directly implicated in the evolution of the capacity to engage in the creation of learnt symbols, it is still true that metarepresentation of other minds is crucial to the practice of conversational discourse. Without it the twin tasks of ostension and inference become more-or-less impossible. After all, how can you be relevant if you are not sensitive to what your interlocutor knows and does not know? Consequently, those that are unable to represent other minds are disarmingly literal-minded, "unable to participate in a conversation in any normal sense" (Baron-Cohen, 1988, p.83-84). On a similar note, "the inability of most animals to recognize the mental states of others distinguishes animal communication most clearly from human language" (Seyfarth & Cheney, 2003, p.145). This dependence of ostensive-inferential communication on the representation of other minds (termed theory of mind) means that even if the stronger hypothesis that $4^{th}$-order intentionality is required for ostensive-inferential communication is not supported, the evolution of social intelligence is still critical for the evolution of pragmatic fluency. Indeed, a recent hypothesis, well supported by the data, is that what is unique about human cognition is what has been termed shared intentionality, the capacity and motivation to participate with others in collaborative activities with shared goals and intentions (Tomasello et al., 2005). This is, unquestionably, an idea predicated on sociality. In fact, it reflects our 'ultra-sociality': the fact that, as was briefly mentioned in §2.1.3, we do not just have to survive in large social groups, but we must also take from the cultural environment into which we are born a whole set of artifacts, symbols and social practices, of which language is just the most obvious example. This cultural intelligence chimes quite nicely with the social emphasis of our discussion of the ECG, and suggests quite clearly that the phylogenetic story of language must include, at a minimum, a pivotal role for the emergence of the ability to represent other minds to at least a reasonable degree of sophistication (Dunbar, 1998b; Tomasello, 2008).

# Seven

# Where now for the social evolution of language?

The development of social evolution theory in the second half of the 20th-century (as personified by e.g. Dawkins, 1976; Maynard Smith, 1982; Mayr, 1963; Tinbergen, 1963; Williams, 1966; Wilson, 1975) provides a rich explanatory framework for the social behaviour of organisms, including communication. This thesis has asked what that body of work can tell us about language, and pragmatics in particular. Following the emphasis typically found in that literature, we first asked, in chapter 2, about the evolutionary function of language. Whilst it is commonly (and not unreasonably) assumed (following Pinker & Bloom, 1990) that the answer to that question is communication, others, in particular Chomsky, have argued otherwise. A verdict on that debate was delayed until a proper, biologically-grounded definition of communication was obtained, which was the goal of chapter 3. That definition allowed us to conclude that language is indeed an adaptation for the purposes of communication, so long as we accept that view that language is, ultimately, a matter of ostension and inference and not the symbol manipulation device that it is within the Chomksyan paradigm. We then turned, in chapter 4, to the question of evolutionary stability. It was observed, using insights from pragmatics, that there are in fact two (analogous) problems rather than one, at least when signals carry propositional content. The possible solutions to these problems were placed into a simple three-way classification (indices, handicaps, deterrents) and it was observed that of these three, deterrents are the most likely explanation in humans, via the mechanism of reputation. Chapter 5 then asked how the functionality of ostensive-inferential communication is implemented in humans. Using the basic definition of communication derived in chapter 3, a very general, functional level model of communication was constructed, and its basic tenets articulated. These tenets were then shown to predict with remarkable precision the central ideas of one particular post-Gricean approach to pragmatics, that of Relevance Theory (Sperber & Wilson, 1986); a mesh that gives the relevance-theoretic enterprise evolutionary support. Finally, in chapter 6, the theoretical apparatus constructed in the rest of the thesis was applied to the problem of the emergence of symbolic communication. The Embodied Communication Game was introduced, whose unique quality is that it demands the communication not just of informative intent, as previous work has done, but also of communicative intent. The experimental work presented suggests that how the latter problem is resolved fundamentally influences the form of the final communication system. Much further work using this approach can be imagined.

Even though we have addressed the questions of both function and mechanism, this does not mean that language has been explored in its entirety. On the contrary, by placing pragmatics at the centre of our conception of language we have ignored many aspects of language that many linguists, and not just Chomskyans, would consider integral (see, e.g., the list of design features in Hockett, 1960). Correspondingly, the emergence of something we would recognise as language is dependent on a great range of factors (Hurford, 1999, 2003), most of which are not discussed in this thesis: the appropriate vocal apparatus (Fitch, 2000); the computational intelligence to engage in the combinatorics required for syntax and phonology (Hauser et al., 2002); biases for one-to-one form-meaning mappings (Smith, 2004); and the formation of basic concepts (Hurford, 2007) are just a few of many possible examples. Consequently it is often argued that language is a "new machine constructed out of old parts" (Bates et al., 1991, p.35). In fact, researchers of quite different theoretical bents, who agree on little else of substance (e.g. Bates et al., 1991 and Hauser et al., 2002), do seem to agree that the most likely course of events was not one in which some magic bullet arrived and from which everything flowed, but rather one in which language emerges from the confluence of several independent evolutionary streams. There is, of course, a sense in which this 'old parts, new machine' hypothesis is trivially true: as we observed in §2.1.1 on exaptation, natural selection inevitably builds on the preadaptations (an unfortunate term, since it appears to grant foresight to the blind process of natural selection; however it is in common use, so I will stick with it) that have gone before. Wings, for example, did not spring from nowhere, and so could also be said to be a new machine born of old parts (feathers, limbs that extend away from the body at a suitable angle, etc.). The same is true of any and all biological phenomena. However there is also a sense in which the metaphor is more apt and less trivial when applied to language, for two reasons. First, the number of preadaptations is particularly large. Different researchers classify the various phenomena in different ways, but however we do this the list is a long one (Hurford, 1999, 2003). Second, their alliance is a particularly powerful one; the increase in net adaptive value that is marked by the onset of language is likely to have been quite dramatic.

Thus a natural next step might be to examine the evolutionary foundations of each of these prerequisites. It is here that the comparative method, by which we can draw inferences about evolution based on what we know about how a trait varies between species, can be a particularly powerful tool for evolutionary linguistics (Fitch, 2005; Hauser, 1996; Hauser et al., 2002; Tomasello, 2003b). In the meantime there are a number of potential projects that arise quite directly from this thesis. Most obviously, there are many directions in which the work with the ECG, presented in the last chapter, could be extended. We have already discussed how the question of how it could be used to address the question of whether the emergence of learnt symbolism might depend on a pre-existing social intelligence. Also mentioned, in passing, was the matter of what qualities a protosignal must possess if it is to be recognised as such, and how the ECG might be used to investigate that. Other avenues for future work with the ECG include: the manipulation of exactly what colours participants think the other player can see, so as to explore in more depth the role of common ground in the emergence of symbolic communication; increasing the number of colours used from four to eight, and simultaneously varying exactly which colours are used in different conditions, so as to ask if perceptual structure is exploited to produce linguistic structure when there are a large number of meanings that need to be expressed; and the creation of Iterated Learning Model (Kirby & Hurford, 2002) and/or multiplayer versions of the ECG, so as to better understand the relative contributions of individual innovation and cultural transmission. Like the work discussed above, the idea here, as with most experimental work on language origins (e.g. Galantucci, 2005; Kirby et al., 2008; Selten & Warglien, 2007), is not to directly replicate the evolutionary history of language, but rather to investigate the cognitive and cultural phenomena that underpin the emergence of shared symbolic communication systems in a systematic way. How participants choose to signal signalhood under these various conditions will provide us with important insights into how learnt symbolism can and most likely did emerge in the human lineage.

There are, then, a number of ways in which the exploratory work in the ECG can be taken forward. However that is not the only source of future research to arise from this thesis. Three other topics that merit further exploration are the use of reputation to enforce stability in human communication (see §4.4), the question of whether and how syntax might vary according to adaptive prediction (see §4.1.2), and an extension of the generalised model of communication (see §5.2). The first of these, whilst highly plausible and intuitive, requires empirical investigation. One obvious way in which this could be done would be in the use of the economic games that have been profitably used to study the effects of reputation in the evolution of cooperation in humans (e.g. Axelrod, 1995; Milinski et al., 2002), but with the independent variable as honesty in a communication game rather than cooperation in a prisoner's dilemma or some similar game. Investigation of whether and how humans might differ from other primates in this regard would also be useful. The second is to take an evolutionary psychological approach to language form. Thus far evolutionary psychologists have paid most attention to pitch (e.g. Puts et al., 2006), and there is some work on vocabulary size (Rosenburg & Tunney, 2008) and semantic creativity (Franks & Rigby, 2005). In particular, demonstration that syntactic form can vary along adaptive lines would be a worthwhile contribution to the formalism-functionalism debate within linguistics. The third would be to extend the model of communication developed in chapter 5. It would be worthwhile to explore whether it could be used to formalise the observations made in §6.1 about the emergence of communication – that cues or coerced behaviour tend to emerge first, as preadaptations that are then co-opted for communication, or perhaps it could be used to frame the problem of evolutionary stability in a general way. Also, it might be asked whether the observations made about communication that were used to argue in favour of RT might also be used to evaluate current theories of dialogue in psycholinguistics (e.g. Pickering & Garrod, 2004). On a tangential note, whilst that model has been used in this thesis to make predictions about human pragmatic behaviour, there is no reason it could not also be used to make predictions about the behaviour of other organisms. Articulation of what those predictions might be, how they could be tested and what explanatory value they might have would be a worthwhile future project. Looking to the particularly long-term, the development of a framework to analyse the pragmatic behaviour of non-humans, and primates in particular, would allow us to use the comparative method to investigate the evolutionary history of pragmatics.

It may be observed that many of these suggestions for future research are empirical in nature. It is in some respects frustrating that there is not more experimental work in this thesis. However, these suggestions have arisen only after a significant amount of theoretical analysis. What are the reasons for this, and was so much theoretical work really necessary? Those working in the hard sciences have (almost by definition) foundations that provide a strong intellectual context for their research: quantum theory and general relativity in physics, the periodic table in chemistry, Darwinism in biology. Such paradigms help to define and shape the questions that are asked, and consequently scientists in those fields can engage in the practice of normal science, in which such questions can be treated as puzzles to be solved (Kuhn, 1962). In contrast, scientists who work in disciplines that are yet to converge on an agreed paradigm (the social sciences, including both linguistics and psychology, are the obvious contemporary examples) will often work within quite different explanatory frameworks, a fact that sometimes makes it difficult to establish where there is and where there is not agreement. Furthermore, what even counts as explanation will often differ within a discipline, a problem that is magnified even further with interdisciplinary work, where cross-disciplinary misinterpretation is frustratingly common. Given all of this, it is very important that researchers in the evolution of language are clear about the exact nature of the problem their research addresses, and of the explanatory framework within which they operate. However, as was discussed in the opening remarks of this thesis, it is undeniable that just such problems have arisen in contemporary work on evolution of language, despite agreement that a Darwinian account of language is desirable. It is in part for these reasons that this thesis has been distinctly theory-driven. Rather than take some aspect of language, and with it the explanatory tools and frameworks from its home discipline, and seek to provide an evolutionary analysis, we have instead built only on the basic foundations of one branch of the Darwinian edifice. Correspondingly, the thesis has focused on pragmatics not because that topic was chosen a priori, but because that is the branch of linguistics that social evolution theory most immediately speaks to. This approach has almost inevitably meant that empirical work had to be delayed until the tools and terminology of the foundational discipline were articulated in ways that could be applied to the explanatory target, but once that was done then we were able to carry out some exploratory and promising experimental work that, in large part because it was born of sound theory, has the potential to be taken further.

None of which is to say that social evolution is the only explanatory framework that evolutionary linguistics should take seriously. Far from it: the biological sciences provide a number of different approaches, all of which should be employed. What should be cautioned against, however, is the sacrifice of Darwinian thinking at the altar of theories that have a weaker paradigmatic base (Kinsella, in press). For example, "if a current theory of language is truly incompatible with the neo-Darwinian theory of evolution, one could hardly blame someone for concluding that it is not the theory of evolution that must be questioned, but the theory of language" (Pinker & Bloom, 1990, p.708). This thesis has been constructed with that sort of sentiment very much in mind; witness the epigraph on the very first page. Having said that, social evolution has been far from the only source of inspiration. The Gricean distinction between communicative and informative intent, for example, has been influential throughout, especially in the discussion of the social evolutionary problem of evolutionary stability (see §4.2), and RT obviously informed the entirety of chapter 5, to pick just two instances. In fact, it is precisely the multitude of approaches and frameworks that one is exposed to in interdisciplinary study that makes the evolution of language such a fascinating topic, and an excellent place in which to develop one's intellectual worldview. Whilst mine remains framed in the powerful explanatory terms of social evolutionary theory, it is now richer and more pluralistic than before I began my postgraduate education. I therefore hope that this thesis has reflected that development, and done my schooling justice.

# References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., et al. (1991). The HCRC map task corpus. Language and Speech, 34(4), 351-366.

Andersson, M. B. (1994). Sexual Selection. Princeton, NJ: Princeton University Press.

Anolli, L., & Ciceri, R. (1997). The voice of deception: Vocal strategies of naive and able liars. Journal of Nonverbal Behaviour, 21, 259-284.

Atlas, J. D. (2005). Logic, Meaning and Conversation: Semantic Underdeterminacy, Implicature, and their Interface. Oxford: Oxford University Press.

Austin, J. L. (1955). How to Do Things with Words. Oxford: Oxford University Press.

Axelrod, R. (1995). The Evolution of Cooperation. New York: Basic Books.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. Science, 211, 1390-1396.

Bach, K. (2006). Impliciture vs. Explicature: What's the Difference? Paper presented at workshop on 'Explicit communication', in honour of Robyn Carston, June

Bach, K., & Harnish, R. M. (1987). Relevant questions. Behavioral and Brain Sciences, 10, 711-712.

Ball, P. (2001). The Self-made Tapestry: Pattern Formation in Nature. Oxford: Oxford University Press.

Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. Cognition, 107, 817-828.

Barclay, P., & Lalumière, M. L. (2006). Do people differentially remember cheaters? Human Nature, 17(1), 98-113.

Barkow, J. H., Cosmides, L., & Tooby, J. (1992). The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford: Oxford University Press.

Baron-Cohen, S. (1988). Without a theory of mind one cannot participate in a conversation. Cognition, 29, 83-84.

Bar-Hillel, Y. (1971). Out of the pragmatic wastebasket. Linguistic Inquiry, 2, 401-407.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), Approaches to the Evolution of

Language: Social and Cognitive Bases (pp. 405-426). Cambridge: Cambridge University Press.

Bates, E., & MacWhinney, B. (1990). Welcome to functionalism. Behavioral and Brain Sciences, 13(4), 727-728.

Bates, E., Thal, D., & Marchman, V. (1991). Symbols and syntax: A Darwinian approach to language development. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch & M. Studdert-Kennedy (Eds.), Biological and Behavioral Determinants of Language Development. Hillsdale, NJ: Lawrence Erlbaum.

Bateson, M., Nettle, D., & Roberts, G. (2007). Cues of being watched enhance cooperation in a real-world setting. Biology Letters, 2(3), 412-414.

Becker, A. L. (1991). Language and languaging. Language and Communication, 11(1-2), 33-35.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. Developmental Science, 8(6), 492-499.

Bennett, J. (1976). Linguistic Behavior. Cambridge: Cambridge University Press.

Benz, A., Jäger, G., & van Rooy, R. (Eds.). (2006). Game Theory and Pragmatics. Basingstoke: Palgrave MacMillian.

Berger, P. L., & Luckmann, T. (1966). The Social Construction of Reality: A Treatise in the Sociology of Knowledge. Harmondsworth: Penguin.

Bickerton, D. (1990). Language and Species. Chicago: University of Chicago Press.

Bickerton, D. (1991). Language origins and evolutionary plausibility. Language and Communication, 11(1-2), 37-39.

Bickerton, D. (1996). I chat, thereby I groom. Nature, 380(6572), 303.

Bickerton, D. (1998). Catastrophic evolution: The case for a single step from protolanguage to full human language. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), Approaches to the Evolution of Language: Social and Cognitive Bases (pp. 341-358). Cambridge: Cambridge University Press.

Bickerton, D. (2003). Symbol and structure. In M. H. Christiansen & S. Kirby (Eds.), Language Evolution (pp. 77-93). Oxford: Oxford University Press.

Bickerton, D. (2007). Language evolution: A brief guide for linguists. Lingua, 117(3), 510-526.

Billen, J., & Morgan, E. D. (1998). Pheromone communication in social insects: Source and secretions. In R. K. Vander Meer, M. D. Breed, K. Espelie & M. L. Winston (Eds.), Pheromone Communication in Social Insects: Ants, Wasps, Bees and Termites. Boulder, CO: Westview.

Blackmore, S. (1999). The Meme Machine. Oxford: Oxford University Press.

Blutner, R. (1998). Lexical pragmatics. Journal of Semantics, 15(2), 115-162.

Bond Jr., C. F., Omar, A., Mahmoud, A., & Bonser, R. N. (1990). Lie detection across cultures. Journal of Nonverbal Behavior, 14(3), 189-204.

Botha, R. P. (2003). Unravelling the Evolution of Language. Oxford: Elsevier.

Boyd, R., & Richardson, R. C. (1985). Culture and the Evolutionary Process. Chicago: University of Chicago Press.

Bradbury, J. W., & Vehrencamp, S. L. (1998). Principles of Animal Communication. Sunderland, MA: Sinauer Associates, Inc.

Bradbury, J. W., & Vehrencamp, S. L. (2000). Economic models of animal communication. Animal Behaviour, 59(2), 259-268.

Brighton, H. (2002). Compositional syntax from cultural transmission. Artificial Life, 8, 25-54.

Brighton, H., Smith, K., & Kirby, S. (2005). Language as evolutionary system. Physics of Life Reviews, 2, 177-226.

Brown, J. L. (1983). Cooperation - a biologist's dilemma. Advances in the Study of Behavior, 13, 1-37.

Buchler, J. (1955). Philosophical writings of CS Pierce. New York, NY: Cornell University Press.

Buller, D. (2005). Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature. Cambridge, MA: MIT Press.

Bullock, S. (1997). Evolutionary Simulation Models: On their Character, and Application to Problems Concerning the Evolution of Natural Signalling Systems. Unpublished DPhil. thesis, University of Sussex.

Burling, R. (2000). Comprehension, production and conventionalisation in the origins of language. In C. Knight, M. Studdert-Kennedy & J. R. Hurford (Eds.), The Evolutionary Emergence of Language (pp. 27-39). Cambridge: Cambridge University Press.

Burling, R. (2005). The Talking Ape: How Language Evolved. Oxford: Oxford University Press.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new pussycat? On talking to babies and animals. Science, 296, 1435.

Buss, D. M., & Dedden, L. A. (1990). Derogation of competitors. Journal of Social and Personal Relationships, 7, 395-422.

Bybee, J. (1985). Morphology: A Study in the Relation Between Meaning and Form. Amsterdam: John Benjamins.

Bybee, J., Perkins, R., & Pagliuca, W. (1994). The Evolution of Grammar. Chicago: University of Chicago Press.

Byrne, R., & Whiten, A. (1989). Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans. Oxford: Clarendon Press.

Calvin, W. H. (1987). The brain as a Darwin machine. Nature, 330, 33-34.

Calvin, W. H., & Bickerton, D. (2000). Lingua ex Machina: Reconciling Darwin and Chomsky with the Human Brain. Cambridge, Mass.: MIT Press.

Cangelosi, A., Smith, K., & Smith, A. D. M. (Eds.). (2006). The Evolution of Language: Proceedings of the 6th International Conference on the Evolution of Language. Singapore: World Scientific.

Carnap, R. (1958). Introduction to Symbolic Logic and its Applications. New York, NY: Dover.

Carroll, S. B. (2005). Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom. London: Norton.

Carstairs-McCarthy, A. (2007). Language evolution: What linguists can contribute. Lingua, 117(3), 503-509.

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

Carter, R. (2004). Language and Creativity: The Art of Common Talk. London: Routledge.

Chiappe, D., Brown, D., Dow, B., Koontz, J., Rodriguez, M., & McCulloch, K. (2004). Cheaters are looked at longer and remembered better than cooperators in social exchange situations. Evolutionary Psychology, 2, 108-120.

Chomsky, N. (1957). Syntactic Structures. The Hague: Mouton.

Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behaviour. Language, 35(1), 26-58.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

Chomsky, N. (1967). Preface to the 1967 reprint of A Review of B. F. Skinner's Verbal Behavior. In L. Jakobovits & M. S. Miron (Eds.), Readings in the Psychology of Language (pp. 142-143). Englewood Cliffs, NJ: Prentice-Hall.

Chomsky, N. (1972). Language and Mind. New York, NY: Harcourt.

Chomsky, N. (1975). Reflections on Language. New York, NY: Pantheon.

Chomsky, N. (1982). Noam Chomsky on the Generative Enterprise: A Discussion with Riny Junybregts and Henk van Riemsdijk. Dordrecht: Foris.

Chomsky, N. (1988). Language and Problems of Knowledge: The Managua lectures. Cambridge, MA: MIT Press.

Chomsky, N. (1991). Linguistics and cognitive sciences: Problems and mysteries. In A. Kasher (Ed.), The Chomskyan Turn. Oxford: Blackwell.

Chomsky, N. (1995). The Minimalist Program. Cambridge, MA: MIT Press.

Chomsky, N. (1996). Language and evolution. New York Review of Books, 41.

Chomsky, N. (2002). On Nature and Language. Cambridge University Press, Cambridge.

Chomsky, N., Huybregts, R., van Riemsdijk, H. C., Fukui, N., & Zushi, M. (2004). The Generative Enterprise Revisited. Berlin: Walter de Gruyter.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. Behavioral and Brain Sciences, 31, 489-558.

Christiansen, M. H., & Kirby, S. (2003a). Language evolution: Consensus and controversies. Trends in Cognitive Science, 7(7), 300-307.

Christiansen, M. H., & Kirby, S. (Eds.). (2003b). Language Evolution. Oxford: Oxford University Press.

Clark, A. (1987). The kludge in the machine. Mind and Language, 2(4), 277-300.

Clark, A. (1997). Being There: Putting Brain, Body and World Together Again. Cambridge, MA: MIT Press.

Clark, A. (2002). That special something: Dennett on the making of minds and selves. In A. Brook & D. Ross (Eds.), Daniel Dennett (pp. 187-205). Cambridge: Cambridge University Press.

Clark, A., & Chalmers, D. (1997). The extended mind. Analysis, 58, 7-19.

Clark, H. H. (1996). Using Language. Cambridge: Cambridge University Press.

Clark, S. R. L. (2000). The evolution of language: Truth and lies. Philosophy, 75, 401-421.

Clutton-Brock, T. H., & Albon, S. D. (1979). The roaring red deer and the evolution of honest advertising. Behaviour, 69, 145-170.

Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. Nature, 373, 209-216.

Corballis, M. C. (2003). From hand to mouth: The gestural origins of language. In M. H. Christiansen & S. Kirby (Eds.), Language Evolution (pp. 201-218). Oxford: Oxford University Press.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. Cognition, 31, 187-276.

Croft, W. (1995). Autonomy and functionalist linguistics. Language, 71(3), 490-532.

Croft, W. (2000). Explaining Language Change: An Evolutionary Approach. London: Longden.

Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. Acta Psychologica, 124, 60-78.

Darwin, C. (1889). The Expression of the Emotions in Man and Animals. London: John Murray.

Dawkins, M. S. (1995). Unravelling Animal Behaviour. Harlow: Longman.

Dawkins, R. (1976). The Selfish Gene. Oxford: Oxford University Press.

Dawkins, R. (1979). Twelve misunderstandings of kin selection. Zeitschrift für Tierpsychologie, 51, 184-200.

Dawkins, R. (1982). The Extended Phenotype. Oxford: Oxford University Press.

Dawkins, R. (1983). Universal Darwinism. In D. S. Bendall (Ed.), Evolution from Molecules to Men (pp. 403-425). Cambridge: Cambridge University Press.

Dawkins, R. (1986). The Blind Watchmaker. Harlow: Longman.

Dawkins, R. (2004). Extended phenotype - but not too extended. A reply to Laland, Turner and Jablonka. Biology and Philosophy, 19, 377-396.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation. In J. R. Krebs & N. B. Davies (Eds.), Behavioural Ecology (1st ed., pp. 282-309). Oxford: Blackwell.

Dawkins, R., & Krebs, J. R. (1979). Arms races between and within species. Proceedings of the Royal Society of London, B, 205, 489-511.

de Backer, C. J. S., Nelissen, M., & Fisher, M. L. (2007). Let's talk about sex: A study on the recall of gossip about potential mates and sexual rivals. Sex Roles, 56, 781-791.

de Boer, B. (2001). The Origins of Vowel Systems. Oxford: Oxford University Press.

de Ruiter, J. P., Noordzij, M. L., Newman-Norland, S., Hagoort, P., & Toni, I. (2007). On the origin of intentions. In P. Haggard, Y. Rossetti & M. Kawato (Eds.), Attention & Performance XXII (pp. 593-610). Oxford: Oxford University Press.

de Saussure, F. (1959). Course in General Linguistics. New York, NY: McGraw-Hill.

Deacon, T. (1997). The Symbolic Species. London: Penguin.

Dennett, D. C. (1987). The Intentional Stance. Cambridge, Mass.: MIT Press.

Dennett, D. C. (1995). Darwin's Dangerous Idea. London: Penguin.

Dennett, D. C. (1996). Kinds of Minds. New York, NY: Basic books.

Dennett, D. C. (2003). Freedom Evolves. London: Penguin.

Dessalles, J.-L. (1998). Altruism, status and the origin of relevance. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), Approaches to the Evolution of Language: Social and Cognitive Bases (pp. 130-147). Cambridge: Cambridge University Press.

Dessalles, J.-L. (2007). Why We Talk: The Evolutionary Origins of Language. Oxford: Oxford University Press.

di Paolo, E. A. (1997). An investigation into the evolution of communication. Adaptive Behavior, 6(2), 285-324.

Dickins, T. E., & Dickins, B. J. A. (2007a). Designed calibration: Naturally selected flexibility, not non-genetic inheritance. Behavioral and Brain Sciences, 30(4), 368-369.

Dickins, T. E., & Dickins, B. J. A. (2007b). Mother nature's tolerant ways: Why non-genetic inheritance has nothing to do with evolution. New Ideas in Psychology, 26(1), 41-54.

Diggle, S. P., Gardner, A., West, S. A., & Griffin, A. S. (2007). Evolutionary theory of bacterial quorum sensing: When is a signal not a signal? Philosophical Transactions of the Royal Society, B, 362(1483), 1241-1249.

Diggle, S. P., Griffin, A. S., Campbell, G. S., & West, S. A. (2007). Cooperation and conflict in quorum sensing bacterial populations. Nature, 450, 411-414.

Dowty, D. (1990). A Semantic Account of Sequence of Tense. Paper presented at the workshop 'Time in Language', March

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. Journal of Human Evolution, 20, 469-493.

Dunbar, R. I. M. (1993). Co-evolution of neocortex size, group size and language in humans. Behavioral and Brain Sciences, 16(4), 681-735.

Dunbar, R. I. M. (1997). Grooming, Gossip, and the Evolution of Language. London: Faber.

Dunbar, R. I. M. (1998a). The social brain hypothesis. Evolutionary Anthropology, 6, 178-190.

Dunbar, R. I. M. (1998b). Theory of mind and the evolution of language. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), Approaches to the evolution of language: Social and cognitive bases (pp. 92-110). Cambridge: Cambridge University Press.

Dunbar, R. I. M. (2004). The human story. London: Faber and Faber.

Dunbar, R. I. M., Duncan, N. D. C., & Marriott, A. (1997). Human conversational behaviour. Human Nature, 8, 231-246.

Dunbar, R. I. M., Duncan, N. D. C., & Nettle, D. (1995). Size and structure of freely forming conversational groups. Human Nature, 6, 67-78.

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. Science, 317, 1344-1347.

Durrant, R., & Haig, B. D. (2001). How to pursue the adaptationist program in psychology. Philosophical Psychology, 14(4), 357-380.

Dusenbery, D. B. (1992). Sensory Ecology: How Organisms Acquire and Respond to Information. New York, NY: WH Freeman.

Edwardes, M. (2007). The Nature of Grammar, its Role in Language and its Evolutionary Origins. Unpublished PhD. thesis, University of East London.

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6, 169-200.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wlebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. Nature, 418, 869-872.

Enquist, M., & Leimar, O. (1993). The evolution of cooperation in mobile organisms. Animal Behaviour, 45(4), 747-757.

Evans, J. S. B. T. (1982). The Psychology of Deductive Reasoning. London: Routledge.

Fay, N., Garrod, S., MacLeod, T., Lee, J., & Oberlander, J. (2004). Design, adaptation and convention: The emergence of higher order graphical representations. In Proceedings of the 26th Annual Conference of the Cognitive Science Society (pp. 411-416).

Fehr, E. (2004). Don't lose your reputation. Nature, 432, 449-450.

Fernandez, A. A., & Morris, M. R. (2007). Sexual selection and trichromatic color vision in primates: Statistical support for the pre-existing bias hypothesis. American Naturalist, 170(1), 10-20.

Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. Cognition, 77, 1-79.

Fitch, W. T. (2000). The evolution of speech: A comparative review. Trends in Cognitive Science, 4(7), 258-267.

Fitch, W. T. (2004). Evolving honest communication systems: Kin selection and mother tongues. In D. K. Oller & U. Griebel (Eds.), The Evolution of Communication Systems: A Comparative Approach (pp. 275-296). Cambridge, MA: MIT Press.

Fitch, W. T. (2005). The evolution of language: a comparative review. Biology and Philosophy, 20, 193-230.

Fitch, W. T., Chomsky, N., & Hauser, M. D. (2005). The evolution of the language faculty: Clarification and implications. Cognition, 97, 179-210.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. Proceedings of the Royal Society of London, B, 268, 1669-1675.

Fodor, J. A. (1975). The Language of Thought. Cambridge, MA: MIT Press.

Franks, B., & Rigby, K. (2005). Deception and mate selection: Some implications for relevance and the evolution of language. In M. Tallerman (Ed.), Language Origins: Perspectives on Evolution (pp. 208-229). Oxford: Oxford University Press.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. Cognitive Science, 29, 737-767.

Galef, B. G., & Giraldeau, L.-A. (2001). Social influence on foraging in vertebrates: Causal mechanisms and adaptive functions. Animal Behaviour, 61, 3-15.

Gallistel, C. R. (1999). The replacement of general-purpose learning models with adaptively specialized learning modules. In M. S. Gazzaniga (Ed.), The Cognitive Neurosciences (pp. 1179-1191). Cambridge, MA: MIT Press.

Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. Science, 165, 664-672.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? Cognitive Science, 31(6), 961-987.

Gazdar, G., & Good, D. (1982). On a notion of relevance: Comments on Sperber and Wilson's paper. In N. Smith (Ed.), Mutual Knowledge (pp. 88-100). London: Academic Press.

Gigerenzer, G., & Todd, P. M. (Eds.). (1999). Simple Heuristics that Make us Smart. Oxford: Oxford University Press.

Gintis, H. (2000). Strong reciprocity and human sociality. Journal of Theoretical Biology, 206, 169-179.

Gintis, H., Alden Smith, E., & Bowles, S. (2001). Costly signaling and cooperation. Journal of Theoretical Biology, 213, 103-119.

Givón, T. (1979). From discourse to syntax: Grammar as a processing strategy. In T. Givón (Ed.), Syntax and Semantics: Discourse and Syntax (pp. 81-112). New York, NY: Academic Press.

Givón, T., & Malle, B. F. (2002). Introduction. In T. Givón & B. F. Malle (Eds.), The Evolution of Language out of Pre-language (pp. vii-ix). Amsterdam: John Benjamins.

Godfrey-Smith, P. (1993). Spencerian Explanation and Constructivism. Paper presented at MIT philosophy colloquium.

Gould, S. J. (1989). Tires to sandals. Natural History, 8-15.

Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. Journal of Social Issues, 47(3), 43-65.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. Proceedings of the Royal Society of London, B, 205, 581-598.

Gould, S. J., & Vrba, E. S. (1982). Exaptation; a missing term in the science of form. Paleobiology 8(1), 4-15.

Gouzoules, H., & Gouzoules, S. (2002). Primate communication: By nature honest, or by experience wise? International Journal of Primatology, 23(4), 821-847.

Grafen, A. (1990). Biological signals as handicaps. Journal of Theoretical Biology, 144, 517-546.

Grafen, A. (1999). Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging. Proceedings of the Royal Society of London, B, 266, 799-803.

Gray, R. D., Heaney, M., & Fairhall, S. (2003). Evolutionary Psychology and the challenge of adaptive explanation. In K. Sterelny & J. Fitness (Eds.), From Mating to Mentality: Evaluating Evolutionary Psychology (pp. 247-268). London: Psychology Press.

Greengross, G., & Miller, G. F. (2008). Dissing oneself versus dissing rivals: Effects of status, personality, and sex on the short-term and long-term attractiveness of self-deprecating and other-deprecating humor. Evolutionary Psychology, 6(3), 393-408.

Gregory, R. L. (1981). Mind in Science: A History of Explanations in Psychology and Physics. Cambridge: Cambridge University Press.

Grice, H. P. (1971). Meaning. In D. Steinburg & L. Jakobovits (Eds.), Semantics: An Interdisciplinary Reader (pp. 53-59). Cambridge: Cambridge University Press.

Grice, H. P. (1975a). Logic and conversation. In P. Cole & J. Morgan (Eds.), Syntax and Semantics III: Speech Acts (pp. 41-58). New York, NY: Academic Press.

Grice, H. P. (1975b). Studies in the Way of Words. Cambridge, MA: Harvard University Press.

Griffin, A. S., & West, S. A. (2002). Kin selection: Fact and fiction. Trends in Ecology and Evolution 17(1), 15-21.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. Cognitive Science, 31, 441-480.

Griggs, R., & Cox, J. (1982). The elusive thematic-materials effect in Wason's selection task. British Journal of Psychology, 73, 407-420.

Griskevicius, V., Cialdini, R. B., & Kenrick, D. T. (2006). Peacocks, Picasso, and parental investment: The effects of romantic motives on creativity. Journal of Personality and Social Psychology, 91(1), 63-76.

Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. Animal Behaviour, 42, 1-14.

Guinet, C. (1992). Predation behaviour of Killer Whales (Orcinus orca) around Crozet islands. Canadian Journal of Zoology, 70, 1656-1667.

Hailman, J. P. (1977). Optical Signals: Animal Communication and Light. Bloomington, IN: Indiana University Press.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evolution and Human Behavior, 26(3), 245-256.

Halliday, M. A. K. (1985). An Introduction to Functional Grammar. London: Edward Arnold.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. Journal of Theoretical Biology, 7, 1-52.

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of Relevance Theory. Cognition, 48, 101-119.

Hardin, G. (1968). The tragedy of the commons. Science, 162, 1243-1248.

Hasson, O. (1994). Cheating signals. Journal of Theoretical Biology, 167, 223-238.

Hauser, M. D. (1996). The Evolution of Communication. Cambridge, MA: MIT Press.

Hauser, M. D., Barner, D., & O'Donnell, T. (2007). Evolutionary linguistics: A new look at an old landscape. Language Learning and Development, 3(2), 101-132.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? Science, 298, 1569-1579.

Haven-Wiley, R. (1983). The evolution of communication: Information and manipulation. In T. R. Halliday & P. J. B. Slater (Eds.), Animal Behaviour II: Communication (pp. 156-189). Oxford: Blackwell.

Haviland, J. B. (1977). Gossip, Reputation and Knowledge in Zinacantan. Chicago, IL: University of Chicago Press.

Healey, P. I. U., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. Cognitive Science, 31, 285-309.

Heine, B. (1997). Cognitive Foundations of Grammar. Oxford: Oxford University Press.

Heine, B., & Kuteva, T. (2002). World Lexicon of Grammaticalization. Cambridge: Cambridge University Press.

Herrmann, E., Call, J., Hernández-Lloreda, M.-H., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. Science, 317, 1360-1366.

Hess, N. H., & Hagen, E. H. (2006). Psychological adaptations for assessing gossip veracity. Human Nature, 17(3), 337-354.

Hockett, C. (1960). The origin of speech. Scientific American, 203, 88-111.

Hoefler, S. H. (2008). The role of Pragmatic Plasticity in the Evolution of Linguistic Communication. Unpublished PhD. thesis, University of Edinburgh.

Hoefler, S. H., & Smith, A. D. M. (2008). Reanalysis vs. metaphor: What grammaticalisation can tell us about language evolution. In A. D. M. Smith, K. Smith & R. Ferrer i Cancho (Eds.), The Evolution of Language: Proceedings of the 7th International Conference (pp. 163-170). Singapore: World Scientific.

Hopper, P. J., & Traugott, E. C. (1993). Grammaticalization. Cambridge: Cambridge University Press.

Horn, L. R. (1984). Towards a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), Meaning, Form, and its Use in Context: Linguistic Applications (pp. 11-42). Washington, DC: Georgetown University Press.

Huang, Y. (2007). Pragmatics. Oxford: Oxford University Press.

Hull, D. L. (1988). Science as a Process: an Evolutionary Account of the Social and Conceptual Development of Science. Chicago, IL: University of Chicago Press.

Humphrey, N. (1976). The social function of the intellect. In P. Bateson & R. Hinde (Eds.), Growing Points in Ecology (pp. 303-317). Cambridge: Cambridge University Press.

Hurd, P. L. (1995). Communication in discrete action-response games. Journal of Theoretical Biology, 174, 217-222.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. Lingua, 77, 187-222.

Hurford, J. R. (1991). The evolution of the critical period for language acquisition. Cognition, 40, 159-201.

Hurford, J. R. (1999). The evolution of language and languages. In R. I. M. Dunbar, C. Knight & C. Power (Eds.), The Evolution of Culture (pp. 173-193). Edinburgh: Edinburgh University Press.

Hurford, J. R. (2000). Social transmission favours linguistic generalisation. In C. Knight, M. Studdert-Kennedy & J. R. Hurford (Eds.), The Evolutionary Emergence of Language (pp. 324-352). Cambridge: Cambridge University Press.

Hurford, J. R. (2003). The language mosaic and its evolution. In M. H. Christiansen & S. Kirby (Eds.), Language Evolution (pp. 38-57). Oxford: Oxford University Press.

Hurford, J. R. (2007). Origins of Meaning. Oxford: Oxford University Press.

Hurford, J. R. (2008). The evolution of human communication and language. In P. D'Ettorre & D. Hughes (Eds.), Sociobiology of Communication: An Interdisciplanry Perspective (pp. 249-264). Oxford: Oxford University Press.

Hurford, J. R. (in press). Universals and the diachronic life cycle of languages. In M. Christiansen, C. Collins & S. Edelman, Language Universals (pp. 40 – 53). Cambridge, MA: MIT Press.

Hurford, J. R., Studdert-Kennedy, M., & Knight, C. (Eds.). (1998). Approaches to the Evolution of Language: Social and Cognitive Bases. Cambridge: Cambridge University Press.

Huxley, J. (1966). Ritualisation of behaviour in animals and men. Philosophical Transactions of the Royal Society of London, B, 251, 249-271.

Hymes, D. (1991). Is Poetics Original and Functional? Language and Communication, 11(1-2), 49-51.

Jablonka, E., & Lamb, M. J. (2005). Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life. Cambridge, MA: MIT Press.

Jablonka, E., & Lamb, M. J. (2007). Précis of Evolution in Four Dimensions. Behavioral and Brain Sciences, 30, 353-392.

Kauffman, S. A. (1993). The Origins of Order: Self-Organisation and Selection in Evolution. Oxford: Oxford University Press.

Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In M. Barlow & S. Kemmer (Eds.), Usage-based Models of Language. Stanford, CA: CSLI Publications.

Kimura, D. (1993). Neuromotor Mechanisms in Human Communication. Oxford: Oxford University Press.

Kinsella, A. R. (in press). Language Evolution and Syntactic Theory. Cambridge: Cambridge University Press.

Kirby, S. (1998). Fitness and the selective adaptation of language. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), Approaches to the Evolution of Language: Social and Cognitive Bases (pp. 359-383). Cambridge: Cambridge University Press.

Kirby, S. (1999). Function, Selection and Innateness: The Emergence of Language Universals. Oxford: Oxford University Press.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, J. R. Hurford & M. Studdert-Kennedy (Eds.), The Evolutionary Emergence of Language (pp. 303-323). Cambridge: Cambridge University Press.

Kirby, S. (2002). Natural language from artificial life. Artificial Life, 8(2), 185-215.

Kirby, S. (2007). The evolution of language. In R. I. M. Dunbar & L. Barrett (Eds.), Oxford Handbook of Evolutionary Psychology (pp. 669-681). Oxford: Oxford University Press.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. Proceedings of the National Academy of Sciences, 105(31), 10681-10686.

Kirby, S., Dowman, M., & Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12), 5241-5245.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the Iterated Learning Model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language* (pp. 121-148). London: Springer Verlag.

Kirby, S., Smith, K., & Brighton, H. (2004). From UG to universals: Linguistic adaptation through iterated learning. *Studies in Language*, 28(3), 587-607.

Kitamura, C., Thanavishuth, C., Burnhama, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behaviour and Development*, 24, 372-392.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, 100, 78-100.

Knight, C. (1998). Ritual/speech coevolution: a solution to the problem of deception. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases* (pp. 68-91). Cambridge: Cambridge University Press.

Knight, C. (2004). Decoding Chomsky. *European Review*, 12(4), 581-603.

Knight, C., Studdert-Kennedy, M., & Hurford, J. R. (Eds.). (2000). *The Evolutionary Emergence of Language*. Cambridge: Cambridge University Press.

Kopytko, R. (1995). Against rationalistic pragmatics. *Journal of Pragmatics*, 23, 475-491.

Krebs, J. R., & Davies, N. B. (1993). *An Introduction to Behavioural Ecology*. Oxford: Blackwell.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (2nd ed., pp. 380-402). Oxford: Blackwell.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Lachmann, M., Számadó, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. Proceedings of the National Academy of Sciences, 98(23), 13189-13194.

Lakoff, G. (1991). Cognitive versus generative linguistics: How commitments influence results. Language and Communication, 11(1-2), 53-62.

Laland, K. N., & Sterelny, K. (2006). Seven reasons (not) to neglect niche construction. Evolution, 60(9), 1751-1762.

Lenneberg, E. H. (1967). Biological Foundations of Language. New York, NY: Wiley.

Leslie, A. (1989). Review of Relevance. Mind and Language, 4(1-2), 147-150.

Leslie, J. C. (2006). Herbert Spencer's contributions to behavior analysis: A retrospective review of Principles of Psychology. Journal of the Experimental Analysis of Behavior, 86(1), 123-129.

Levinson, S. C. (1983). Pragmatics. Cambridge: Cambridge University Press.

Levinson, S. C. (1989). A review of Relevance. Journal of Linguistics, 25, 455-472.

Lewis, D. (1969). Convention. Cambridge, MA: Harvard University Press.

Lewis, D. B., & Gower, D. M. (1980). The Biology of Communication. London: Blackie.

Lieberman, P. (1991). Preadaptation, natural selection and function. Language and Communication, 11(1-2), 63-65.

Lieberman, P. (2003). Motol control, speech, and the evolution of human language. In M. H. Christiansen & S. Kirby (Eds.), Language Evolution (pp. 255-271). Oxford: Oxford University Press.

Lightfoot, D. (1991). Subjacency and sex. Language and Communication, 11, 67-69.

Lock, A. (1978). The emergence of language. In A. Lock (Ed.), Action, Gesture and Symbol: The Emergence of Language (pp. 3-18). New York, NY: Academic Press.

Locke, J. L. (2001). Rank and relationships in the evolution of spoken language. Journal of the Royal Anthropological Institute, 7, 37-50.

Longa, V. M. (2001). Sciences of complexity and language origins: An alternative to natural selection. Journal of Literary Semantics, 30(1), 1-17.

Lorenz, K. (1939). Vergleichende Verhaltenforschung. Zoologische Anzeiger, 12, 69-102.

Lorenz, K. (1965). Evolution and Modification of Behaviour. Chicago, IL: University of Chicago Press.

Loyau, A., Petrie, M., Saint Jalme, M., & Sorci, G. (2008). Do peahens not prefer peacocks with more elaborate trains? Animal Behaviour, 76, e5-e9.

Mameli, M. (2004). Nongenetic selection and nongenetic inheritance. British Journal for the Philosophy of Science, 55(1), 35-71.

Marcus, G. (2008). Kluge: The Haphazard Construction of the Human Mind. Boston, MA: Houghton-Mifflin.

Mather, K. (1953). Genetic control of stability in development. Heredity, 7, 297-336.

Maynard Smith, J. (1956). Fertility, mating behaviour and sexual selection in Drosophila subobscura. Journal of Genetics, 54, 261-279.

Maynard Smith, J. (1976). Sexual selection and the handicap principle. Journal of Theoretical Biology, 57, 239-242.

Maynard Smith, J. (1978). Optimization theory in evolution. Annual Review of Ecological Systems, 9, 31-56.

Maynard Smith, J. (1982). Evolution and the Theory of Games. Cambridge: Cambridge University Press.

Maynard Smith, J. (1991). Honest signaling – the Philip Sydney game. Animal Behaviour 42, 1034-1035

Maynard Smith, J. (1995). Genes, memes and minds. New York Review of Books, 46-48.

Maynard Smith, J., & Harper, D. G. C. (1995). Animal signals: Models and terminology. Journal of Theoretical Biology, 177, 305-311.

Maynard Smith, J., & Harper, D. G. C. (2003). Animal Signals. Oxford: Oxford University Press.

Maynard Smith, J., & Szathmáry, E. (1995). The Major Transitions in Evolution. Oxford: W. H. Freeman.

Mayr, E. (1963). Animal Species and Evolution. Cambridge, MA: Harvard University Press.

Mayr, E. (1982). The Growth of Biological Thought. Cambridge, MA: Harvard University Press.

McAndrew, F. T., Bell, E. K., & Maria Garcia, C. (2007). Who do we tell and whom do we tell on? Gossip as a strategy for status enhancement. Journal of Applied Social Psychology, 37(7), 1562-1577.

McAndrew, F. T., & Milenkovic, M. A. (2002). Of tabloids and family secrets: The evolutionary psychology of gossip. Journal of Applied Social Psychology, 32, 1064-1082.

McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), Machine Intelligence IV. Edinburgh: Edinburgh University Press.

McComb, K., & Semple, S. (2005). Coevolution of vocal communication and sociality in primates. Biology Letters, 1, 381-385.

McGeer. (1990). Passive dynamic walking. The International Journal of Robotics Research, 9(2), 62-82.

McGregor, P. K. (2005). Communication. In J. J. Bolhuis & L.-A. Giraldeau (Eds.), The Behavior of Animals: Mechanisms, Function and Evolution (pp. 226-250). Oxford: Blackwell.

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. Ethology and Sociobiology, 17, 119-128.

Mehler, J. (1985). Review of P. Lieberman's 'The biology and evolution of language'. Journal of the Acoustic Society of America, 80, 1558-1560.

Milinski, M., Semmann, D., & Krambeck, H.-J. (2002). Reputation helps solve the 'tragedy of the commons'. Nature, 415, 424-426.

Miller, G. F. (2000). The Mating Mind. London: BCA.

Møller, A. P. (1993). Patterns of fluctuating asymmetry in sexual ornaments predict female choice. Journal of Evolutionary Biology, 6, 481-491.

Newmeyer, F. J. (1991). Functional explanation in linguistics and the origins of language. Language and Communication, 11, 3-28.

Newmeyer, F. J. (1998). Language Form and Language Function. Cambridge, MA: MIT Press.

Newmeyer, F. J. (2003). What can the field of linguistics tell us about the origin of language? In M. H. Christiansen & S. Kirby (Eds.), Language Evolution (pp. 58-76). Oxford: Oxford University Press.

Noble, J. (1998). The Evolution of Animal Communication Systems: Questions of Function Examined through Simulation. Unpublished PhD. thesis, University of Sussex.

Noble, J. (2000). Cooperation, competition and the evolution of prelinguistic communication. In C. Knight, M. Studdert-Kennedy & J. R. Hurford (Eds.), The Evolutionary Emergence of Language (pp. 40-61). Cambridge: Cambridge University Press.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. Proceedings of the National Academy of Sciences, 96, 8028-8033.

Nowak, M. A., & Komarova, N. L. (2001). Towards an evolutionary theory of language. Trends in Cognitive Science, 5(7), 288-295.

Nowak, M. A., Plotkin, J. B., & Jansen, V. A. (2000). The evolution of syntactic communication. Nature, 404, 495-498.

Nowak, M. A., Plotkin, J. B., & Krakauer, D. C. (1999). The evolutionary language game. Journal of Theoretical Biology, 200, 147-162.

Oda, R. (1997). Biased face recognition in the prisoner's dilemma games. Evolution and Human Behavior, 18, 309-315.

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). Niche Construction: The Neglected Process in Evolution. Princeton, NJ: Princeton University Press.

Ogden, C. K., & Richards, I. A. (1923). The Meaning of Meaning. New York, NY: Harcourt Brace.

Oliphant, M. (1997). Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language. Unpublished PhD. thesis, University of California, San Diego.

Oliphant, M. (2002). Learned systems of arbitrary reference: the foundation of human linguistic uniqueness. In E. Briscoe (Ed.), Linguistic Evolution through Language

Acquisition: Formal and Computational Models (pp. 23-52). Cambridge: Cambridge University Press.

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

Oudeyer, P.-Y. (2006). Self-Organization in the Evolution of Speech. Oxford: Oxford University Press.

Panksepp, J. (1998). Affective Neuroscience: The Foundations of Human and Animal Emotions. Oxford: Oxford University Press.

Parker, G. A. (1974). Assessment strategy and the evolution of animal conflicts. Journal of Theoretical Biology, 47, 223-243.

Pentland, A. (2008). Honest Signals: How they Shape our World. Cambridge, MA: MIT Press.

Piatelli-Palmarini, M. (1989). Evolution, selection and cognition: From 'learning' to parameter setting in biology and the study of language. Cognition, 31, 1-44.

Piatelli-Palmarini, M. (1990). An idealogical battle over modals and quantifiers. Behavioral and Brain Sciences, 13(4), 752-754.

Piazza, J., & Bering, J. M. (2008). Concerns about reputation via gossip promote generous allocations in an economic game. Evolution and Human Behavior, 29, 172-178.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences, 27, 1-22.

Pinker, S. (1994). The Language Instinct. London: Penguin.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M. H. Christiansen & S. Kirby (Eds.), Language Evolution (pp. 16-37). Oxford: Oxford University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. Behavioral and Brain Sciences, 13(4), 707-727.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? Cognition, 95, 201-236.

Pinker, S., Nowak, M., & Lee, J. J. (2008). The logic of indirect speech. Proceedings of the National Academy of Sciences 105(3), 833-838.

Power, C. (1998). Old wives' tales: The gossip hypothesis and the reliability of cheap signals. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), Approaches to the Evolution of Language: Social and Cognitive Bases (pp. 111-129). Cambridge: Cambridge University Press.

Power, C. (2000). Secret language use at female initiation. In C. Knight, M. Studdert-Kennedy & J. R. Hurford (Eds.), The Evolutionary Emergence of Language (pp. 81-98). Cambridge: Cambridge University Press.

Premack, D. (1971). Language in chimpanzees? Science, 172, 808-822.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. Evolution and Human Behavior, 27(4), 283-296.

Quine, W. V. O. (1960). Word and Object. Cambridge, MA: MIT Press.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In J. Kelemen & P. Sosík (Eds.), Advances in Artificial Life: ECAL6 (pp. 357-366). Berlin: Springer.

Quinn, M. (2004). The Evolutionary Design of Controllers for Minimally-equipped Homogeneous Multi-robot Systems. Unpublished DPhil. Thesis, University of Sussex.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of Red Deer stags. Animal Behaviour, 65, 317-329.

Recanati, F. (2004). Literal Meaning. Cambridge: Cambridge University Press.

Rice, S. H. (2004). Evolutionary Theory: Mathematical and Conceptual Foundations. Sunderland, MA: Sinauer.

Richardson, R. C. (1996). The prospects for evolutionary psychology: Human language and human reasoning. Minds and Machines, 6, 541-557.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. Cognitive Brain Research, 3(2), 131-141.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. Ohio State University Working Papers in Linguistics, 49, 91-136.

Roberts, G. (in press). Language and the freerider problem: An experimental paradigm. Biological Theory.

Rohwer, S. (1975). The social significance of avian winter plumage variability. Evolution, 29, 593-610.

Rohwer, S., & Rohwer, F. C. (1978). Status signalling in Harris' Sparrows: experimental deceptions achieved. Animal Behaviour, 26, 1012-1022.

Rose, S. (1997). Lifelines: Biology, Freedom, Determinism. London: Penguin.

Rose, S. (2007). Thinking allowed. Guardian, February 17th.

Rosenburg, J., & Tunney, R. J. (2008). Human vocabulary use as display. Evolutionary Psychology, 6(3), 538-549.

Ryan, M. (1990). Sexual selection, sensory systems, and sensory exploitation. Oxford Surveys in Evolutionary Biology, 5, 156-195.

Sahlins, M. (1977). The Use and Abuse of Biology. Ann Arbor, MI: University of Michigan Press.

Samuels, R. (1998). Evolutionary psychology and the massive modularity hypothesis. The British Journal for the Philosophy of Science, 49, 575-602.

Schelling, T. C. (1960). The Strategy of Conflict. Oxford: Oxford University Press.

Scherer, K. R., Feldstein, S., Bond, R. N., & Rosenthal, R. (1985). Vocal cues to deception: A comparative channel approach. Journal of Psycholinguistic Research, 14, 409-425.

Schiffer, S. (1972). Meaning. Oxford: Clarendon Press.

Schillinger, K. (2008). A Pilot Study on the Cumulative Cultural Transmission of Emerging Communication Systems. Unpublished MSc. Dissertation, University of Edinburgh.

Searcy, W. A., & Nowicki, S. (2007). The Evolution of Animal Communication. Princeton, NJ: Princeton University Press.

Searle, J. (1969). Speech Acts. Cambridge: Cambridge University Press.

Searle, J. (1979). Expression and Meaning: Studies in the Theory of Speech Acts. Cambridge: Cambridge University Press.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. Proceedings of the National Academy of Sciences, 104(18), 7361-7366.

Semmann, D., Krambeck, H.-J., & Milinski, M. (2004). Strategic investment in reputation. Behavioral Ecology and Sociobiology, 56, 248-252.

Semmann, D., Krambeck, H.-J., & Milinski, M. (2005). Reputation is valuable within and outside one's own social group. Behavioral Ecology and Sociobiology, 57(6), 611-616.

Seuren, P. A. M. (1986). The self-styling of Relevance Theory. Journal of Semantics, 5(2), 123-143.

Seyfarth, R. M., & Cheney, D. L. (2003). Signallers and receivers in animal communication. Annual Review of Psychology, 54, 145-173.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. Science, 210, 801-803.

Shannon, C. E., & Weaver, W. (1949). The Mathematical Theory of Information. Urbana, IL: University of Illinois Press.

Shettleworth, S. J. (1998). Cognition, Evolution, and Behavior. Oxford: Oxford University Press.

Silk, J. B., Kaldor, E., & Boyd, R. (2000). Cheap talk when interests conflict. Animal Behaviour, 59, 423-432.

Skinner, B. F. (1953). Science and Human Behaviour. New York, NY: MacMillan.

Skinner, B. F. (1957). Verbal Behavior. New York, NY: Appleton-Century-Crofts.

Slater, P. J. B. (1983). The study of communication. In T. R. Halliday & P. J. B. Slater (Eds.), Animal Behaviour II: Communication (pp. 9-42). Oxford: Blackwell.

Smith, A. D. M. (2008). Protolanguage reconsidered. Interaction Studies, 9(1), 98-114.

Smith, A. D. M., Smith, K., & Ferrer i Cancho, R. (Eds.). (2008). The Evolution of Language: Proceedings of the 7th International Conference on the Evolution of Language. Singapore: World Scientific.

Smith, K. (2004). The evolution of vocabulary. Journal of Theoretical Biology, 228, 127-142.

Smith, K., & Kirby, S. (2008). Cultural evolution: Implications for understanding the human language faculty and its evolution. Philosophical Transactions of the Royal Society of London, B, 363, 3591-3603.

Smith, W. J. (1977). The Behavior of Communicating. Cambridge, MA: Harvard University Press.

Sober, E. (1984). The Nature of Selection: Evolutionary Theory in Philosophical Focus. Cambridge, MA: MIT Press.

Sommerfeld, R. D., Krambeck, H.-J., & Milinski, M. (2008). Multiple gossip statements and their effect on reputation and trustworthiness. Proceedings of the Royal Society of London, B, 275, 2529-2536.

Sommerfeld, R. D., Krambeck, H.-J., Semmann, D., & Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. Proceedings of the National Academy of Sciences, 104(44), 17435-17440.

Spence, M. (1973). Job market signalling. Quarterly Journal of Economics 87, 355-374.

Sperber, D. (1990). The evolution of the language faculty: A paradox and its solution. Behavioral and Brain Sciences, 13(4), 756-758.

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

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. Cognition, 57(1), 31-95.

Sperber, D., & Girotto, V. (2002). Use or misuse of the selection task? Rejoinder to Fiddick, Cosmides and Tooby. Cognition, 85, 277-290.

Sperber, D., & Origgi, G. (2004). A pragmatic perspective on the evolution of language. Interdisciplines, from http://www.interdisciplines.org/coevolution/papers/6

Sperber, D., & Wilson, D. (1982). Mutual knowledge and relevance in theories of comprehension. In N. Smith (ed.), Mutual Knowledge (pp. 61-68). New York, NY: Academic Press.

Sperber, D., & Wilson, D. (1986). Relevance: Communication and Cognition. Oxford: Blackwell.

Sperber, D., & Wilson, D. (1987). Précis of: Relevance: Communication and cognition. Behavioral and Brain Sciences, 10(4), 697-754.

Sperber, D., & Wilson, D. (1995). Relevance: Communication and cognition (2nd. ed.). Oxford: Blackwell.

Sperber, D., & Wilson, D. (2002), Pragmatics, modularity and mind-reading. Mind & Language 17, 3-23.

Sperber, D., & Wilson, D. (2008). Epistemic Vigilance. Paper presented at the workshop on pragmatics and social cognition, UCL, London, April.

Stam, J. H. (1976). Inquiries into the Origin of Language: The Fate of a Question. New York, NY: Harper & Row.

Steels, L. (2003). Evolving grounded communication for robots. Trends in Cognitive Science, 7(7), 308-312.

Stegmann, U. E. (2005). John Maynard Smith's notion of animal signals. Biology and Philosophy, 20(5), 1011-1025.

Strawson, P. (1964). Intention and convention in speech acts. Philosophical Review, 73, 439-460.

Számadó, S., & Szathmáry, E. (2006). Selective scenarios for the emergence of natural language. Trends in Ecology and Evolution, 21(10), 555-561.

Takahashi, M., Arita, H., Hiraiwa-Hasegawa, M., & Hasegawa, T. (2008). Peahens do not prefer peacocks with more elaborate trains. Animal Behaviour, 75(4), 1209-1219.

Tallerman, M. (Ed.). (2005). Language Origins: Perspectives on Evolution. Oxford: Oxford University Press.

Taylor, P. W., Hasson, O., & Clark, D. L. (2000). Body postures and patterns as amplifiers of physical condition. Proceedings of the Royal Society of London, B, 267, 917-922.

Teller, D. Y., & Bornstein, M. H. (1987). Infant color vision and color perception. In P. Salapatek & L. Cohen (Eds.), Handbook of Infant Perception; vol. 1: From Sensation to Perception, (pp. 185-236). Orlando, Florida: Academic Press.

Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence? Science, 206, 891-902.

Tinbergen, N. (1952). "Derived" activities; their causation, biological significance, origin, and emancipation during evolution. Quarterly Review of Biology, 27(1), 1-32.

Tinbergen, N. (1963). On the aims and methods of ethology. Zeitschrift für Tierpsychologie, 20, 410-433.

Tomasello, M. (1999). The Cultural Origins of Human Cognition. Cambridge, MA: Harvard University Press.

Tomasello, M. (2003a). Constructing a Language: A Usage-based Theory of Language Acquisition. Cambridge, MA: Harvard University Press.

Tomasello, M. (2003b). On the different origins of symbols and grammar. In M. H. Christiansen & S. Kirby (Eds.), Language Evolution (pp. 94-110). Oxford: Oxford University Press.

Tomasello, M. (2008). Origins of Human Communication. Cambridge, MA: MIT Press.

Tomasello, M., & Call, J. (1997). Primate Cognition. Oxford: Oxford University Press.

Tomasello, M., George, B., Kruger, A., Farrar, J., & Evans, E. (1985). The development of gestural communication in young chimpanzees. Journal of Human Evolution, 14, 175-186.

Tomasello, M., Gust, D., & Frost, G. T. (1989). The development of gestural communication in young chimpanzees: A follow-up. Primates, 30, 35-50.

Tomasello, M., & Zuberbühler, K. (2002). Primate vocal and gestural communication. In M. Bekoff, C. Allen & M. Burghardt (Eds.), The Cognitive Animal (pp. 293-299). Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. Behavioral and Brain Sciences 28(5), 675-735.

Tooby, J., & Cosmides, L. (1990a). The past explains the present: Emotional adaptations and the structure of ancestral environments. Ethology and Sociobiology, 11, 375-424.

Tooby, J., & Cosmides, L. (1990b). Toward an adaptationist psycholinguistics. Behavioral and Brain Sciences, 13(4), 760-762.

Tooby, J., & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D. M. Buss (Ed.), The Handbook of Evolutionary Psychology (pp. 5-67). Hoboken, NY: Wiley.

Tricks of the traders. (2008). Guardian Magazine, May 3rd.

Trivers, R. L. (1971). The evolution of reciprocal altruism. Quarterly Review of Biology, 46, 35-57.

van der Henst, J.-B., & Sperber, D. (2004). Testing the cognitive and communicative principles of relevance. In I. A. Noveck & D. Sperber (Eds.), Experimental Pragmatics (pp. 141-171). London: Palgrave Macmillan Ltd.

van Rooij, R. (2003). Being polite is a handicap: Towards a game theoretical analysis of polite linguistic behaviour. In M. Tennenholtz, ed., Proceedings of the 9th Conference on Theoretical Aspects of Rationality and Knowledge (pp.45-58). New York, NY: ACM Press.

Veblen, T. (1899). The Theory of the Leisure Class. London: MacMillian.

Versluis, M., Schmitz, B., von der Heydt, A., & Lohse, D. (2000). How Snapping Shrimps snap: Through cavitating bubbles. Science, 289, 2114-2117.

von Stackelberg, H. (1934). Marktform und Gleichgewicht. London: Springer Verlag.

Vygotsky, L. (1986). Thought and Language. Cambridge, MA: MIT Press.

Wason, P. (1966). Reasoning. In B. M. Foss (Ed.), New Horizons in Psychology. Harmondsworth: Penguin.

Wason, P. (1968). Reasoning about a rule. Quarterly Journal of Experimental Psychology, 20, 273-281.

Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. Science, 288, 850-852.

Wedgwood, D. (2005). Shifting the Focus: From Static Structures to the Dynamics of Interpretation. Oxford: Elsevier.

Weinburg, S. (1992). Dreams of a Final Theory. New York, NY: Pantheon.

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism and strong reciprocity. Journal of Evolutionary Biology, 20, 415-432.

Whitfield, D. P. (1987). Plumage variability, status signalling and individual recognition in avian flocks. Trends in Ecology and Evolution, 2, 13-18.

Wilkins, J. F., & Godfrey-Smith, P. (in press). Adaptationism and the adaptive landscape, Biology and Philosophy.

Wilks, Y. (1986). Relevance and beliefs. In T. Myers, K. Brown & B. McGonigle (Eds.), Reasoning and Discourse Processes (pp. 265-289). London: Academic Press.

Williams, G. C. (1966). Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought. Princeton, NJ: Princeton University Press.

Wilson, D., & Sperber, D. (1981). On Grice's theory of conversation. In P. Werth (Ed.), Conversation and Discourse (pp.158-178). London: Croom Helm.

Wilson, D., & Sperber, D. (1986). The self-appointment of Seuren as censor: A reply to Pieter Seuren. Journal of Semantics, 5(2), 145-162.

Wilson, D., & Sperber, D. (2002a). Relevance theory. In UCL Working Papers in Linguistics XIV (pp. 249-287).

Wilson, D., & Sperber, D. (2002b). Truthfulness and relevance. Mind, 111, 583-632.

Wilson, E. O. (1975). Sociobiology: The New Synthesis. Cambridge, MA: Harvard University Press.

Wilson, E. O. (1998). Consilience: The Unity of Knowledge. New York, NY: Vintage.

Wray, A. (Ed.). (2002). The Transition to Language. Oxford: Oxford University Press.

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, S., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from co-operators. Evolution and Human Behavior, 24, 290-301.

Yus Ramos, F. (1998). A decade of Relevance Theory. Journal of Pragmatics, 30(3), 305-345.

Zahavi, A. (1975). Mate selection: A selection for a handicap. Journal of Theoretical Biology, 53, 205-214.

Zahavi, A., & Zahavi, A. (1997). The Handicap Principle: A Missing Piece of Darwin's Puzzle. Oxford: Oxford University Press.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun & K. Obermayer (Eds.), Advances in Neural Information Processing Systems XV (proceedings of NIPS'02) (pp. 51-58). Cambridge, MA: MIT Press.