Cognition 113 (2009) 226–233

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Signalling signalhood and the emergence of communication

Thomas C. Scott-Phillips *, Simon Kirby, Graham R.S. Ritchie

School of Psychology, Philosophy and Language Sciences, University of Edinburgh, Edinburgh EH8 9AD, United Kingdom

ARTICLE INFO

Article history:
Received 21 August 2008
Revised 11 August 2009
Accepted 11 August 2009

Keywords:
Communication
Emergence of communication
Common ground
Language
Evolution
Symbolism
Communicative intent
Dialogue
Embodiment
Embodied communication game

ABSTRACT

A unique hallmark of human language is that it uses signals that are both learnt and symbolic. The emergence of such signals was therefore a defining event in human cognitive evolution, yet very little is known about how such a process occurs. Previous work provides some insights on how meaning can become attached to form, but a more foundational issue is presently unaddressed. How does a signal signal its own signalhood? That is, how do humans even know that communicative behaviour is indeed communicative in nature? We introduce an experimental game that has been designed to tackle this problem. We find that it is commonly resolved with a bootstrapping process, and that this process influences the final form of the communication system. Furthermore, sufficient common ground is observed to be integral to the recognition of signalhood, and the emergence of dialogue is observed to be the key step in the development of a system that can be employed to achieve shared goals.

1. Introduction

Human language is the only communication system in the natural world where the signals are both learnt and symbolic (Deacon, 1997). These twin features 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 in the first place (Oliphant, 2002)? Almost nothing is known about the answer to this question. Previous experimental (de Ruiter, Noordzij, Newman-Norland, Hagoort, & Toni, 2007; Fay, Garrod, MacLeod, Lee, & Oberlander, 2004; Galantucci, 2005; Healey, Swoboda, Umata, & King, 2007; Selten & Warglien, 2007), computational (e.g. Hurford, 1989; Noble, 2000; Nowak & Krakauer, 1999; Smith, 2004) and theoretical studies (e.g. Lewis, 1969) offer some insights; but all have, with one exception (Quinn, 2001), assumed that at the very earliest stages of a system's development individuals are able to detect that a given behaviour is intended to be communicative. Yet this cannot be taken for granted: before potential receivers can access the problem of what a communicative behaviour must mean, they must first recognise that the behaviour is indeed communicative.

The recognition of informative intent is a fundamental component of (non-natural) meaning (Grice, 1971). Yet previous work, whether it is concerned with learnt or innate symbolism, has avoided the question of how this is achieved. This has been done 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 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, Fay, Lee, Oberlander, & MacLeod, 2007; Selten & Warglien, 2007). Although this does not make communicative behaviour quite so salient as a pre-defined communication channel,

* Corresponding author. Tel.: +44 0131 650 3956.
E-mail address: thom@ling.ed.ac.uk (T.C. Scott-Phillips).

0010-0277/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2009.08.009

it nevertheless primes the receiver to interpret the signaller's behaviour in communicative terms. Finally, the possible forms that a signal might take may be pre-specified by the researcher, which renders it is immediately recognisable as a signal. 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; Noble, 2000; Nowak & Krakauer, 1999; Smith, 2004) and experimental (e.g. Selten & Warglien, 2007) work. All of these scenarios mean that the problems that are investigated are how to map form onto meanings, and in some cases (e.g. de Ruiter et al., 2007) how to construct forms, but if we wish to study the question of emergence we must address an even more foundational issue: how do (potential) receivers even know that there is a signal? Put another way, *how does a signal signal its own signalhood*? There is one previous study using evolved robots that directly addresses this question (Quinn, 2001), but that work studied the emergence of an innate, iconic system. We, on the other hand, are interested in the emergence of a learnt, symbolic system[1].

If we wish to address this question our investigative set-up must allow communicative behaviour either to emerge from non-communicative behaviour or be created *de novo*. This means, at a minimum, that we must not pre-define the communication channel, the roles of signaller and receiver, or the form space. 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. In general, to embody is to make concrete or to give physical form to some entity. For cognition, this means that the bodies that are controlled by brains are themselves an integral part of the cognitive process (see Wilson, 2002 for a review of the various ways in which this point may play out). For communication, it means, minimally, that there should be no *a priori* distinction between communicative and non-communicative behaviour. The act of communication must be situated in the world (as that world is defined by the investigative approach). There is at least one previous piece of experimental work with human participants that satisfies this condition (de Ruiter et al., 2007), and that study correspondingly offers insights into the origins of our communicative intentions. However, it is not ideally suited to the present task for two reasons. First, as mentioned above, it pre-defines the roles of signaller and receiver. Second, iconic solutions are possible, and indeed they are found by participants (this is also the case in Galantucci, 2005). Thus in addition to embodiment and the other constraints mentioned above, we also demand that iconicity (and indeed indexicality) be impossible.

This paper introduces the embodied communication game (ECG), an interactive, cooperative two-player game which satisfies these conditions. Pairs of participants must coordinate their behaviour to solve a simple task where they lack shared information, yet they have no interaction with each other except for their movements[2] 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 just 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 the results section 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.

2. The embodied communication game

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 is 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 finished 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 Fig. 1.

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

[1] By *iconic* we mean systems in which the sign bears a resemblance (physical, auditory, etc.) to its referent; *symbolic* systems, in contrast, exhibit arbitrary relationships. As an example of an innate, iconic system we would suggest the aspect of bee dance that refers to the direction of the nectar; and as an example of a learnt, symbolic system we would point to human language.

[2] Our sense of *movement* is actually slightly more broad than just visible movements, and should be construed as 'game moves' which include physical movements and also end-of-turn indicators. These are both embodied in the sense that they are actions required by the player to traverse the space described by the ECG's world.

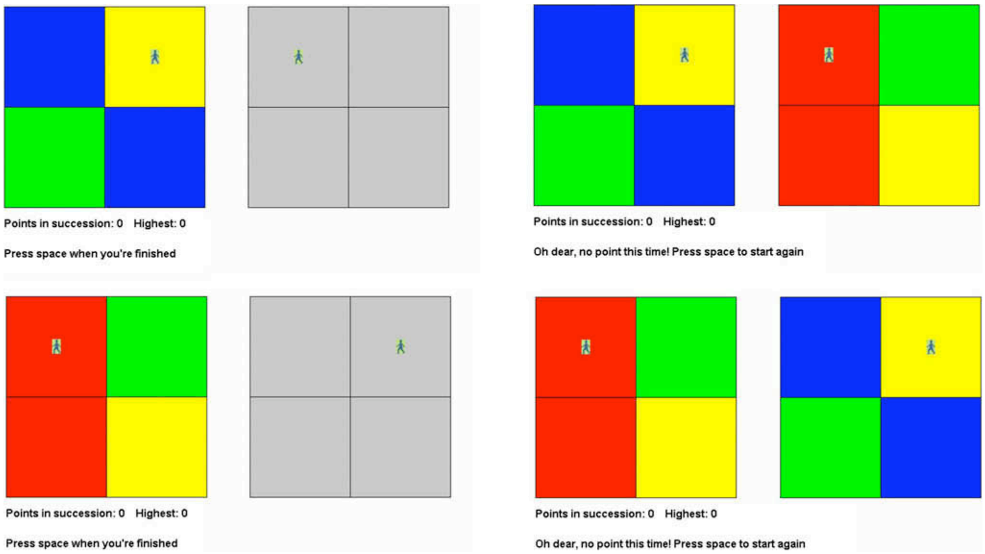

Fig. 1. Screen-shots of the game. Participants play multiple rounds of the game on networked computers. These screen-shots show the view of both players, one on each row, both before (left-hand side) and after (right-hand side) both participants have pressed space to finish their turn. Participants can see their own colours but not the other participants'. 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 to finish their turn, and when they do so all colours are revealed to both participants. Participants score a point if they finish on the same colour. 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). Succeeding at the game requires finding some way to communicate the intended destination colour each round. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

principle possible for the players to score a point. Consequently the initial stages of the ECG, before any communication occurs, can be thought of as a type of Schelling Game (Schelling, 1960). In such games 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. In the case of the ECG the choice is between the colours available, and the prize one point in the game. 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 games played; they must instead find a way to communicate reliably and hence coordinate their behaviour with each other.

The instructions were explicit that the colours would be randomly distributed, since pilots suggested that otherwise participants would look for patterns rather than attempt to communicate. Following basic instructions, which were given in writing, participants were given a 3-min familiarisation period in which to play the game. Further, clarifying instructions, also given in writing, were then given and any queries addressed. Participants then played the game for 40 min uninterrupted. Over the two conditions (described below), pairs played an average of 193.5 rounds of the ECG in the 43 min. At the end of each game subjects were asked about the communication systems they developed or attempted to develop. 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.

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, the communication channel is not 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. They are then (implicitly) charged with the creation of a viable channel. The fact that many pairs failed to communicate with each other at all (see below) shows that to co-opt one's movement for the purpose of communication is no trivial task.

3. Results

3.1. Emergence

Successful pairs typically converged upon a system like that described in Fig. 2, 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 the players score a point. Note that passing through all four colours during dialogue in this way is rare, simply because it is likely that one of the first three suggested colours will be shared. A video of such dialogue using the system described in Fig. 2 is supplied as supplementary information at http://www.lel.ed.ac.uk/~simon/ecg/.

Of interest is the way in which such systems emerge. In debrief interviews most pairs reported that such systems are not created fully-formed by one or the other player. Instead they follow a more organic process, which typically runs as follows. First, the participants choose a default colour to which they will always travel 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[3], or a loop around the entire box; in short, it is something that differentiates it from direct travel to a quadrant, which is what participants do when they have a default colour to travel to (this is discussed further in the section on signalling signalhood, below). Signallers report that this behaviour is designed to mean "No red!", "Not plan A!" or something similar. This behaviour must then be noticed by the other participant. This stage is marked by the other player choosing a colour that is not the default colour, even though the default colour is available to them. The recipient of the signal does not know which colour the signal refers to, but they do recognise that it is a signal, and that all relevant meanings of that signal share one thing in common, namely that the signaller does not have the default colour. A signal has now been established, but it does not yet have fixed meaning. At this point players may choose different colours to each other, 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 "No red!" 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 above. Fig. 3 reports the entire process, none of which is a post-hoc analysis of ours; it is what the participants themselves describe in debriefing interviews after the event.

In all cases participants reported the same story as their partner in terms of (i) whether or not communication was achieved; (ii) the communicative system employed, if any; and (iii) the process by which such a system emerged. This consistency allows us to take the self-reports as reliable, and use them as a guide to breakdown each pair's run according to when they passed through each of the stages described above. A specific sequence of events was defined to be diagnostic of the onset of each stage (for example, the criterion for the establishment of a default colour was that both players choose the same colour for three successive occasions on which it is available). The full details of the number and proportion of rounds played until each stage was reached, the final system employed and other additional details are listed as supplementary information at

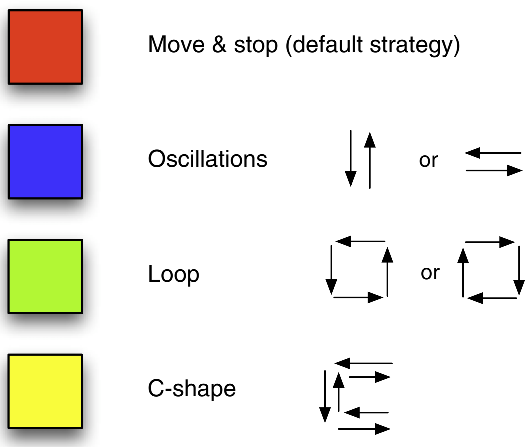

Fig. 2. A typical emergent system. In this communication system red is the default colour. If participants have a red square, they move to it and wait. If they do not have red they will signal one of the other colours by using the movements indicated. If one participant signals a colour that the other 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

[3] It might be suggested that such oscillations could be used as icons, for example to mean "yes" (if they were up-down) or "no" (if they were left-right), reflecting a convention of nodding or shaking one's head accordingly. No players reported this to us, and such behaviours are no more common as signals than any other (see supplementary information at http://www.lel.ed.ac.uk/~simon/ecg/). We therefore think such a use of iconicity was unlikely to have affected the systems in any significant way.

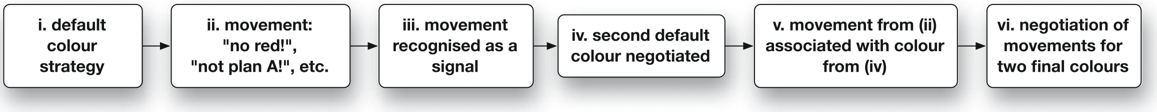

Fig. 3. Stages in the development of successful communication systems. First, in (i), the participants converge upon some shared default colour, usually (in 4 of 5 cases) red. In (ii) one participant performs some movement that would be otherwise unexpected – typically oscillations or circles around the box. This is designed to tell the other participant that this participant does not have the default colour available. This movement must then (iii) be recognised as a signal by the other player. As a result different colours to the default are chosen, and soon (iv) 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 (v), the movement used in (ii) comes to mean, through repeated use, the colour chosen in (iv). Finally, (vi) 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 Fig. 2. This enables them to score in every round and hence build a very high points-in-succession score.

http://www.lel.ed.ac.uk/~simon/ecg/. In addition, Fig. 4 gives a graphical representation of one pair's entire run, and marks the stages at which behaviours associated with each of the colours were developed for that pair.

An important conclusion to take from this initial study is that the final systems that are observed, of which Fig. 2 is representative, are fundamentally affected by the process by which they emerge. They do not, in general, resemble any system that one might invent on one's own (as reported below, systems that are invented by one player on their own take a quite different form, typically associating a number of movements with each colour). On the contrary, they exhibit clear vestiges of the process of emergence, in the form of the default colour. This demonstrates that the problem of how to signal signalhood is not orthogonal to questions of signal form; answers to the former will directly influence answers to the latter.

Initially, 24 participants were assigned into 12 pairs. Despite the fact that all participants were fluent users of a learnt, symbolic communication system, namely natural language, 5 of the 12 pairs reported that they had failed to achieve any communication at all, while 7 did report some success. The accuracy of these self-reports in 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 206.92 rounds with a standard deviation of 108.82, 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).

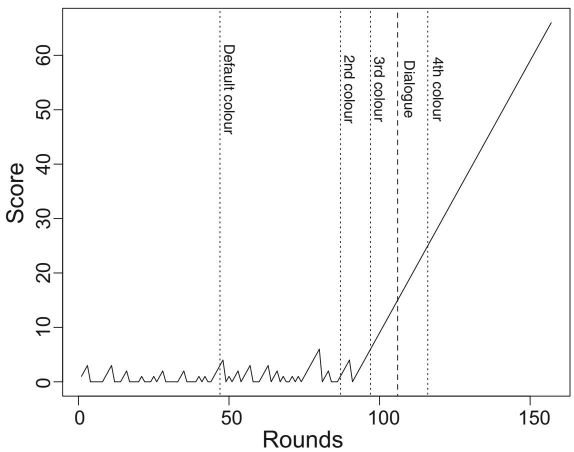

Fig. 4. An example of one pair's progress. Along the *x*-axis is the total number of rounds played and along the *y*-axis the points-in-succession score. As can be seen, initially the pair does not score significantly above chance, but as they establish behaviours for each colour they achieve better points-in-succession scores, eventually hitting upon a full-proof system that is able to score a point in every round.

Of the seven pairs that succeeded, five (final scores 83, 66, 49, 17, and 14) built their system in the way described above, or some close variant of it, although not all pairs actually reached the end of this process. The two other successful pairs (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 these cases the system was created in its entirety by one of the participants who then used it until the other player detected it.

3.2. The importance of initial conventions

It seems, then, that the possibility of creating some initial convention (the default colour) is an aid to the emergence of communication. We tested this hypothesis 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. We predicted that fewer pairs would be able to construct communication systems than did so under the original set-up. This is despite the fact that any 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 the use of any particular system once established.

The players in this condition played an average of 180.08 rounds, with standard deviation 111.02; this is not significantly different from the previous condition ($t_{22} = 0.598$, $p = .556$). Two of the twelve pairs reported success. In one of these (score: 38) the system was fully created by one player and detected by the other. In the other case (score: 14) the process described in Fig. 3 was used: even though the default colour strategy could never score more than one point in succession, that does not mean that it cannot be established, only that it will be unsuccessful in its own right, and thus less likely to emerge. As before, the full

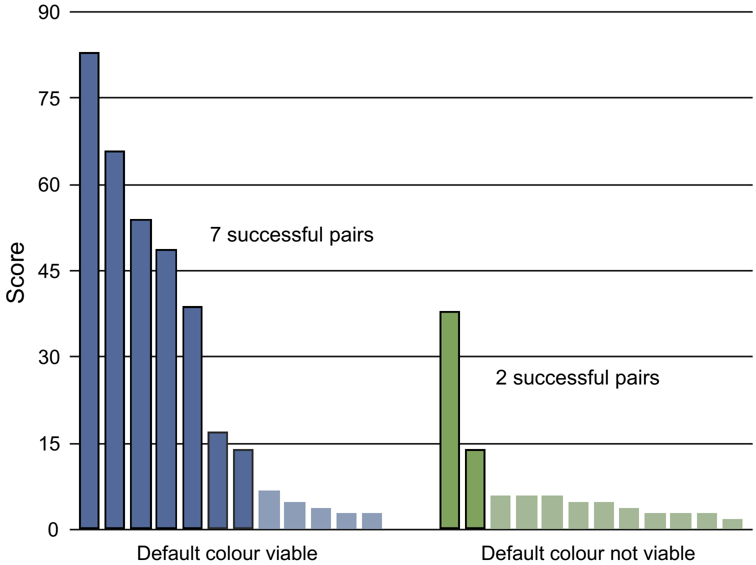

Fig. 5. Comparison of performance between original condition and condition where default colour could not achieve success. Each bar refers to one pair, with their final score on the *y*-axis. The darker-coloured bars are those pairs that reported success; the lighter-coloured ones those that reported failure. The difference between the two conditions is significant both in terms of the number of pairs that achieved success ($\chi^2_1 = 4.44$, $p = .035$) and the average score achieved in each condition ($t_{22} = 2.39$, $p = .026$).

details of the process of emergence can be seen in the supplementary information at http://www.lel.ed.ac.uk/~simon/ecg/. The other ten pairs in this condition reported failure (scores: 6, 6, 6, 5, 5, 4, 3, 3, 3, 2; in all cases $p > .12$ in a Monte Carlo simulation with 10,000 runs). These scores achieved in this condition 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 ($\chi^2_1 = 4.44$, $p = .035$) and the average score of pairs in each condition ($t_{22} = 2.39$, $p = .026$); see Fig. 5. We conclude 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.

3.3. The role of common ground, of dialogue, and of perceptual biases

As discussed, pairs that built their system in the way described by Fig. 3 were able to successfully signal signalhood; this occurs at stage (ii) of Fig. 3. What about the other pairs? Based on the performances of the pairs who are already known to have signalled signalhood, we took any unbroken repetitious movement that is performed in two successive rounds as diagnostic of an attempt to signal signalhood. This property is shared by all the instances of signalling signalhood in the successful pairs except one[4], and is otherwise not seen in those trials. This diagnosis is complicated somewhat by the fact that in some cases players used repetitive movement in an iconic way, to illustrate that they had the same colour on each of the quadrants that they moved between (information that is, of course, useless). This is therefore not necessarily an attempt to signal signalhood – the form of the signal is entirely a function of its meaning. Pairs in which only this type of repetitious movement was observed were therefore not counted as signalling signalhood ($n = 3$). Additionally, four pairs showed no repetitious movement at all. This leaves 17 (of 24) pairs that did attempt to signal signalhood.

We diagnose the detection of signalhood as the subsequent establishment of any form of communication system that makes use of the movement used to signal signalhood. It is possible that a player could detect signalhood but fail to attach the correct meaning to it, but this should not reoccur over and over, since the players can converge upon a shared meaning for the signal through simple trial and error. Therefore, the establishment of a system that uses that signal will necessarily follow the detection of signalhood. Of the 17 pairs that signalled signalhood, nine satisfied this criterion. Details can be found with the supplementary information at http://www.lel.ed.ac.uk/~simon/ecg/.

The pairs that signalled signalhood can be divided into those that did so within the context of a default colour strategy, and those that did not. The first group has six pairs, and in all cases signalhood was detected, and a communication system subsequently built. The second group has 11 pairs, but in only three of these was signalhood detected and a communication system built. The difference between these two groups is significant ($\chi^2_1 = 8.24$, $p = .009$), and can be explained by the importance of context and common ground. With the default colour strategy

[4] The pair that did not do this instead used a circuitous movement to their final destination (i.e. up-left-down, rather than just left on its own). However this irrationality cannot be used to diagnose signalhood, since it is a common feature of the essentially aimless background wandering that is typical of participants that are yet to work out how they might succeed at the ECG. It worked for this pair because it occurred within the context of an established default colour strategy.

both participants expect the other player to travel directly to whichever quadrant has that colour, and so deviations from this are taken to be meaningful. Without this expectation signalhood is harder to detect.

What else determines a pair's final score? Obviously they need to develop some conventional behaviours, but these alone are insufficient to achieve a high score. With only a default colour, for example, high scores are not possible since sooner or later one player or the other will not have the default colour. Nor is it sufficient to have a behaviour that is performed in the absence of a default colour: whilst an alternative colour can be suggested, there is no way to know if the other player has that suggested colour. This point in fact applies to all movements that take on a communicative function; it is only with feedback that a participant can know whether the suggested colour actually appears in the other player's box. However once such feedback is in place and is accompanied by a shared convention for each colour, success is assured (subject to human error). We therefore predict that the onset of such feedback will correlate with a pair's final score. We define dialogue as those occasions when one player performs an established movement to indicate a particular colour (but not the default colour) and the other player, not having that colour in their box, then performs a different established movement, which refers to a different colour (or, in one case, pair A10, to "no – I do not have that colour"). We recorded the proportion of the rounds elapsed before each successful pair displayed such behaviour, and plotted this against their final score. The resultant correlation is significant ($r = -.753$, $p = .019$)[5]; see Fig. 6. We conclude that it is only after dialogue has been established that a pair can be considered to have developed a complete, full-proof system that, so long as concentrations errors are avoided, is guaranteed to succeed. The importance of dialogue in the emergence of communication, well-established in the psycholinguistic literature (Anderson et al., 1991; Garrod et al., 2007; Healey et al., 2007) is thus reinforced.

It may also be of interest that our participants appear to exhibit a preference for red over other colours. For example, in all three pairs where a system was imposed unilaterally by one of the participants, the colour with the fewest number of moves associated with it was red. Similarly, of the six pairs that used a default colour, three used red, two blue, one green and none yellow. In all, red is chosen as the initial colour significantly more often than by chance ($t_1 = 2.50$, $p = .037$). This can be explained by the existence of a human perceptual bias for red (see Fernandez & Morris, 2007; Teller & Bornstein, 1987). We suggest that such a bias can act as common ground within the ECG, and thus influence the emergence of communication systems. It increases the likelihood that, when a new movement is used, the two players will converge upon the same colour as the meaning to be associated with that movement. The speed of emergence is thus increased. Of course, if all the colours carried equal perceptual salience then the task would still be possible, but more negotiation may be needed to agree on the meaning of particular movements. Furthermore, the very fact that such negotiation is possible is in significant part an artefact of the constrained nature of the task. In a more open-ended environment individuals would have far fewer constraints on possible meanings, and the role of shared sensory preferences and other factors that contribute to common ground would likely be greater.

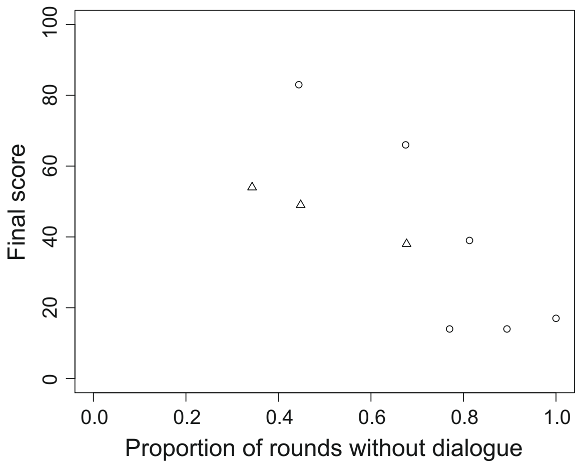

Fig. 6. The importance of dialogue. The *x*-axis measures each pair's final points-in-succession score, and the *y*-axis the proportion of the total number of rounds that have passed when a pair establishes dialogue. Circles refer to those pairs that built their system in the gradual way described in Fig. 3; triangles to those pairs in which one player invented a system that was later detected by the other participant. A negative correlation is observed ($r = -.753$, $p = .019$). This suggests that it is only once dialogue has been established that a pair can be considered to have developed a complete, full-proof system that, so long as concentrations errors are avoided, is guaranteed to succeed.

4. Discussion

Recent years have witnessed a growing interest in the emergence of communication. One particularly interesting strand of that development is the use of experiments using human participants who are charged with the creation of a novel communication system (de Ruiter et al., 2007; Galantucci, 2005; Garrod et al., 2007; Healey et al., 2007). Of these, (de Ruiter et al., 2007) is closest to the present paper in spirit: it too is focused on the origins of communicate intent (although that paper offers a mostly neurological perspective), and the approach used demands, as the ECG does, that there be no *a priori* distinction between movements for travel and movements for communication. However, there are, as discussed in the introduction, two differences between that study and our own: first, that participants are pre-defined to be either signallers or receivers; and second, that iconic solutions are possible. We suggest these differences, and in particular the near-impossibility of iconicity in the ECG (see footnote 3), explain the large disparity between the two studies in terms of the number of pairs of participants that were successful.

[5] Two pairs reported a loss in concentration after the entire system had been established (see supplementary information at http://www.lel.ed.ac.uk/~simon/ecg/). If their scores are adjusted to ignore those losses in concentration (pair A10 would have scored 139; and pair A11 would have scored 55; see supplementary information at http://www.lel.ed.ac.uk/~simon/ecg/) then the correlation remains significant ($r = .854$, $p = .003$).

The current paper thus adds to this developing literature, and reiterates some of the conclusions that are beginning to emerge. For example, interaction is found to be crucial to the evolution of graphical sign systems (Garrod et al., 2007) and is suggested to be an important constraint on representational form (Healey et al., 2007). The current paper not only finds that interaction is integral to the process of emergence (see Fig. 3), but also that that the emergence of a particular form of interaction, dialogue, may be the hallmark of when a communication system is sufficiently developed to be useful in the coordination of behaviour in a changing environment.

The importance of common ground, emphasised in the theoretical literature (Clark, 1996; Lewis, 1969; Schelling, 1960), is also reinforced. This is most obvious in the process of emergence described in Fig. 3, in which the establishment of the default convention provides the common ground from which a signal may be created and inferred. As already observed, the ECG is a type of Schelling Game, and common ground is necessary to succeed at above chance levels in such games (Clark, 1996). In the example given earlier, in which two players are asked to pick either a basketball, a football, and a squash ball, the players may happen to be squash partners. 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 establishment of the default colour strategy provides just such assumptions, since both players have an expectation that the other will travel directly to whichever quadrant has the default colour. Furthermore, universal colour preferences may provide additional common ground that speeds the process of emergence in the ECG.

In addition to these commonalities with the existing literature, this paper makes what we think is an important theoretical point: that study of the emergence of communication must take seriously the question of how individuals recognise that a given behaviour is communicative. Our findings suggest that this question is not orthogonal to those raised by other studies. Instead, the constraint of embodiment is observed to fundamentally affect the form of the final communication system. 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, Perkins, & Pagliuca, 1994). Since the final systems observed in the ECG do not generally correspond to any obvious solution that might be designed beforehand, our work suggests that the same may be true of emergence. This conclusion could not be reached if communication was not an embodied act, and we therefore suggest that future work on the emergence of communication systems will be most productive when the constraint of embodiment is met.

Acknowledgements

We wish to thank three anonymous reviewers for valuable comments on a previous version of this paper. TSP was funded by the AHRC; and GR by the ESRC.

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.

Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar*. Chicago: University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge 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 and performance XXII* (pp. 593–610). Oxford: Oxford University Press.

Deacon, T. (1997). *The symbolic species*. London: Penguin.

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).

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.

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. (1971). Meaning. In D. Steinburg & L. Jakobovits (Eds.), *Semantics: An interdisciplinary reader* (pp. 53–59). Cambridge: Cambridge University 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.

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

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

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.

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.

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.

Schelling, T. C. (1960). *The strategy of conflict*. Oxford: Oxford 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.

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

Teller, D. Y., & Bornstein, M. H. (1987). Infant color vision and color perception. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception. From sensation to perception* (Vol. 1, pp. 185–236). Orlando, Florida: Academic Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review, 9*(4), 625–636.