The evolution of communication

Humans may be exceptional

Thomas C. Scott-Phillips
University of Edinburgh

Communication is a fundamentally interactive phenomenon. Evolutionary biology recognises this fact in its definition of communication, in which signals are those actions that cause reactions, and where both action and reaction are designed for that reason. Where only one or the other is designed then the behaviours are classed as either cues or coercion. Since mutually dependent behaviours are unlikely to emerge simultaneously, the symmetry inherent in these definitions gives rise to a prediction that communication will only emerge if cues or coercive behaviours do so first. They will then be co-opted for communication. A range of case studies, from animal signalling, evolutionary robotics, comparative psychology, and evolutionary linguistics are used to test this prediction. The first three are found to be supportive. However in the Embodied Communication Game, a recent experimental approach to the emergence of communication between adult humans, communication emerges even when cues or coerced behaviours are not possible. This suggests that humans are exceptional in this regard. It is argued that the reason for this is the degree to which we are able and compelled to read and interpret the behaviour of others in intentional terms.

Keywords: communication, emergence, cues, coercion, signal, response, evolution, Embodied Communication Game, interaction, intentionality

1. Introduction

Recent years have witness a growing body of research that uses novel experimental techniques to investigate the emergence of communication between pairs or groups of interacting human adults (de Ruiter et al., this volume; Galantucci, 2005; Garrod et al., 2007; Healey et al., 2007; Scott-Phillips et al., 2009; Selten & Warglien, 2007). Collectively these studies have helped to define and shape the space of theories about how dyads and populations converge upon shared meanings. They do this not because they replicate the evolutionary history of

Interaction Studies 11:1 (2010), **78–99**. DOI 10.1075/is.11.1.07sco
ISSN 1572–0373 / E-ISSN 1572–0381

language (how could they?), but because they allow us to investigate the nature of both the communicative systems themselves and cognitive phenomena that underpin their emergence. As such, they offer a body of data that can be used to explore specific hypotheses in the evolution of language, a field that has historically been more speculative than empirical.

This paper presents and explores one such hypothesis. The two-step hypothesis of the emergence of communication, which states that communication can emerge only if it co-opts pre-existing cues or coerced behaviours, logically follows the oft-made observation that communication is fundamentally interactive. Indeed, this interactivity is so central that several disciplines recognise it in the very definition of communication. The next section takes one such definition, and shows that its most basic form not only reflects the interactivity of communication but also gives rise to the sort of definitions employed by other disciplines. The two-step hypothesis that arises from this is then described. The third section explores the hypothesis with a number of case studies. These are: (i) a study from evolutionary robotics in which pairs of simulated Khepera robots evolve communication even in the absence of dedicated communication channels; (ii) ethological accounts of the evolution of animal signals; and (iii) ontogenetic ritualisation, in which pairs of great apes (including human adult-infant dyads) shape each other's behaviour through a series of repeated interactions. All are observed to be in accord with the two-step hypothesis.

However none of these studies involve the emergence of communication between adult humans. It is here that the data from the recent body of work mentioned above is of use. The Embodied Communication Game (ECG) (Scott-Phillips et al., 2009) allows for a direct test of the two-step hypothesis, as cues and coerced behaviours are impossible within it. If the hypothesis applies to humans then this should mean that they would be unable to communicate within the ECG, but this is in fact not the case. Humans thus appear to be exceptional. In the final section of the paper it is suggested that the reason for this is our capacity and inclination to interpret the behaviour of others in intentional terms.

2. The nature of communication; and the two-step hypothesis of its emergence

Communication is a fundamentally interactive phenomenon. This fact has been long recognised by researchers in a wide range of disciplines, including evolutionary biology (Krebs & Dawkins, 1984; Maynard Smith & Harper, 2003), psycholinguistics (Clark, 1996; Pickering & Garrod, 2004), the philosophy of language

(Lewis, 1969; Wittgenstein, 1953) and pragmatics (Grice, 1975; Sperber & Wilson, 1986). Evolutionary biology captures this interactivity in its definition of communication (Maynard Smith & Harper, 2003; Scott-Phillips, 2008). There, 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 then the successful completion of a signalling act (Maynard Smith & Harper, 2003; Scott-Phillips, 2008). This definition correctly matches our *prima facie* intuitions about what is and what is not communication (Stegmann, 2005). Moreover, its symmetrical format, which reflects the interactivity that lies at the heart of communication, allows for a straightforward distinction between communication, cues and coercion. The second clause, that the signal be evolved to affect the other organism, is used to distinguish a signal from a *cue* (Hasson, 1994); and the third, that the effect be evolved to be affected by the signal, is used to distinguish signals from *coercive* behaviours (for further discussion and examples see Maynard Smith & Harper, 2003; Scott-Phillips, 2008). The situation is summarised in Table 1.

Table 1. The relationship between communication, cues and coercion in biological communication

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

Can this definition be generalised? That is, can it be applied outside of evolutionary biology? In particular, can it be used to think about human communication? There is an immediate and obvious problem: individual utterances do not themselves evolve, and hence it is not immediately clear how the definition might apply. However, this is not the major problem it first appears. Consider any animal signal; the red deer roar (Clutton-Brock & Albon, 1979), 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 in fact mean by that is the evolution of the capacity and instinct to roar. Humans, similarly, have a capacity and instinct to use language. Thus when applied to humans the shorthand employed by evolutionary biology might misleadingly lead us to talk about the (cultural) evolution of human utterances when what we in fact mean is the evolution of the capacity to produce utterances and the pragmatics of doing so. Thus although essentially harmless in

the case of animal signals, to insist upon such shorthand for humans is to invite misunderstanding. To overcome this we may replace *evolved for* in our definition with the more general notion of *design*. After all, natural selection is, ultimately, the only source of design in nature (Dawkins, 1986; Dennett, 1995), and thus this substitution accounts for precisely the same phenomena as the original definition. However the substitution allows us to think about communication in terms of the proximate sources of design, for example human intentionality (see Dennett, 1987; 1995 on why intentionality is a type of designed behaviour), which may be more explanatorily satisfactory than to think only in terms of ultimate evolutionary function (for more on the proximate/ultimate distinction see Mayr, 1963). In the case of linguistic utterances that proximate source will be human cognition. Indeed, 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:

Communication *occurs when an act or structure:*

1. *produces a reaction in another organism;*
2. *was designed to produce such a reaction; and*
3. *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 Table 1 above, but with 'evolved' replaced by 'designed'. Examples of how animal signals satisfy this definition can be found elsewhere (e.g. Maynard Smith & Harper, 2003). For humans, consider the following. If I tell my friend to meet me in George Square at 7pm this evening, and they do 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. Failed communication is also possible. If, for example, my instructions to my friend were ambiguous or confused such that she turned up at 8pm rather than 7pm, then this would be an instance of failed communication. Here, although the stimulus was designed to achieve a goal, it was not sufficiently well designed

for that end. Similarly, my friend may fail to appear at the anointed time, in which case her reaction would not have been well-designed. Note that these scenarios are not the same thing as cues or coerced behaviours. There, either the stimulus (in the case of a cue) or the reaction (in the case of coercion) is *not* designed; here, both are designed, but one or the other is not *well* designed. This difference will be important later, when we discuss why humans may be an exception to trends witnessed elsewhere. (To complete the range of examples, if a second friend overheard my conversation with the first friend, and subsequently also arrived at George Square at 7pm, perhaps without telling us that she would be there, then my instructions to the first friend would have acted as a cue to the second friend. If, on the other hand, there was some acquaintance of mine who did not want to come along at all, but I had, for whatever reason, physically dragged them to George Square, this would be an instance of coercion.)

This definition is very much in accord with accounts of meaning and communication that arise from pragmatics, the philosophy of language, and psycholinguistics. Grice posited (1975; see also Levinson, 1983; 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); 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 and psycholinguists (e.g. Clark, 1996; Tomasello, 2003) 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.

There is a sense in which this revised definition 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. Evolutionary biology places great emphasis on the organism's relationship with its environment, and this is reflected in the definition we have obtained. However, other disciplines may not

so readily see the need to place so much emphasis on external factors. An account based on information transfer might thus be preferred (but see Scott-Phillips, 2008 for reasons why such accounts actually reduce to the definition given here), and if so then 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 already mentioned, this synergy is inherent in communication. It is desirable that signals and responses be explanations for each other. An alternative way to capture this symmetry is to refer to causality (Oliphant, 1997): an interaction is communicative if signals and responses can be seen as causes of each other. As before, this symmetrical causality highlights that what is crucial about communication is that its two halves – signals and responses – are explanations for each other.

What implications does this definition have for matters of evolution and emergence? Signals and responses are mutually dependent: each depends for its utility on the existence of the other. Whilst one might occur for its own reasons, independent of its value when combined with the other, it is *a priori* unlikely that both will appear simultaneously. This is most clearly seen if we think in terms of underlying genetic changes. Adaptive mutations are rare enough on their own. Correspondingly, simultaneous mutations in different but interacting organisms, which depend for their adaptive value upon the existence of the other, are vanishingly unlikely. The obvious way in which this problem might be overcome is if one or the other behaviour has its own adaptive value independent of the existence of the other. That is, either a cue (where the reaction is already adapted) or a coercive behaviour (where the action is already adapted) could exist beforehand, for its own reasons. This could then be co-opted for communication. This observation gives rise to a two-step hypothesis regarding the emergence of communication; namely that it will involve the co-option of design that has been built into one side of the interaction first, for reasons independent of communication. To test this prediction, four case studies of the evolution and emergence of communication will be briefly reviewed: the evolution of communication in simulated Khepera robots; ethological accounts of the evolution of animal signals; ontogenetic ritualisation; and the Embodied Communication Game, a recent experimental approach with human participants. This range of literatures allows for a diverse testing of the two-step hypothesis: two of these case studies describe phylogenetic emergence, while the other two are ontogenetic; at the same time, two involve naturalistic observation, and two are experimental; see Table 2.

Table 2. Classification of the case studies of the emergence of communication

		type of study	
		experimental	**naturalistic observation**
timescale of emergence	phylogenetic	simulated Khepera robots	animal signals
	ontogenetic	Embodied Communication Game	ontogenetic ritualisation

3. Three case studies of the emergence of communication

3.1 Simulated Khepera robots

In this first case study (Quinn, 2001), pairs of simulated robots must 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 infrared sensors, which allow them to detect objects in front and to the side of them, but not behind. Their behaviours are controlled by an 'innate' (that is, fixed and not able to learn) neural network. Pairs of robots are placed in an obstacle-free environment, and their task is to move their mutual centre of gravity (i.e. the point half way between them) as far as possible. Put another way, they must travel as far as possible away from their starting positions, 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 their partner will score poorly. Collisions with the other robot also carry a fitness penalty. 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.

Despite the difficulty of the task, and the absence of any dedicated communication channel, some pairs of robots do evolve to communicate with one another. The systems that are employed take the following form. First, the robots rotate counterclockwise. This allows the robots to detect the location of the other robot. They continue to rotate until they are facing the other robot. The first robot to do this then begins to oscillate back and forth. The other robot, meanwhile, continues to rotate until it, too, is facing its partner. When that occurs then the second robot begins to travel backwards and the first forwards. They are now, in effect, travelling off together

in the same direction. The communicative behaviour here is the oscillations of the first robot, which we can anthropomorphically gloss as meaning 'after you' (Kirby, 2002). This signals to the still rotating robot that it the first robot is aligned and ready to travel (Quinn, 2004). The whole process is depicted in Figure 1.

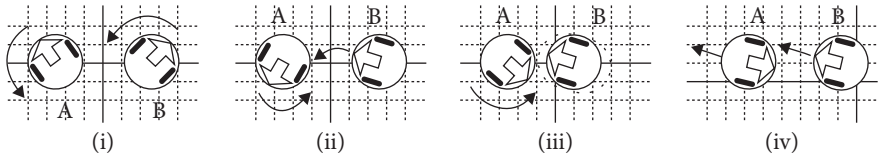

Figure 1. (reproduced from Quinn, 2001) Communication between two Khepera robots. In (i) the two robots rotate in an anticlockwise direction. This continues until (ii) one of the robots (in this case B) is aligned with the other. Then in (iii) the aligned robot oscillates back and forth while the other 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 they will score positively. However it equally often resulted in collisions, which, recall, 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 behaviours, both of which are adaptive: remain close to the other robot; and do not collide with the other robot. Crucially, both are cued behaviours. That is, both the presence and the non-presence of the other robot is a cue: they alter the robot's behaviour and that reaction is one that has been designed, in this case by natural selection. 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.

This deadlock is broken, some generations later, when one robot evolves, by random mutation, to advance when faced with such a situation. The other robot, faced with imminent collision, is now under strong selection pressure 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 1. 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. This case study, then, is exactly consonant with the two-step hypothesis: the robots first evolve cues, and these then provide the scaffolding that allows design to evolve on the other half of the communicative dynamic.

3.2 The evolution of animal signals

Systematic study of the evolution of animal signals has been ongoing for at least 50 years, since the seminal work of some of the founding fathers of ethology (e.g. Huxley, 1966; Lorenz, 1965; Tinbergen, 1952). Ritualisation is the name given by this literature to the evolution of a cue into a signal (Bradbury & Vehrencamp, 1998; Hauser, 1996; Maynard Smith & Harper, 2003), precisely the process described by the two-step hypothesis. Moreover, the general consensus is that "most signals probably evolved by the ritualisation of cues that other individuals were already using to gain information" (Maynard Smith & Harper, 2003, p. 68). Some examples will serve to illustrate.

Dogs (amongst others) bare their teeth as a sign of aggression. Why so? Initially, receivers would have used the open mouth as a cue, since it displays the dog's teeth and hence provides important information about its fighting ability. Once common, this use of cues places a pressure for some dogs, namely those with the highest RHP (Resource Holding Potential; a composite measure of all factors that affect fighting ability (Parker, 1974)), to display their teeth, as it will prevent weaker individuals initiating a fight with them. Yet now there is a pressure on the dogs with the next highest RHP levels to do the same, so as to differentiate themselves from the remainder of the pack. This logic then cascades its way through the whole population, and so all dogs evolve to display their teeth in aggressive scenarios. Many non-human primates do the same, and there are many more examples in the literature (for extensive detail and discussion see Bradbury & Vehrencamp, 1998; Hauser, 1996; Maynard Smith & Harper, 2003).

A second example is the use of urine (and/or faeces) to mark territorial boundaries. Many mammals relieve themselves when they experience extreme fear, which may occur as they leave the safe environment of their own territory. If so, then conspecifics could use the presence of urine as a guide to the area within which the focal animal feels safe. The urine would then be a cue. However there is now a pressure for the focal animal to urinate so as to inform the conspecific that this is their natural territory, even when they are not fearful. This would then result in a behaviour that has been selected to inform the other animal, and in which the

reaction was also selected for. In other words, urination can act as a signal. This account (Lorenz, 1970) is clear case of ritualisation, and as such is consistent with the two-step hypothesis.

There is, however, some equivocation in the already quoted statement about the ubiquity of ritualisation: "*most* signals *probably* evolved…" (italics added). This is because there does exist an alternative process, *sensory exploitation*, in which an organism's receptive abilities are exploited by some other organism for its own ends (see Ryan, 1990). For example, 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, then, be an instance of coerced behaviour. With that coercion comes new evolutionary pressures for the female. 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. This alternative to ritualisation is, then, a process by which coerced behaviours, rather than cues, are co-opted for communication; which is precisely what the two-step hypothesis predicts. (A terminological note: there will be scenarios where *exploitation*, with its anthropomorphically negative connotations, is not such a good term. For this reason some authors prefer the term *sensory bias*.)

In general, then, the interactivity that is part of the definition of communication gives rise to a chicken-and-egg problem that ritualisation and sensory exploitation resolve by selecting first for either a cue (ritualisation) or a coerced behaviour (sensory exploitation) and then later, once that behaviour is established, for the other half of the equation. Are there any other ways in which natural selection can give rise to communication? The only apparent alternative is drift. This is possible in principle, but as already discussed, it 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. These conditions make this an extremely unlikely occurrence. We may thus conclude that this phylogenetic case study, like that with the simulated Khepera robots, is very much in alignment with the two-step hypothesis. We turn now to an ontogenetic case study of the emergence of communication.

3.3 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 where 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 finally
iv. *A* anticipates *B*'s anticipation of *X* and hence produces *X* in ritualised form so as to elicit *Y*.

Does this process fit the general picture that cues or coerced behaviours precede communication? 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). At this point the infant is coercing the mother, and it is hard to see how steps (iii) and (iv) constitute the co-option of that coercion into a signal. However, separate to that coercion, we see the appearance of a cue in (iii): the infant's initial movements act as a cue to the mother that the infant wishes to feed. That is, they alter the mother's behaviour and the mother exhibits a designed reaction to that movement. 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. Only now can we label the interaction communicative. The important observation for the present purposes is that a cue was established first (in (iii)), 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 (which is, indeed, occasionally called *phylogenetic ritualisation* (e.g. 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*). As with the nursing poke, this is coercion, but it is not this coercion that will be co-opted for communication. As above, what is co-opted is the cue that emerges in (iii), when the child begins to recognise the initial movements that are associated with lifting and so lifts their arms in anticipation. This is a cue because the adult's movements 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 likely. 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, we should bear 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 aspect of the process of emergence. The child's use of 'arms up' to indicate their desire to be lifted may therefore 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 is a stepping-stone on the way to communication. Thus both of the classic instances of ontogenetic ritualisation fit the picture described by the two-step hypothesis: design is built into one side of the interaction and consequently creates a pressure for design to appear on the other side.

This case study serves to illustrate that the two-step hypothesis applies just as well to ontogeny as it does to phylogeny. There is, then, no *a priori* reason to suppose that the two-step hypothesis will not apply to the emergence of communication between pairs or groups of interacting humans, which is the topic to which we now turn.

4. The Embodied Communication Game

As discussed in the introduction, there has been a recent interest in the use of experimental games to study the emergence of communication between pairs or groups of interacting individuals (de Ruiter et al., this volume; Galantucci, 2005; Garrod et al., 2007; Healey et al., 2007; Scott-Phillips et al., 2009; Selten & Warglien, 2007). This body of work thus offers a number of candidate case studies with which to further investigate the two-step hypothesis, this time with adult humans. However, despite their valuable contributions to other questions related to the emergence of communication, most of these studies do not speak directly to this specific task. That is, they do not explicitly investigate the problem with which the two-step hypothesis is concerned – the question of how pairs of behaviours that depend upon each other for their function but which emerge independently can come into being at all; instead they focus on the (equally important) question of how signallers and receivers can agree on shared meaning-form pairs once the communicative relationship is already established. Of course, answers to the latter question can inform the former (and vice versa), but since the specific present objective is to investigate the two-step hypothesis, this section will focus only on the study that most explicitly addresses the latter problem.

The Embodied Communication Game (ECG) (Scott-Phillips et al., 2009) is an interactive, cooperative two-player game played over a computer network. The basic idea is that 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 also what behaviours will be communicative at all. The question can then be asked: Is it possible for communication to emerge under such circumstances; and if so, how does that occur?

In the ECG each player is represented as a stick man, each located in his own 2×2 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; and both players know that the experience was the same for the other player. At the beginning of each round each players' stick man begins in one of the quadrants of his/her box. This starting point is chosen at random in each round. The players can move between quadrants at will, but each move is 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.

Each press of the arrow buttons takes the stick man directly to the centre of the new quadrant at a fixed speed. The players 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 2. The pairs' final score is the highest number of points scored in succession. In order to succeed at the ECG, participants must recognise that they are able to use their movements to signal to each other, and then find some way to signal the fact that some of their movements are communicative in nature (see Scott-Phillips et al., 2009 for further details).

What happens? Pairs of participants who find a way to communicate in the ECG typically do so by first establishing a 'default colour strategy': they always travel to the same colour (say, red) if it is available. This is not a full-proof strategy, however, since in some rounds that default colour may not appear in one or other box. Under such circumstances, one of the players will perform some otherwise idiosyncratic behaviour (for example, repeatedly oscillating along one axis) – this is intended to communicate that they do not have the default colour available to them. This behaviour is then recognised as communicative (although not always immediately), and an alternative colour is chosen when the default is unavailable. Over time, the behaviour that initially meant 'no red' (or whatever the default colour is) comes to mean 'blue' (or whatever the second colour is). The participants then negotiate over movements for the remaining two colours.

Does this process follow to the same pattern as those discussed above, and thereby conform to the two-step hypothesis? No, it does not. It may intuitively be thought that the default colour strategy is a cue. However that would be to confuse cues with the notion of common ground. Common ground is knowledge that we both share, and which we both know that we both share (Clark, 1996). Cues, in contrast, are behaviours that induce a calibrated reaction, even though that was not the intention, or design, behind them. Both cues and common ground are informative, but for different reasons: cues are incidentally informative, while common ground is informative by virtue of some set of shared assumptions. The fact that something (for example the default colour) is in common ground informs both participants of important facts (that they can rely upon the other participant to behave in a way that takes advantage of the knowledge contained in common ground), but it does not do so incidentally, as is required for it to be a cue. Yet it is common ground that is seen to be important to the process of emergence observed in the ECG: the assumption that the other player will go to red provides

Figure 2. Screen-shots of the ECG. Participants play multiple rounds of the game on networked computers. These four screen-shots show each player's view (one player on the top row; the other on the bottom) both 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). Success at the game requires finding some way to communicate the intended destination colour each round

a background against which idiosyncratic behaviour can be identified as having some communicative function (Clark, 1996).

A second intuitive thought is that although cues are not possible within rounds, they are possible between rounds. One player could notice that the other pursued some recognisable pattern of behaviour, and this may be used to increase success rates. Consider again the default colour strategy. If may be pursued by one participant, and then noticed by the other, who then does the same. This allows the pair to score at above chance levels. However the use of the default colour strategy is in fact not a cue. To see why, we must ask why one of the players would pursue a strategy of always travelling to the same quadrant. This strategy is of no use unless the other player does the same; therefore the only reason to pursue such a strategy would be in the hope or expectation that the other player will do likewise. As such, it is an attempt at communication, which may or may not be successful. However unsuccessful communication is not the same thing as a cue: the former is not sufficiently well designed, while the latter is not designed at all. Cues require that the act or structure in question not be designed to cause a reaction, yet that is precisely why the default colour strategy would be pursued.

In fact, cues are near impossible in the ECG. This is because success is mutual: participants receive a joint score, one that is contingent on achieving communication. In order for cues or coerced behaviours to be possible, participants must be able to notice and exploit some regularity in another's behaviour, one that is not itself designed to achieve adaptive ends. This would require separate (but not necessarily conflicting) reward structures. However this is not how the ECG is set up; the two participants have a joint reward structure. This fact means that any regularity in one or the other participant's behaviour is either intended to influence the other behaviour, or is pure coincidence. If the regularity is intended to influence, then it is not a cue, and plainly it cannot be coercive: it does not force the other participant into any particular behaviour. A participant could choose the same colour in successive rounds purely coincidentally, but this is *a priori* unlikely. This nascent possibility is why cues are considered near impossible rather than outright impossible.

However in the experimental work with the ECG, this route to communication is not observed. Some pairs of participants are able to build communication systems, often in the way described above, in which a default colour strategy provides some common ground that can then be used to bootstrap a communication system. This is contrary to the predictions of the two-step hypothesis. Since cues and coercive behaviours are impossible within the ECG, the two-step hypothesis would expect that communication would not emerge. Yet it does. In the standard version of the ECG (described above), 7 of 12 pairs are able to build communication systems. A second version differs from the standard version in that when a point is scored then the colour on which the point is scored does not appear in

both boxes in the next round. This makes the default colour strategy non-profitable), but even under these difficult circumstances, 2 of 12 pairs are able to build a communication system.

5. Discussion

It seems, then, that humans are able to build communication systems in a way that makes them exceptional in the natural world. This conclusion might be resisted on the grounds that only three other case studies were considered. However one of those, the ethological perspective on the evolution of animal signals, is in fact a general statement about how communication evolves in many other species, and so the number of instances of emergence that have actually been considered is quite large. In addition, the work with simulated Khepera robots shows that these conclusions hold in a laboratory setting, and the observations of ontogenetic ritualisation show that the same basic process applies to ontogeny just as it does to phylogeny. This is, then, a wide range of scenarios. We should therefore expect humans to follow the same pattern as is observed in these literatures. That they do not does suggest exceptionality.

Why might this be so? As already discussed, in the typical process of emergence observed in the ECG some common ground is established, and then used as the contextual foundation for the communication of a participant's informative intention. It seems, then, that the ability to interpret the behaviour of others in functional, goal-oriented terms (in a sufficiently rich context) is necessary for success at the ECG. Certainly, without it the task is more-or-less impossible, since there would then be no way to distinguish between movement whose purpose is communication and movement whose purpose is travel. More generally, linguistic communication poses a number of interactional challenges that we routinely perform with ease (Levinson, 1995); and the detection and interpretation of goal-directed behaviour is an utterly central feature of human cognition; we are not just able but compelled to interpret the behaviour of others in such a way (Csibra & Gergely, 2007; Dennett, 1987).

To what extent is this quality uniquely human? Some other species (in particular non-human primates) show some impressive capabilities in this respect (Tomasello, 2008). The exact nature and degree of these capabilities remains an open empirical question, but plainly there is some difference between humans and other species. Moreover, that difference is a crucial one; "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 suggests that the human exceptionality with respect to the two-step hypothesis is one of the degree to which we are able to detect and interpret intentional behaviour.

After all, how can you converse if you are not sensitive to what your interlocutor knows and does not know? 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). Correspondingly, these skills represent the starting point for some of the recent experimental studies into the emergence of communication that have motivated this special issue (e.g. de Ruiter et al., 2007). Sophisticated forms of intention reading thus present themselves as a plausible candidate explanation as to why humans are able to break the trends of emergence observed in the other case studies.

This is, however, an inexact and thus somewhat unsatisfying conclusion. I believe we can be more precise. Much human communication, and all linguistic communication, is ostensive-inferential (Grice, 1975; Levinson, 1983; Sperber & Wilson, 1986): it involves the provision and interpretation of evidence for the meaning that the speaker intends to convey. Moreover, this is the only sort of communication possible in the ECG – the alternative of directly coded symbols is plainly not possible, since there is no pre-existing code. Yet ostensive-inferential communication is not simply just another type of intentional behaviour. Ostension (the speaker's provision of evidence for the meaning they wish to convey) requires that speakers embed intentions within other intentions (Bennett, 1976; Dennett, 1987; Grice, 1975; Sperber & Wilson, 1986). Specifically, a communicative intention is an intention that the listener understand that the speaker has an informative intention; which is, in turn, an intention that the listener understands the content of the utterance (Grice, 1975; Sperber & Wilson, 1986). All of which means that communicative intentions embed several other layers of intentionality, a fact that makes their performance and detection a more complex task than the performance and detection of other sorts of intentions. It demands that speakers and listeners not just represent their interlocutor's representations, but that they represent their interlocutors' representations of those representations. The metarepresentational nature of ostensive-inferential communication makes it a particularly demanding task, and may explain why humans appear to be exceptions to the two-step hypothesis.

Enculturation offers an alternative explanation to the above argument that the biological capacity for ostension and inference is what makes humans an exception with respect to the emergence of communication. Participants in the ECG were fluent language users, and as such were already users of a vast repertoire of symbolic associations that can be and are used in everyday communication. This, rather than any particular cognitive predisposition, may explain whatever success they achieved in the ECG. However, highly enculturated non-human primates are only able to acquire a few hundred different symbolic associations (Savage-Rumbaugh et al., 1996), and even these are sometimes used inappropriately. Furthermore, non-human primates compare poorly with 2.5 year-old

human infants on a range of social tasks that includes the detection of communicative intent, but compare favourably with the same group on a range of physical tasks (Herrmann et al., 2007). These results suggest that while enculturation plays a role in the acquisition of symbolic associations, non-human primates lack some important cognitive capacities that underpin the emergence of learnt, symbolic communication systems.

To summarise, this article makes four key points. The first is that there are clear trends in the literatures that deal with the emergence of communication; namely that cues and coercive behaviours appear to be pre-requisite. This applies to both ontogenetic and phylogenetic emergence, and to both natural observation and laboratory-based studies. The second is that these trends are entirely predictable from the way that communication is defined. The definition employed was taken, initially, from evolutionary biology, but was shown to capture the essential features of communication as identified by philosophers of language and psycholinguists. The third point is that humans are an exception to these trends. This is most clearly demonstrated by work with the ECG, although it should be mentioned that other experimental work on the emergence of communication (in particular de Ruiter et al., this volume) leads us to similar conclusions. The fourth and final point is that the most likely candidate explanation for this exceptionality is, in general, our capacity to read the intentions of others; and more specifically, our capacity for ostensive-inferential communication, which requires several layers of mental metarepresentation.

There are at least two predictions that follow directly from this analysis. The first is that although humans can succeed at the ECG, populations of agents who lack our mind-reading abilities should not be able to, either through learning or through natural selection. A computational model of the ECG could be developed to test this idea. The second is that the simulated Khepera robots discussed in Section 3.1 would not evolve communication if they could not have first evolved the cued behaviours that they do. Since these cues are a consequence of the fact that the robots incur a fitness penalty for collisions, the analysis presented here would predict that if this penalty were removed, then cues would not appear, and communication would hence no longer evolve. Both of these predictions are potentially profitable avenues for future research.

Acknowledgements

TSP was sponsored in this work by grants from the AHRC and ESRC. He wishes to thank Andy Smith, Monica Tamariz and three anonymous reviewers for valuable discussion and comments on a previous draft.

References

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

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

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

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.

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

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

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

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

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.

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

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

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

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.

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

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

Hauser, M.D. (1996). *The Evolution of Communication*. Cambridge, Mass.: MIT 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.

Herrmann, E., Call, J., Hernàndez-LLoreda, M.V., Hare, B. & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*(5843), 1360–1366.

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

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

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.

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

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

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

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

Lorenz, K. (1970). *Studies in Animal and Human Behaviour*, vol. 1. London: Methuen.

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

Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, MA: Harvard University Press.
Oliphant, M. (1997). *Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language.* Unpublished Ph.D. thesis, University of California, San Diego.
Parker, G.A. (1974). Assessment strategy and the evolution of animal conflicts. *Journal of Theoretical Biology, 47*, 223–243.
Pickering, M.J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 1–22.
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 D.Phil. thesis.
Ryan, M. (1990). Sexual selection, sensory systems, and sensory exploitation. *Oxford Surveys in Evolutionary Biology, 5*, 156–195.
Savage-Rumbaugh, E.S., Shanker, S. & Taylor, T.J. (1996). *Apes, Language and the Human Mind*. Oxford: Oxford University Press.
Scott-Phillips, T.C. (2008). Defining biological communication. *Journal of Evolutionary Biology, 21*(2), 387–395.
Scott-Phillips, T.C., Kirby, S., & Ritchie, G.R.S. (2009). Signalling signalhood and the emergence of communication, *Cognition, 113*, 226–233.
Searle, J. (1969). *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.
Seyfarth, R.M., & Cheney, D.L. (2003). Signallers and receivers in animal communication. *Annual Review of Psychology, 54*, 145–173.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.
Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp.117–137). Oxford: Oxford University Press.
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.
Tinbergen, N. (1952). "Derived" activities; their causation, biological significance, origin, and emancipation during evolution. *Quarterly Review of Biology, 27*(1), 1–32.
Tomasello, M. (2003). *Constructing a Language: A Usage Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, Mass.: 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: MIT Press.
Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

Authors' Address

Thomas C. Scott-Phillips
Language Evolution and Computation Research Unit
School of Philosophy, Psychology and Language Sciences
University of Edinburgh

thom@ling.ed.ac.uk
Dugald Stewart Building
3 Charles Street
Edinburgh
EH8 9AD

Biographical note

Thom Scott-Phillips is an ESRC-sponsored postdoctoral research fellow at the Language Evolution and Computation Research Unit at the University of Edinburgh, where he also completed his Ph.D. He has published journal articles and conference proceedings on defining communication, on the social evolution of language, and on the emergence of symbolic communication. At the 2008 Evolang conference he was awarded the Hurford Prize for the best student presentation, for his work Signalling Signalhood and the Emergence of Communication. In 2010 he will take up a Leverhulme Early Career Fellowship entitled Change and Emergence in Communication Systems.