Evolved Signals: Expensive Hype vs. Conspiratorial Whispers

Jason Noble
School of Cognitive and Computing Sciences
University of Sussex
Brighton BN1 9QH, U.K.
jasonn@cogs.susx.ac.uk

Abstract

Artificial life models of the evolution of communication have usually assumed either cooperative or competitive contexts. This paper presents a general model that covers signalling with and without conflicts of interest between signallers and receivers. Krebs & Dawkins (1984) argued that a conflict of interests will lead to an evolutionary arms race between manipulative signallers and sceptical receivers, resulting in ever more costly signals; whereas common interests will lead to cheap signals or "conspiratorial whispers". Simple game-theoretic and evolutionary simulation models suggest that signalling will evolve only if it is in the interests of both parties. In a model where signallers may inform receivers as to the value of a binary random variable, if signalling is favoured at all, then signallers will always use the cheapest and the second-cheapest signal available. Costly signalling arms races do not get started. A more complex evolutionary simulation was constructed, featuring continuously variable signal strengths and reception thresholds. As the congruence of interests between the parties became more clear-cut, the evolution of successively cheaper signals was observed. The findings are taken to support a modified version of Krebs & Dawkins's argument.

Artificial life models of communication

Artificial life (AL) models of the evolution of communication are often constructed such that honest signalling is in the interests of both signallers and receivers—any communication systems that evolve can therefore be described as cooperative. For example, Werner & Dyer (1992) postulated blind, mobile males and sighted, immobile females: the evolution of a signalling system was in the interests of both parties as it allowed mating to take place at better-than-chance frequencies. In MacLennan & Burghardt's (1994) model, signallers and receivers were rewarded if and only if they engaged in successful communicative interactions.

Other AL models (Ackley & Littman 1994; Oliphant 1996) have looked at the special case where communication would benefit receivers, but the potential signallers are indifferent. Oliphant argues that this is a good way to model the evolution of alarm calls, for example: if one bird in a flock spots an approaching hawk, it is clear that its conspecifics would benefit from an alarm call. However, why should the bird in question, considered as a product of its selfish genes, give the call? The models suggest that signalling will not evolve in these cases unless a mechanism such as reciprocal altruism or (spatially induced) kin selection is in place. Note that such mechanisms have no mystical effect: they simply shift the expected fitness payoffs for particular strategies such that communication is mutually beneficial.

Finally, some AL work considers the evolution of communication in situations where the two parties appear to have conflicting interests. Wheeler & de Bourcier (1995) modelled aggressive territorial signalling. Bullock (1997) constructed a general model in which signallers of varying degrees of quality solicited receivers for a favourable response; receivers were rewarded for responding positively only to high-quality signallers. A conclusion drawn in both studies was that if signals were sufficiently costly (e.g., long, elaborate tails or energetic ritual displays) then reliable communication could evolve and persist over time. Bullock made the more specific prediction that in order for communication to be stable, the net cost of signalling must be lower for higher-quality signallers (see also Grafen, 1990). However, it could be argued that such differential signal costs effectively render honest signalling mutually beneficial. We will return to this notion below.

One goal of the current paper is to position previous AL work in an overarching theoretical context. To this end some general models of the evolution of simple signalling systems will be presented; the models will cover situations with and without a conflict of interests between the two interacting agents.

Manipulative and cooperative signals

Krebs & Dawkins (1984) discuss the behavioural ecology of animal signals—they view signalling as a typically competitive affair involving mind-reading and manipulation. Mind-reading consists of one animal exploiting tell-tale predictors about the future behaviour of another, e.g., a dog noticing the bared teeth of an opponent, concluding that it is about to attack, and fleeing in order to avoid injury. Manipulation is what happens when those being mind-read fight back, influencing the behaviour of the mind-readers to their own advantage. For example, a dog could bare its teeth despite not having the strength or inclination to attack, and thus scare off its mind-reading opponent. The authors predict evolutionary arms races between manipulative signallers and sceptical receivers: "selection will act simultaneously to increase the power of manipulators *and* to increase resistance to it" (p. 390). The result will be increasingly costly signals.

Krebs & Dawkins admit, however, that not all interactions are competitive in nature. They suggest that when the reliable transmission of information is to the benefit of both parties (e.g., bee dances indicating the location of nectar), a different kind of signal co-evolution will result. Specifically, there will be selection for signals that are as cheap as possible while still being detectable: "conspiratorial whispers".

Krebs & Dawkins's argument has been influential but no formal justification (i.e., model) of it exists. A second goal of the current paper is to test their prediction that evolved signals will necessarily be more costly when there is a conflict of interests than when the participants have common interests. In order to do so, it will be necessary to determine whether communication should be expected *at all* when signallers and receivers have a genuine conflict of interests.

Conflicts of interest

The first requirement in constructing a general model of communication is a classification scheme for determining when a conflict of interests exists between signallers and receivers—Figure 1 shows such a scheme, adapted from Hamilton (1964). Assume that a successful instance of communication in a particular scenario has fitness implications for both participants. The fitness effect on signallers, P_S, and the fitness effect on receivers, P_R, together define a point on the plane in Figure 1. For example, consider a hypothetical food call, by which one animal alerts another to the presence of a rich but limited food source. By calling and thus sharing the food, the signaller incurs a fitness cost; by responding to the call, the receiver benefits through obtaining food it would otherwise have missed. Thus, the call would be located in the "altruism" quadrant. The situations modelled by Ackley & Littman (1994) and Oliphant (1996), where receivers benefit but signallers are ambivalent, can be thought of as points on the positive vertical axis, i.e., where $P_S = 0$ and $P_R > 0$.

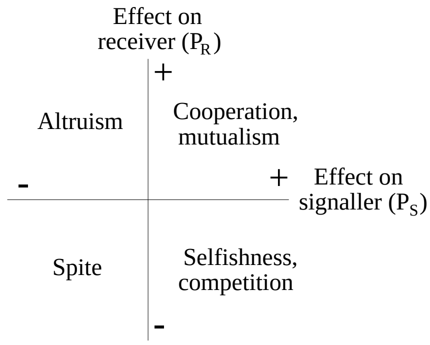

Figure 1: Possible communication scenarios classified by their effects on the fitness of each participant.

Conflicts of interest can be defined as interactions in which natural selection favours different outcomes for each participant (Trivers 1974), or in which participants place the possible outcomes in a different rank order (Maynard Smith & Harper 1995). Conflicts of interest therefore exist when P_S and P_R are of opposite sign, i.e., in the upper-left and lower-right quadrants. Selection will, by definition, favour actions that have positive fitness effects. In the upper-left and lower-right quadrants, one agent[1] but not the other will be selected to participate in the communication system: their interests conflict. The "spite" quadrant does *not* represent a conflict of interests because agents will be mutually selected not to communicate.

If the specified fitness effects of participating in a communicative interaction are truly *net* values, and already include such factors as the cost of signalling and the cost of making a response (as well as inclusive fitness considerations and costs due to exploitation of the signal by predators, etc.), then predicting the evolution of the communication system is trivial. Reliable communication requires, on average, honest signallers and trusting receivers, and thus will only develop when $P_S > 0$ and $P_R > 0$, i.e., when both agents are selected to participate. However, real animals sometimes communicate despite apparent conflicts of interest (Hinde 1981). Recent models (Grafen 1990; Bullock 1997) have established that, in certain situations where communication would otherwise be unstable, increasing the production costs of the signal can

[1] The term "agent" is used to refer to an entity that may be playing a signalling or a receiving role.

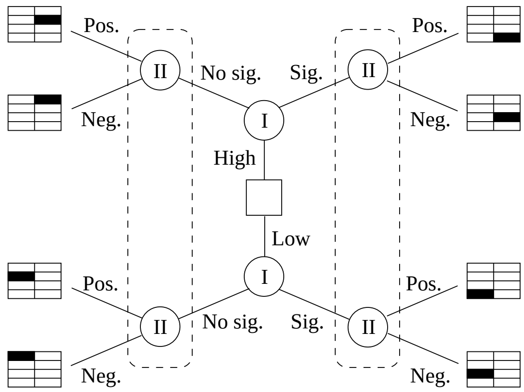

Figure 2: Extended form of the simple signalling game. Chart icons index payoffs in Table 1.

	State of environment	
	Low	High
No signal		
Neg. response	0 , 0	0 , 0
Pos. response	0 , $-C_R$	P_S , $P_R - C_R$
Signal		
Neg. response	$-C_S$, 0	$-C_S$, 0
Pos. response	$-C_S$, $-C_R$	$P_S - C_S$, $P_R - C_R$

Table 1: Payoff matrix for the simple game. Entries in the table represent the payoff to the sender and receiver respectively.

lead to a prediction of evolutionarily stable signalling. Therefore, in the current model, P_S and P_R refer to gross fitness effects before the specific costs of producing the signal, C_S, and making the response, C_R, have been taken into account.

A simple signalling game

If the signalling interaction is to involve information transmission, and allow for the possibilities of deception and manipulation, it must be modelled as a game of imperfect information, in which the signaller knows something that the receiver does not. Figure 2 shows the extended form of a simple action-response game that fulfils this requirement. The game begins with a chance move (the central square) in which some state is randomly determined to be either "high" or "low". The signaller has access to this state, and we can suppose that it represents either a feature of the environment that only the signaller has detected (e.g., noticing an approaching predator), or a hidden internal state of the signaller (e.g., ovulation). Based on this state, the signaller (player I) must decide whether or not to send an arbitrary signal of cost C_S. The receiver (player II) is ignorant of the hidden state and only knows whether or not a signal was sent—the dashed rectangles show the receiver's information sets. The receiver can respond either positively, i.e., perform some action "appropriate" to the high state, or negatively, i.e., not respond at all. Positive responses incur a cost, C_R. If and only if the hidden state is high, a positive response results in the payoffs P_S and P_R to the signaller and receiver respectively. Table 1 specifies the payoff matrix. Hurd (1995), Oliphant (1996), and Bullock (1997) used similar games with different payoff structures.

The game models a range of possible communicative interactions. For example, suppose that the high state represents the signaller's discovery of food. Sending a signal might involve emitting a characteristic sound, while not sending a signal is to remain silent. For the receiver, a positive response means approaching the signaller and sharing the food, whereas a negative response means doing nothing. Various possibilities exist besides honest signalling of the high state: the receiver might *always* approach the signaller in the hope of obtaining food, regardless of whether a signal was sent. The signaller might be uninformative and never signal, or only signal when food was *not* present. One important feature of the game is that the signaller is ambivalent about the receiver's response in the low state—in terms of the example, this represents the assumption that when no food has been discovered, the signalling animal does not care about whether the receiver approaches or not.

The strategies favoured at any one time will depend on the relative values of P_S, P_R, C_S and C_R, as well as on what the other members of the population are doing.[2] Allowing the base fitness effects P_S and P_R to vary across positive and negative values will allow the payoff space of Figure 1 to be explored, and thus determine whether changes in signal and response cost can produce stable signalling in situations that would otherwise involve conflicts of interest. This will be a first step towards assessing Krebs & Dawkins's conspiratorial whispers theory.

Stable strategies in the simple game

A signalling strategy in the simple game specifies whether to respond with no signal (NS) or a signal (Sig) to low and high states respectively. Likewise, a response strategy specifies whether to respond negatively (Neg) or positively (Pos) when faced with no

[2] Another parameter of interest in the signalling game is the relative frequency of high and low states; in the models presented here each state occurred 50% of the time.

signal and when faced with a signal. A strategy pair is the conjunction of a signalling and a response strategy; e.g., (NS/NS, Pos/Pos) is the strategy pair that specifies never signalling and always responding positively.

The strategy pair (NS/Sig, Neg/Pos) specifies signalling only in the high state, and responding positively only to signals—call this the honest strategy. It can be shown that honesty will be an evolutionarily stable strategy (ESS; Maynard Smith 1982) if:

$$P_S > C_S > 0$$
$$P_R > C_R > 0.$$

That is, honest signalling is stable if the costs of signalling and responding are both positive, and if the payoffs in each case outweigh the costs. The requirement that P_S and P_R must both be positive means that the honest strategy is only expected to be stable when the interests of the parties do not conflict.

Of the 16 possible strategy pairs, there are three besides the honest strategy that involve the transmission of information, in that the receiver responds differently to different hidden states. None of these three strategy pairs are ESSs if C_S and C_R are both positive; these two values represent energetic costs and so cannot sensibly be negative. If $C_S = 0$, i.e., if giving a signal is of negligible cost, then the reverse honesty strategy (Sig/NS, Pos/Neg) can be stable, although P_S and P_R must still be positive. It is also worth noting that any mixed strategy involving (NS/NS, Pos/Pos) and (NS/NS, Pos/Neg), both non-signalling strategies where the receiver always responds positively, can be an ESS if the payoff to the receiver is large enough, i.e., if:

$$C_S > 0$$
$$P_S > -C_S$$
$$P_R > 2C_R > 0.$$

The analysis indicates that while the cost of signalling plays some role in stabilizing the honest strategy, there are no circumstances in which stable communication is predicted when a conflict of interests exists. This is despite the fact that we have separated the costs of signalling and responding from the base fitness payoffs of a communicative interaction.

Evolutionary simulation model

Game theory is limited to describing equilibria; an evolutionary simulation model of the simple game was also constructed in order to determine whether communicative behaviour might sometimes be found outside the range of identified ESSs.

A straightforward genetic algorithm (GA) was used. Each individual could play both signalling and receiving roles; a strategy pair was specified by a four-bit genotype as shown in table 2. The population size was 100, the mutation rate was 0.01 per locus, and, due to the trivially small genome, crossover was not used. Each generation, 500 games were played between randomly selected opponents. An agent could therefore expect to play 5 games as a signaller and 5 as a receiver. The agent's fitness score was the total payoff from these games. For breeding purposes, the fitness scores were normalized by subtracting the minimum score from each. Proportionate selection was then applied to the normalized scores. The genetic algorithm was run in this manner for 500 generations. In the results presented below, the games played in the final, i.e., 500th, generation have been used as a snapshot of the evolved signalling strategies.

	Bit value	
	0	1
If low state...	No signal	Signal
If high state...	No signal	Signal
Response to no signal	Negative	Positive
Response to signal	Negative	Positive

Table 2: Genetic specification of strategies.

An attempt was made to investigate evolutionary dynamics, in that the initial populations were not determined randomly but started as either "honest" or "non-signalling". Honest initial populations were made up entirely of individuals who played the honest strategy, i.e., a genome of '0101'. Non-signalling populations underwent 100 generations of preliminary evolution in which their receiving strategies were free to evolve but their signalling strategies were clamped at '00', i.e., no signalling. For each class of initial conditions, a simulation run was performed for all combinations of integer values of P_S and P_R between -5 and +5, making 121 runs in all. Each run was repeated 25 times with different random seeds. The values of C_S and C_R were fixed at 1.

Communication was indexed by cross-tabulating the hidden state value with the receiver's response and calculating a chi-squared statistic. The receiver has no direct access to the hidden state, so any reliable correspondence between state and response indicates that information has been transmitted and acted upon. Values of the χ^2 statistic close to zero indicate no communication, and values close to the maximum (in this case $\chi^2_{max} = 500$, due to the 500 games played in the final, snapshot generation) indicate near-perfect communication.

Figure 3 shows the average values of the commu-

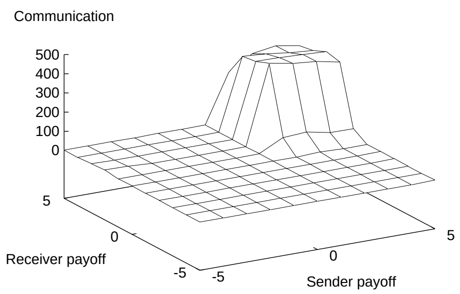

Figure 3: Mean communication index by P_S and P_R; honest initial conditions. Each point is a mean calculated over 25 runs. Mean standard error = 2.96.

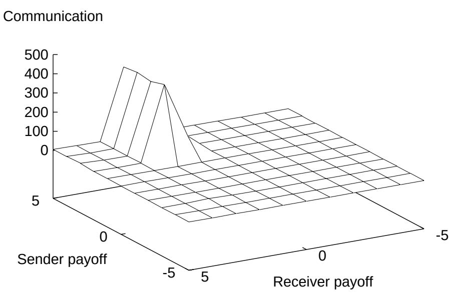

Figure 4: Mean communication index by P_S and P_R; non-signalling initial conditions. Each point is a mean calculated over 25 runs. Mean standard error = 2.75. Graph rotated for clarity.

nication index for honest initial conditions. Seeding the population with honesty tests the stability of honest signalling given a particular payoff pair, much as a game-theoretic analysis does. The results are compatible with the conditions outlined in the previous section: honesty is stable when the payoffs to signalling and receiving are positive and greater than the respective costs. However, there is some suggestion of intermittent or imperfect communication when $P_R = C_R = 1$, indicating that ambivalent receivers may occasionally cooperate.

Figure 4 shows the average values of the communication index for non-signalling initial conditions. Starting the GA with a non-signalling population tests the likelihood that communication will emerge, given a particular payoff pair. Clearly the conditions for emergence and stability-once-present are not the same. If $P_S > 1$ and $P_R = 2$ communication develops but when $P_S > 1$ and $P_R > 2$ it does not. In the latter region $P_R > 2C_R$ and the population remains at the non-signalling ESS described in the previous section. Despite the fact that communication would result in a higher average fitness, the high value of P_R keeps the receivers responding positively all the time, removing any incentive for the signallers to bother signalling.

The difference in results between the two classes of initial conditions is interesting, but should not obscure the fact that no communication was observed under conditions of conflicting interests. We must conclude that, at least in the simple model discussed so far, stable communication is only to be expected when it is in the interests of both parties.

A game with variable signal costs

In the simple signalling game, signallers can choose between a costly signal or no signal at all. The model does not allow for a range of possible signals with differing costs, and in this respect it is unrealistic. It may be that Krebs & Dawkins's implicit prediction, that signalling can occur when a conflict of interests exists, is in fact true, but can only be demonstrated in a more complex game with a range of signal costs. The simple signalling game (see Figure 2) was therefore extended to incorporate signals of differing costs.

In the extended game, the signalling player has three options: not signalling, which costs nothing; using the "soft" signal, which costs C_S, and using the "loud" signal, which costs $2C_S$. Strategies in the extended game require specifying the signal to give when the hidden state is low, the signal to give when it is high, and the response to give to each of no-signal, soft and loud. The two strategies representing conspiratorial whispers or cheap signalling are (NS/Soft, Neg/Pos/Pos) and (NS/Soft, Neg/Pos/Neg). Both strategies call for the soft signal to be used in the high state, and for positive responses to the soft signal; the strategies differ only in the response to loud signals. Neither of these strategies can strictly be considered an ESS on its own (because neutral drift can take the population from one to the other) but it can be shown that the set of all mixed strategies involving these two is an ESS under the familiar conditions:

$$P_S > C_S > 0$$
$$P_R > C_R > 0.$$

Costly signalling would involve the use of the loud signal for the high state, and either the soft signal or no signal to denote the low state, with a corresponding response strategy. None of the four strategies in this category can be an ESS. For example, (NS/Loud, Neg/Pos/Pos) cannot be an ESS assuming positive costs of signalling and responding. The similar strategy (NS/Loud, Neg/Neg/Pos) is almost stable if $P_S > 2C_S$, but can drift back to the previous strategy which can in turn be invaded by the cheap strategy (NS/Soft, Neg/Pos/Pos).

Analysis of the extended game indicates that if signalling is favoured at all, then at equilibrium the signallers will always use the cheapest and the second-cheapest signal available (i.e., no signal and the soft signal). Further extensions of the game, by adding ever more costly signalling options, do not alter this conclusion. None of the costly signalling strategies can even be an ESS, let alone support communication in the face of a conflict of interests. The possibility of expensive signalling arms races starts to look remote. However, it may be that the discrete signals used in the games presented so far have had an unwarranted effect on the results. Certainly discrete and continuous models of the same biological phenomenon can lead to different conclusions—compare Maynard Smith (1991) and Johnstone & Grafen (1992).

Simulation model with continuous signal costs and reception threshold

A second evolutionary simulation was constructed, in which the cost of signalling was continuously variable. Signalling strategies were represented by two positive real numbers C_{low} and C_{high}: the cost of the signals given in the low state and in the high state respectively. Response strategies were represented by a real-valued threshold T; positive responses were given to signals with costs greater than the receiver's threshold value. Note that threshold value could be negative, indicating a positive response to any signal.

A real-valued GA was used to simulate the evolution of strategies over time. Generally, the same parameters were used as in the previous simulation model, e.g., a population of 100. Mutation was necessarily a different matter: each real-valued gene in each newborn individual was always perturbed by a random gaussian value, $\mu = 0$, $\sigma = 0.05$. If a perturbation resulted in a negative cost value the result was replaced by zero. In addition, 1% of the time (i.e., a mutation rate of 0.01) a gene would be randomly set to a value between 0 and 5 for signal costs, or between -5 and +5 for the threshold value. This two-part mutation regime ensured that offspring were always slightly different from their parent, and occasionally very different.

The C_S parameter was no longer relevant, but C_R, the cost of responding, remained fixed at 1. Honest initial conditions were implemented by setting $C_{low} = 0$, $C_{high} = 1.0$ and $T = 0.5$. Non-signalling initial conditions were implemented by setting T to a random

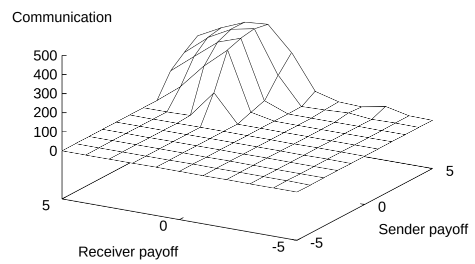

Figure 5: Mean communication index by P_S and P_R in the continuous simulation; honest initial conditions. Each point is a mean calculated over 25 runs. Mean standard error = 3.54. Graph rotated for clarity.

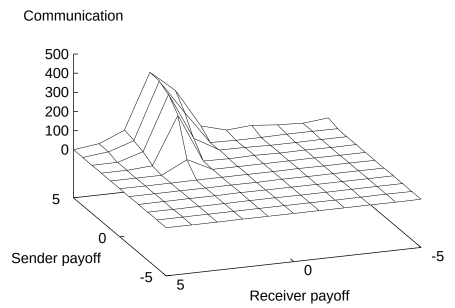

Figure 6: Mean communication index by P_S and P_R in the continuous simulation; non-signalling initial conditions. Each point is a mean calculated over 25 runs. Mean standard error = 2.81. Graph rotated for clarity.

gaussian ($\mu = 0$, $\sigma = 1$) and then clamping $C_{low} = C_{high} = 0$ for 100 generations of preliminary evolution.

The use of continuous values immediately suggests the possibility of random noise, and in trial experiments gaussian noise was added to both the signalling channel (i.e., to the signal's cost value before it was "perceived" by the receiver) and to the payoff values P_S and P_R. It was felt that these measures might introduce some realistic uncertainty to the game. However, the results below were found to be robust with respect to the presence of noise; results from noise-free runs only are reported.

Figures 5 and 6 show the average values of the communication index for honest and non-signalling initial conditions respectively. The results are qualitatively similar to those of the discrete simulation model: com-

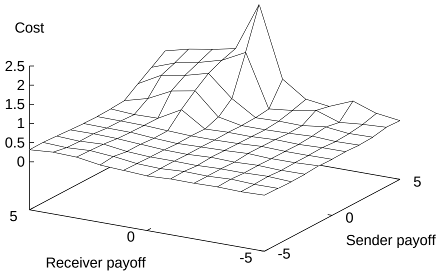

Figure 7: Mean cost of high-state signals by P_S and P_R; honest initial conditions. Each point is a mean calculated over 25 runs. Mean standard error = 0.032. Graph rotated for clarity.

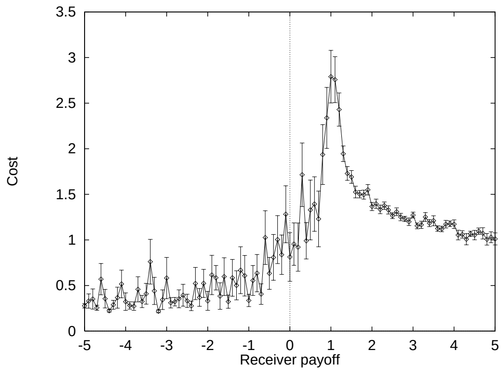

Figure 8: Cross-sectional means (± 1 s.e.) for high-state signal costs with $P_S = 5$; honest initial conditions. Each point is a mean calculated over 25 runs.

munication occurs in both cases, but in a more limited range of the payoff space for non-signalling conditions. In neither case does communication occur outside the "cooperative" quadrant.

The continuous model also allows investigation of the cost and threshold values over the payoff space. C_{low}, the cost of the signal given in response to the low state, always remained close to zero—this was unsurprising as signallers are ambivalent about the receiver's response to the low state. However, the value of C_{high} varied both inside and outside the region where communication was established: Figure 7 shows the mean values of C_{high} for honest initial conditions. The signals given in response to the high state are most costly when P_S, the payoff to the sender, is high and when the receiver's net payoff is marginal, i.e., $P_R \approx 1$. In order to study this effect more closely, additional simulation runs were performed, with P_S fixed at 5 and P_R varied between -5 and +5 in increments of 0.1. These runs can be thought of as exploring the cross section through $P_S = 5$ in Figure 7. Figure 8 shows the cross-sectional mean values of C_{high}. Note that the "energy" devoted to signalling is at a maximum around $P_R = 1$ and drops off as P_R increases—it can be seen from Figure 5 that $P_R = 1$ is approximately the point where significant communication is established. The same pattern was observed for non-signalling initial conditions (not shown for reasons of space).

The threshold values showed corresponding variation. Figure 9 shows the mean value of T across the payoff space. The threshold values are typically very high (a "never respond" strategy) or very low (an "always respond" strategy), but in the region where communication evolved, receivers become progressively less

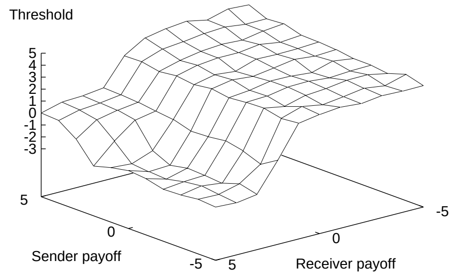

Figure 9: Mean threshold value by P_S and P_R; honest initial conditions. Each point is a mean calculated over 25 runs. Mean standard error = 0.18. Graph rotated for clarity.

demanding, i.e., T gets lower, as P_R increases. Figure 10 shows the cross-sectional results for $P_S = 5$.

Figure 11 plots the mean cost of high and low signals and the mean reception threshold all on one graph. This makes the relationship between costs and threshold clear: at approximately $P_R = 1$, the threshold falls to a level where the mean high-state signal will generate a positive response. As P_R increases, i.e., as the two players' payoffs approach each other, the signallers become less extravagant and the receivers less "sceptical". This is *contra* the game-theoretic result of the previous section, which implies that when signals of varying costs are available, either the cheapest pair of signals will be used, or no signalling will occur—something like Figure 12 would be expected if the soft-loud signalling game accurately modelled the continuous case.

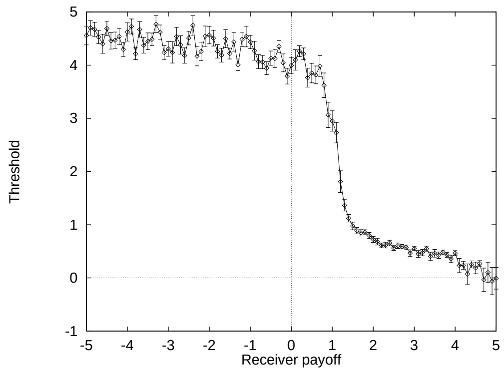

Figure 10: Cross-sectional mean threshold values (± 1 s.e.) with $P_S = 5$; honest initial conditions. Each point is a mean calculated over 25 runs.

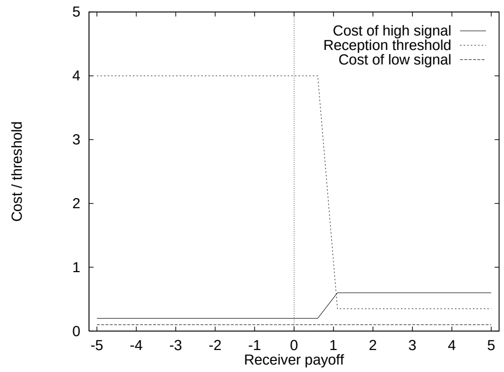

Figure 12: Approximate predicted results for Figure 11 according to discrete-cost game-theoretic model.

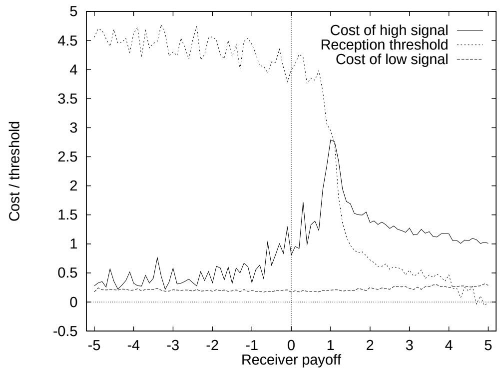

Figure 11: Cross-sectional means: cost of high and low signals, and reception threshold. $P_S = 5$, honest initial conditions. Each point is a mean calculated over 25 runs.

Note that the initial values of C_{high} and T under honest initial conditions were 1.0 and 0.5 respectively. For all but the highest values of P_R, C_{high} has increased on average over the 500-generation run. This rules out any explanation of the results of Figure 11 in terms of there having been insufficient evolutionary time for a cheaper signalling equilibrium to have been reached when the profit for receivers ($P_R - C_R$) was marginal. Evolution has taken the populations *away* from the cheap signalling solution.

Discussion

In all of the models presented, communication evolved or was predicted to evolve only within the cooperative region of the signaller-receiver payoff space. This means that no signalling at all (costly or otherwise) was observed when the signaller and the receiver were experiencing a conflict of interests. The second game-theoretic model, in which discrete signals of varying costs are available, suggests that communication, if selected for, will involve the cheapest pair of signals available. However, the second simulation model, incorporating the more realistic assumption that signals can vary continuously in cost, implies that cheap signals will only be used when both parties stand to gain a high payoff from effective communication. When the net payoff to the receiver is marginal, evolved signals will be more costly than strictly necessary to convey the information. The relationship is not symmetrical: when the net payoff to the signaller is marginal, a non-signalling equilibrium, in which the receiver always responds positively, is likely to occur.

Krebs & Dawkins (1984) predicted that signalling would be costly if a conflict of interests existed; strictly speaking the results do not support nor contradict their prediction, as no signalling occurred in the conflict-of-interest cases. It might be the case that conflicts of interest in the context of a different signalling game would indeed result in costly signals. However, it will be argued below that the simple signalling game used in the current models is plausible, and thus the failure to evolve communication given conflicts of interest in this simple game strongly suggests that in many natural contexts (e.g., food calls, alarm calls) reliable signalling should not be expected unless it is in the interests of both parties. This conclusion is not altered by separate consideration of the specific costs of producing a signal and of making an appropriate response to that signal.

The results from the second simulation model do not confirm Krebs & Dawkins's conspiratorial whispers theory, but they definitely suggest a modification of it. As Figure 11 shows, when the net payoff to the receiver is marginal, receivers will be sceptical and express "sales-resistance" by responding only to costly signals; signallers in turn will be prepared to invest more energy in "convincing" receivers to respond positively. When communication is unambiguously good for both parties, signals are cheaper and response thresholds lower. Therefore both expensive hype and conspiratorial whispers are expected to evolve, but in a much smaller region of the payoff space than Krebs & Dawkins's theory suggests, i.e., within the cooperative region. Expensive hype is what happens when honest signalling is highly profitable to the signaller, but only marginally so to the receiver. For example, if a juvenile benefits by honestly signalling extreme hunger to its parent (because the parent responds by feeding it), but the net inclusive-fitness payoff to the parent is only slight, then costly signals by the juvenile are expected.

The evolutionary simulation models presented were unusual in their use of non-random initial conditions. The use of non-signalling initial conditions in particular can be seen as an attempt to get at the origin or emergence of communication rather than just studying the conditions for its stability, as does orthodox game theory. Non-signalling initial conditions embody the assumption that communication must emerge from a non-communicative context—the un-clamping of signalling strategies after a period of preliminary evolution can be seen as the introduction of a mutation that allows the *possibility* of signalling. To the extent that this paradigm is seen as plausible, results from the two simulations suggest that sometimes real-world signalling will not evolve despite a cooperative context: receivers may fall into blindly optimistic strategies (i.e., always responding positively) that are less efficient than the communicative equilibrium but nevertheless stable. This is particularly likely to occur when the net payoff to the receiver is high. (The expected payoff for always responding positively will of course depend on the relative frequency of high and low hidden states, a factor that was not varied in the models presented).

There are several qualifications that must be made concerning the results. Firstly, the way that conflicting and congruent interests have been defined may be too simplistic. In the simple signalling game, it is true that with positive net payoffs to the signaller and the receiver, and if the hidden state is high, both agents will benefit from a positive response, and they therefore have congruent interests. However, if we consider the moment before the hidden state has been determined, it is not clear whether the interests of the two agents conflict or not. If the signaller, for example, could somehow choose the strategy of its opponent, the receiver, it would want the opponent to play an "always respond positively" strategy—that way the signaller would always receive the payoff and would not have to expend energy in signalling. However, the receiver, if similarly allowed to determine the signaller's strategy, would prefer that the signaller used an honest strategy, precisely so that the receiver could avoid the costs of responding positively to the low hidden state. Recall that Trivers (1974) defined a conflict of interests as an interaction in which natural selection favours a different outcome for each participant. It seems that the signaller and receiver in this situation favour different strategies in their opponent, and thus have a conflict of interests, even though a high value of the hidden state would mean that their interests became congruent. If this strategy-based definition of conflicting interests were adopted, any situation in the cooperative payoff region, assuming signalling had a positive cost, would involve a conflict of interests—this would in turn mean that *all* of the signalling observed in the simulation models evolved despite a conflict of interests. The problem is perhaps that Trivers's (1974) and Maynard Smith & Harper's (1995) definitions are not specific enough about just what constitutes an "outcome" of the signalling game. The simpler definition of conflicting interests, as used in the body of the paper, is useful in isolating the cooperative region of payoff space as the place to expect signalling. It is not yet clear how the results should be interpreted if the strategy-based definition of conflicting interests was pursued.

A second limitation of the results is that the signalling game used is not likely to be a universal model of all possible communicative interactions. In particular, and despite having the same basic structure with two signals possibly used to transmit information about a binary hidden state, the signalling game is different from those employed by Hurd (1995) and Oliphant (1996). Hurd's game models sexual signalling, and the male signaller is *not* ambivalent about the female receiver's response when the hidden state is low; the signaller always prefers a positive response. A low hidden state maps to low male quality, a positive response represents a copulative episode, and even low-quality males want mating opportunities. The current signalling game, in contrast, cannot model so-called "handicap" signalling, because low-state signallers do not care about what the receiver does. Furthermore, in both Hurd's and Oliphant's games, receivers are explicitly rewarded for accuracy in discerning the hidden state, but the game presented here allows the ecologically plausible outcome that receivers simply become disinterested in the signal. The current game is a reasonable model of situations such as alarm calls[3] and food calls, in which potential signallers have no reason to care about what receivers do when no predator has been sighted or no food source has been found. Whereas Hurd's game serves as a (discrete) model of situations where signallers vary on some dimension, the current game models situations where signallers fall into two groups, only one of which is relevant to the potential response. Hurd's game has been used to model the signalling of mate quality, while the current game could be used to model the signalling of sexual maturity. Future work could certainly look at games like Hurd's, where signallers always want a positive response, in order to determine whether the apparent conflict of interests is real, and under what circumstances signalling evolves. Bullock's (1997, this volume) work considers these questions.

Finally, it must be stressed that the simple games and simulations described here are in one sense an unfair way to test Krebs & Dawkins's (1984) conspiratorial whispers hypothesis. Krebs & Dawkins were discussing the likely evolution of signals in complex real-world cases, and could therefore appeal to the effects of differing mutation rates in signallers and receivers, and the exploitation of behaviours that had originally been selected for other purposes, etc. Communication in the predicted costly signalling arms races was not expected to be stable. For example, in a real-world situation where it was not in the interests of receivers to respond positively to a particular signal from a predator, they might nevertheless continue to do so for some time if the signal was structurally similar to a mating signal made by members of the same species. The manipulative signalling system would break down as soon as an appropriate sequence of mutations resulted in organisms that could distinguish between the predator's signal and the conspecific mating signal. In the simple signalling model all this complexity is abstracted into the base fitness payoffs for signallers and receivers, and there is no guarantee that any transient, unstable evolved communication systems will be detected. The results suggest that in the long run signalling will not be stable unless it is to the mutual advantage of both parties, but this is not to deny that costly signalling arms races under conditions of conflicting interest could occur in the relatively short term. AL models of communication are uniquely equipped to investigate such issues further.

[3]Excepting those cases in which false alarm calls are given in order to frighten off other animals and give the caller a brief period of exclusive access to a food source.

Acknowledgements

I am grateful to the Association of Commonwealth Universities and the British Council for financial support. I would also like to thank Seth Bullock and Ezequiel di Paolo for valuable discussions.

References

Ackley, D. H., and Littman, M. L. 1994. Altruism in the evolution of communication. In Brooks, R. A., and Maes, P., eds., *Artificial Life IV*, 40–48. MIT Press, Cambridge, MA.

Bullock, S. 1997. An exploration of signalling behaviour by both analytic and simulation means for both discrete and continuous models. In Husbands, P., and Harvey, I., eds., *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97)*, 454–463. MIT Press / Bradford Books, Cambridge, MA.

Grafen, A. 1990. Biological signals as handicaps. *Journal of Theoretical Biology* 144:517–546.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. *Journal of Theoretical Biology* 7:1–52.

Hinde, R. A. 1981. Animal signals: Ethological and games-theory approaches are not incompatible. *Animal Behaviour* 29:535–542.

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

Johnstone, R. A., and Grafen, A. 1992. The continuous Sir Philip Sidney game: A simple model of biological signalling. *Journal of Theoretical Biology* 156:215–234.

Krebs, J. R., and Dawkins, R. 1984. Animal signals: Mind reading and manipulation. In Krebs, J. R., and Davies, N. B., eds., *Behavioural Ecology: An Evolutionary Approach*. Oxford: Blackwell, second edition. 380–402.

MacLennan, B. J., and Burghardt, G. M. 1994. Synthetic ethology and the evolution of cooperative communication. *Adaptive Behavior* 2(2):161–188.

Maynard Smith, J., and Harper, D. G. C. 1995. Animal signals: Models and terminology. *Journal of Theoretical Biology* 177:305–311.

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

Maynard Smith, J. 1991. Honest signalling: The Philip Sidney game. *Animal Behaviour* 42:1034–1035.

Oliphant, M. 1996. The dilemma of Saussurean communication. *BioSystems* 37:31–38.

Trivers, R. L. 1974. Parent-offspring conflict. *American Zoologist* 14:249–264.

Werner, G. M., and Dyer, M. G. 1992. Evolution of communication in artificial organisms. In Langton, C. G.; Taylor, C.; Farmer, J. D.; and Rasmussen, S., eds., *Artificial Life II*, 659–687. Addison-Wesley, Redwood City, CA.

Wheeler, M., and de Bourcier, P. 1995. How not to murder your neighbor: Using synthetic behavioral ecology to study aggressive signaling. *Adaptive Behavior* 3(3):273–309.