Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 246 (2007) 225–233
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

The evolution of functionally referential meaning in a structured world

Matina C. Donaldson[a,b,*], Michael Lachmann[b], Carl T. Bergstrom[a]

[a]*University of Washington, Seattle, WA, USA*
[b]*Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

Received 6 March 2006; received in revised form 8 December 2006; accepted 31 December 2006
Available online 9 January 2007

Abstract

Animal communication systems serve to transfer both motivational information—about the intentions or emotional state of the signaler—and referential information—about external objects. Although most animal calls seem to deal primarily with motivational information, those with a substantial referential component are particularly interesting because they invite comparison with words in human language. We present a game-theoretic model of the evolution of communication in a "structured world", where some situations may be more similar to one another than others, and therefore require similar responses. We find that breaking the symmetry in this way creates the possibility for a diverse array of evolutionarily stable communication systems. When the number of signals is limited, as in alarm calling, the system tends to evolve to group together situations which require similar responses. We use this observation to make some predictions about the situations in which primarily motivational or referential communication systems will evolve.

Keywords: Communication; Evolution; Alarm call; Functional reference; ESS

1. Introduction

Most of the signals that animals use to communicate with one another do not seem to have a specific meaning in the same sense that nouns in human language do. Rather, these signals communicate about the intentions, emotional state, or identity of the sender. For example, the song of a male Darwin's finch is thought to identify him as such to conspecific females (Grant and Grant, 1997). Little blue penguins use calls to signal their readiness to escalate a fight (Waas, 1991). Even the alarm calls given by ground squirrels, which were once thought to indicate the type of predator, have been shown instead to relate to the degree of urgency perceived by the caller (Robinson, 1981). However, there are other animal communication systems in which the signals really do seem to refer to some external stimulus. Most famously, vervet monkeys use three qualitatively different alarm calls to distinguish between leopards, eagles and snakes (Cheney and Seyfarth, 1990). Similar predator-specific calls have been found in other primate species (Macedonia, 1990; Zuberbühler et al., 1997) as well as suricates (Manser, 2001) and chickens (Evans et al., 1993). This type of system is not limited to predator warnings alone: toque macaques (Dittus, 1984) and chickens (Evans and Evans, 1999) produce specific calls which alert others to the presence of food.

Semantic communication has been suggested as one of the fundamental differences between animals and humans (e.g. Bickerton, 1990). The communication systems described above, though relatively rare, are of special significance because they hint at the ability of animals to communicate about external objects and events. But does a leopard alarm call really refer to a leopard, in the same sense that the word "leopard" does? Philosophers of language contend that understanding how an utterance is used is insufficient to determine its meaning (Grice, 1957; Quine, 1960); according to this view we can never discover the true meaning of any animal signal. Ethologists have instead focused on demonstrating that some animal signals have the property of *functional reference*: the way in which they are used, and the responses that they engender, give the appearance of referring to some external stimulus (Marler et al., 1992; Macedonia and Evans, 1993). The

*Corresponding author. Tel.: +49 341 3550 549.
E-mail address: matina@u.washington.edu (M.C. Donaldson).

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.12.031

notion that animal signals may have some external referent is not diametrically opposed to the idea that they convey motivational information; rather, it is now well recognized that, like human language, animal signals may simultaneously do both. Still, it is possible to differentiate between systems like the vervet monkeys', which primarily refer to external objects, and systems like the prarie dogs', which primarily reflect the degree of urgency; we are interested in the evolutionary reasons behind this kind of difference.

In this report, we present a model for the evolution of functionally referential meaning in animal communication systems. We begin with a simple action–response model in which selective pressure on the production of the signal is produced by the reactions of those who respond to it, and vice versa. Selection on signals and selection on responses will often work towards one another, eventually leading to a stable and coherent communication system, as has been demonstrated previously with similar models (Hurford, 1989; Wärneryd, 1993; Nowak and Krakauer, 1999). However, these models invariably assume that the world itself takes on a very simple structure: each situation requires a particular, unique response, and all possible alternatives are equally inappropriate. Although this may be an adequate representation of certain economic games, it does not describe animal signalling interactions very well. For example, when a vervet monkey is approached by a leopard, the typical response to an eagle—looking up and running into cover—is much more dangerous than the typical response to a snake—scanning the area (Seyfarth et al., 1980).

In our model of communication in a "structured world", we are able to represent the distinction between not-quite-optimal actions and utterly disastrous ones. We find that a wider variety of signalling systems are evolutionarily stable in our model than in the unstructured worlds of previous models, and this diversity of equilibria more accurately reflects the diversity of modern animal communication. In addition, our model suggests that evolved communication systems may facilitate the categorization of events or situations by appropriate responses, rather than by shared physical characteristics. This may explain why primarily motivational alarm call systems, like that of ground squirrels, are so common, while primarily referential ones, like the vervets', are relatively rare. If motivational states (like fear, arousal, or hunger) have evolved to help organisms make advantageous decisions, then in many cases they may be sufficient to predict an appropriate response to the situation, and thus sufficient to determine which signal to produce. Only in special cases, where the possible reactions are too complex to be determined simply by the urgency of the situation, will a system evolve the characteristic of functional reference.

2. A model for the evolution of communication

Since we are interested in modeling the way that a signal, through use, may come to represent an object or a situation, we begin with a simple sender–receiver game. One individual responds to a stimulus in some observable way; another individual observes that response and reacts in turn. The first individual's action has no power to affect her payoff, while the second individual's reaction affects the payoff of both. In this sense, the first individual's action may be seen as a potential signal to the second individual; it is only through natural selection that these actions gain the status of true signals, as defined by Maynard Smith and Harper (2003, p. 3): "an act or structure which alters the behavior of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved." Once natural selection begins to shape the behavior of individuals in both roles, all of the potential signals that are in use become real signals. Some of these signals may later fall out of use, preventing selection on the response. However, as long as some tendency to respond remains—however it may change through drift—they retain their power to be used as signals.

Now we can define the game more rigorously. The first player, the *signaller*, observes the state of the world $t \in \mathscr{T} = \{t_1, t_2, \ldots, t_l\}$, and selects a signal $s \in \mathscr{S} = \{s_1, s_2, \ldots, s_m\}$. The second player, the *signal receiver*, does not know the state of the world directly, but instead observes the signal s and chooses an action $a \in \mathscr{A} = \{a_1, a_2, \ldots, a_n\}$. Note that the number of distinct signals, m, may be different from the number of states, l, or the number of possible actions, n; we discuss the biological factors affecting the relative numbers of each at the end of this section. We will (conventionally, if somewhat unrealistically (Lachmann et al., 2001)) assume a purely cooperative game: both signaller and receiver obtain the same payoff $\pi(t, a)$, which depends only on the state of the world and the selected response. Since the payoffs are independent of the signal used, all signals are in this sense equivalent to one another. For simplicity, we also assume that all signals are transmitted without error.

In this sender–receiver game, the signaller's strategy can be represented by a matrix $\mathbf{P}$ which contains the conditional probabilities $p(s|t)$ of producing each signal s, given each world state t. Similarly, the receiver's strategy is represented as a matrix $\mathbf{Q}$ that provides the conditional probabilities $q(a|s)$ of selecting an action a, given signal s. Each individual can play both signalling and receiving roles, so a complete strategy R consists of both a $\mathbf{P}$ matrix and a $\mathbf{Q}$ matrix.

We can calculate expected payoffs, given a probability distribution on world states $p(t)$. If we further assume that each individual spends half the time as signaller and half the time as receiver, the expected payoff to an individual with strategy $R = (\mathbf{P}, \mathbf{Q})$ of interacting with an individual with strategy $R' = (\mathbf{P}', \mathbf{Q}')$ will be

$$\bar{\pi}(R, R') = \frac{1}{2}\sum_{t\in\mathscr{T}}\sum_{s\in\mathscr{S}}\sum_{a\in\mathscr{A}} p(t)p(s|t)q'(a|s)\pi(t,a) + \frac{1}{2}\sum_{t\in\mathscr{T}}\sum_{s\in\mathscr{S}}\sum_{a\in\mathscr{A}} p(t)p'(s|t)q(a|s)\pi(t,a).$$

Here the first summand is the expected payoff to the individual when acting as a signaller, and the second summand is her expected payoff as a receiver.

Since an individual's ability to communicate depends not only on her own strategy, but also on those of others around her, selection for communicative ability is frequency-dependent. Evolutionary game theory gives us a way to model these sorts of interactions. In particular, the concept of an *evolutionarily stable strategy* (ESS) (Maynard Smith and Price, 1973) provides a way to characterize the long-term behavior of a population without getting into the details of the evolutionary dynamics away from equilibrium. We assume that individuals reproduce asexually according to fitness, which is proportional to payoff in the game, and that offspring learn or otherwise inherit the strategy of their parents with some chance of error. Under a wide range of evolutionary dynamics, a population of individuals playing an ESS will be invulnerable to invasion by mutant strategies. In fact, such a population is also a long-term endpoint of the evolutionary process (Eshel and Feldman, 1984; Hammerstein, 1996).

Lewis (1969) uses a similar signaller–receiver game to describe the conventionalization of meaning in natural language, but he does not address evolutionary questions. Hurford (1989) uses computer simulations to look at the evolutionary process and Trapa and Nowak (2000) find the evolutionarily stable states of a related model. These models differ from ours in that the receivers choose an interpretation from the original set of world states, rather than choosing an action. Fitness is determined by the proportion of correct interpretations. However, the very idea that signals have interpretations presupposes that the communication system is used to convey referential meaning. Since we are interested in the evolution of reference, we prefer to extend a model developed by Wärneryd (1993). His paper is not really about the evolution of communication in itself; his primary goal is to show how cost-free, arbitrary signals can stabilize equilibria in a cooperative game. However, since it makes no assumptions about the "meaning" of a signal, it provides an ideal framework for exploring the evolution of motivational and referential communication.

Wärneryd's framework is more general than the Hurford–Nowak model mentioned earlier, because the players respond to a signal by choosing an action, rather than simply inferring which situation gave rise to the signal. However, he assumes a special form of payoff matrix on signals and actions which makes his model functionally equivalent to theirs. In his model, each world state has a unique best response, and the payoff for all other actions is zero. We relax this assumption to permit arbitrary structural relationships between world states. This representation allows for much more realistic models of animal communication systems. While each state has an optimal response, we allow some of the remaining responses to be better than others. In addition, some actions may be reasonably good for several situations, without being ideal for any (see Fig. 1).

a

Π	a_1	a_2	a_3
t_1	1	0	0
t_2	0	1	0
t_3	0	0	1

b

Π	a_1	a_2	a_3	a_4
t_1	8	1	1	3
t_2	0	7	1	3
t_3	0	0	6	3

Fig. 1. Sample payoff matrices for (a) the Hurford–Nowak model (Hurford, 1989; Nowak and Krakauer, 1999; Trapa and Nowak, 2000) and (b) our model. In (a), each state has one appropriate response, and all others are useless. In (b), actions a_1, a_2, and a_3 are best responses to t_1, t_2, and t_3 but general-purpose action a_4 may be better when the state of the world is uncertain.

Wärneryd also assumes that there are at least as many signals as there are states or actions. We do not, and we will be particularly interested in cases where the number of signals is smaller than either. These cases seem most similar to real animal communication systems over a wide range of taxa, in which the assortment of distinct signal types is surprisingly limited (Moynihan, 1970). Why should this be so? One limitation is imposed by the receivers, who must be not only able but also likely to perceive the signaler's action. That is, we can restrict our attention to the domain, or domains, in which actions cause others to react. This could be, for example, sounds within a certain range of frequencies, or the position of the tail feathers. Another limitation is that the receivers must be able to reliably discriminate different signals. The effects of a noisy environment can create a tradeoff between increasing the number of signals and being able to distinguish between them (Nowak et al., 1999). Since we are interested in the evolution of the function rather than the form of the signal, instead of explicitly modeling this process we will simply assume a fixed number of signals (but see Zuidema, 2003, for a computational approach to modeling both processes together).

3. Evolutionary stability of communication systems

A Nash equilibrium strategy is one which is a best reply to itself; when such a strategy is common, though no alternate strategy can be selected for, some may drift in neutrally. In contrast, a *strict* Nash equilibrium strategy outperforms all other strategies when playing against itself, so no strategy can neutrally invade. The conditions for an ESS lie in between these two extremes: some strategies can invade neutrally, as long as the ESS is strictly superior once the invading strategy becomes common. So, in general, a strict Nash equilibrium is a special type of ESS. However, Selten (1980) showed that for role-asymmetric games (in which players are assigned different roles), *every* ESS must be a strict Nash equilibrium. In this game, therefore, the signalling strategy in an ESS must be uniquely optimal against the receiving strategy, and vice versa.

The following four conditions are necessary for the signalling system $R = (\mathbf{P}, \mathbf{Q})$ to be a strict Nash equilibrium, and therefore an ESS. The first two properties follow directly from Selten's (1980) proof.

Property 1. *The signalling strategy* **P** *must be binary*; *that is*, *each state gives rise to exactly one signal.*

Property 2. *The receiving strategy* **Q** *must be binary*; *that is*, *each signal results in exactly one action.*

Property 3. *The signalling strategy* **P** *must be onto*; *that is*, *every signal must be used.*

Proof. Suppose that **P** is not onto; the ith column in **P**, corresponding to the production of signal s_i, is all zeros. Then the ith row in **Q**, corresponding to the response to s_i, can be altered without changing the expected payoff. Thus **Q** is not the unique best reply to **P**, so R cannot be a strict Nash equilibrium. □

Property 4. *The receiving strategy* **Q** *must be one-to-one*; *that is*, *no two signals may give rise to the same action.*

Proof. Suppose that **Q** maps two signals to the same action. Since **Q** is binary, there must then be two identical rows in **Q**, say those indicating the response to signals s_i and s_j. Then we can swap the ith and jth columns in **P**, which are the production conditions for the two signals, without changing the expected payoff. The resulting $\mathbf{P}'$ must differ from **P** because, by Properties 1 and 3, no two columns are identical. Therefore **P** is not the unique best reply to **Q**, and R cannot be a strict Nash equilibrium. □

These properties limit the multiplicity that is allowable in the signalling and receiving mappings. There are four possible types of multiplicity: (1) one situation leads to multiple signals; (2) one signal leads to multiple actions; (3) multiple situations lead to the same signal; and (4) multiple signals lead to the same action. The first two, as stated in Properties 1 and 2, are addressed by Selten's theorem: an ESS can have only one possible response to each circumstance. Even if some responses may sometimes perform better than others, as long as the player has no further information, the best she can do is to calculate the response which gives the highest payoff on average. The fourth multiplicity is also disallowed, as stated in Property 4: if more than one signal gave rise to the same action, then signallers could use the two interchangeably. However, as we saw above, using two signals in the same situation is never part of a stable strategy. The third type of multiplicity, on the other hand, is perfectly okay: if signallers use the same signal in multiple situations, the signal comes to "mean" to the receivers that one of several situations has occurred, each with some specified probability. As long as the payoffs are asymmetrical, it is still possible to calculate the action with maximal payoff.

While all evolutionarily stable communication systems must meet these four conditions, there are some systems which display these properties and yet are not stable. Next we add two additional properties which fill out the set of sufficient conditions for evolutionary stability.

When there are more signals than states, no strategy fulfills the four conditions above; Properties 1 and 3 cannot hold simultaneously. Similarly, when there are more signals than actions, Properties 2 and 4 are in conflict. When there are equal numbers of states and signals or signals and actions, these conditions impose an exact correspondence between them.

Perhaps the most interesting case is when there are fewer signals than states, because this seems to reflect what we see in most animal communication systems today. In this case, multiple states map to a single signal, which in turn maps to just one action. This divides the set of all world states into smaller, non-overlapping *pools* (see Fig. 2). There is one pool for each signal, and every world state is included in some pool. Note that this usage of the term pool is similar in spirit to the notion of semi-pooling equilibria in costly signalling theory (Lachmann and Bergstrom, 1998; Bergstrom and Lachmann, 1998). However, the reason for the grouping of signaller types in the costly signalling models is quite different: the conflict of interest between signaller and receiver means that in some cases a compromise can be reached in which only partial information is sent. In the current model, we assume that signallers and receivers share the same interests, so the pooling is due only to a limitation in the signals.

The following definitions and properties assume a strategy $R = (\mathbf{P}, \mathbf{Q})$ which satisfies the conditions in Properties 1–4.

Definition 1. The pool of states $\tau(s)$ associated with a signal s is the set of states mapping to that signal under **P**: $\tau(s) = \{t : p(s|t) = 1\}$.

Definition 2. A best response to a pool of states is an action which maximizes the expected payoff for all members of the pool:

$$BR(\tau) = \arg\max_{a \in \mathcal{A}} \sum_{t \in \tau} p(t)\pi(t, a).$$

If there is a unique such action, it is termed the *strict best response* (SBR) to the pool.

a

P	s_1	s_2
t_1	1	0
t_2	1	0
t_3	1	0
t_4	0	1
t_5	0	1

Q	a_1	a_2	a_3	a_4
s_1	0	1	0	0
s_2	0	0	0	1

b

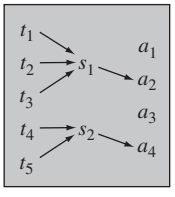


Fig. 2. An illustration of pooling. An example strategy represented by (a) the signalling and receiving matrices; and (b) a schematic diagram. Here, signal s_1 is associated with the pool of states $\{t_1, t_2, t_3\}$ and s_2 is associated with $\{t_4, t_5\}$.

Property 5. *Every pool must have a SBR, and the signal corresponding to that pool must map to it*: $q(BR(\tau(s))|s) = 1$ *for all* $s \in \mathscr{S}$.

Proof. If there is a SBR to a pool of states grouped under one signal by **P**, clearly an optimal reply **Q** must map the signal to that action. On the other hand, suppose there is no SBR to a pool $\tau(s)$. Then at least two different actions maximize the expected payoff for the pool. Any one of these actions can be chosen as the response to the signal s without changing the overall payoff. Thus there is no strict optimal reply to **P**, and R cannot be a strict Nash equilibrium strategy. □

Property 6. *For each member of a pool of states, the SBR for that pool must be a better response than the SBR of any other pool. That is, for all* $t \in \tau(s_i)$ *and* $s_j \neq s_i$,

$$\pi(t, BR(\tau(s_i))) > \pi(t, BR(\tau(s_j))).$$

Proof. We assume every signal maps to the SBR for its pool of states. Suppose that for one state t within a pool $\tau(s_i)$, the SBR of another pool $\tau(s_j)$ provides an equally good or better response; then the signalling strategy can be changed to map t to s_j instead of s_i. The resulting $\mathbf{P}'$ will perform just as well or better against **Q** than **P** does, so R cannot be a strict Nash equilibrium. □

Theorem 1. *A strategy* $R = (\mathbf{P}, \mathbf{Q})$ *is evolutionarily stable if and only if the six properties listed above hold.*

Proof. We have already shown necessity; we now show sufficiency. Given that Properties 1–4 hold, Property 5 ensures that **Q** is the single best reply strategy to **P**, and Property 6 ensures that **P** is the single best reply strategy to **Q**. Therefore, any other strategy will do strictly worse against R than R does against itself, so R must be an ESS. □

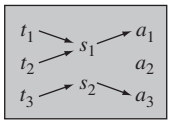
Example 1. Consider the case where there are equal numbers of states, signals and actions, and each state has a unique best response. Then an ESS will assign a signal to each state, and map that signal to the state's best response. The assignment of signal to state is arbitrary, so there will be one such ESS for every possible permutation of signals; functionally, however, all these strategies are equivalent.

In the papers by Wärneryd (1993) and Trapa and Nowak (2000), such communication systems are the only possible ESSs. Constraining the probability of states and the payoff matrix to be completely symmetrical means that no pool of states bigger than one can have a unique best response; this means that under such a model an ESS cannot exist if there are fewer signals than states. In real biological systems, however, two different things are almost never exactly equally likely, nor do they give exactly the same fitness. In a model which removes these unrealistic constraints, we will see that evolutionary stability is not only possible, but even likely.

a

Π	a_1	a_2	a_3
t_1	7	0	2
t_2	4	6	0
t_3	0	5	10

b



c

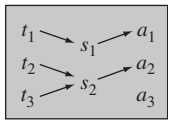


Fig. 3. Different poolings can yield different ESSs for the same system. In this example, all states are equally likely. The evolutionarily stable communication system shown in (b) has a payoff of 7 and is payoff-maximizing, while the evolutionarily stable communication system shown in (c) has a payoff of 6. The third possible pooling, not shown, is not evolutionarily stable.

Example 2. Consider the system shown in Fig. 3. When there are fewer signals than states, an ESS must group some of the states together. The most efficient grouping, shown in Fig. 3b, maps the states t_1 and t_2 to a single signal, while t_3 is differentiated from the others. This is an ESS because both pooling properties are satisfied. While a_1 is not the optimal action for t_2, it is the best of the limited possibilities created by the receiver's strategy.

The communication system illustrated in Fig. 3c is also stable, but non-optimal. The pooling in this strategy creates a kind of evolutionary cul-de-sac: no improvement is possible by changes in either the signalling strategy or the receiving strategy alone.

In some cases, there may be no strategy which is strictly superior to all invaders. What then can we expect to happen, after enough evolutionary time? One possibility is that a *set* of strategies exists such that any strategy in the set is neutrally stable with regard to any other strategy in the set, but which as a set is invulnerable to invasion from outside the set. This is called an *evolutionarily stable set*, or ES set (Thomas, 1985; Balkenborg and Schlag, 2001). In this case, the system never reaches a true equilibrium, but can drift neutrally among the strategies in the set without leaving.

The conditions described above for an ESS need be modified only minimally in order to characterize an evolutionarily stable set. Rather than requiring a single uninvasible strategy, we look for a set of strategies, all of which generate the same payoff against one another, but which are otherwise uninvasible. Just as the role asymmetry in the game ensures that any ESS must be a strict Nash equilibrium, it also guarantees that any ES set must be a strict equilibrium set (Balkenborg, 1994). A strict equilibrium set is a set of Nash equilibria which is closed under best replies. For this game, this means that an ES set must consist of a pair of strategy sets, where each signalling strategy has as its set of best replies the receiving strategy set, and each receiving strategy has as its set of best replies the signalling strategy set.

Example 3. Consider a system with three states and three actions, and a payoff matrix as pictured in Fig. 4a. If there are only two signals, one might expect an optimal strategy to group t_1 and t_2 together with one signal. However, there

a

Π	a_1	a_2	a_3
t_1	2	1	0
t_2	1	2	0
t_3	0	0	2

b

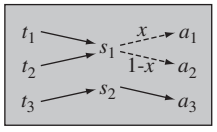


Fig. 4. An example of a system with an ES set consisting of a single signalling strategy and multiple receiving strategies. The payoff matrix is displayed in (a), and all states are equally likely. The strategy displayed in (b) is payoff-maximizing for any $x \in [0, 1]$, where x is the probability of a response a_1 to the signal s_1. The set of all such strategies is evolutionarily stable.

is then no unique best response to that signal, because either a_1 or a_2 will result in the same payoff. There can therefore be no ESS grouping the first two states; in fact, no ESS exists for this system. However, the set of strategies shown in Fig. 4b, consisting of each pair $(\mathbf{P}, \mathbf{Q}(x))$ for all x between 0 and 1, is evolutionarily stable. The set of best responses to the pool of states $\{t_1, t_2\}$ is $\{a_1, a_2\}$ and the best response to t_3 is a_3, so the set of matrices $\{\mathbf{Q}(x)\}$ is the best reply set to the $\mathbf{P}$ matrix displayed. Additionally, both t_1 and t_2 are better represented by either a_1 or a_2 than by a_3, so $\mathbf{P}$ is the strict best reply to any member of the set of $\mathbf{Q}$ matrices.

Just as an ES set may have multiple receiving strategies, an ES set can also have multiple signalling strategies; this occurs when more than one signal is a best response to some state. An ES set can even have both multiple signalling strategies and multiple receiving strategies. Evolutionarily stable sets thus consist of a set of signalling strategies and a set of receiving strategies, where each member of each strategy set has as its set of best replies the entire opposing strategy set. This is still a fairly restrictive condition (particularly if one demands multiple signalling strategies and multiple receiving strategies), and there is no guarantee that an ES set will exist if an ESS does not.

Even when no ES set exists, there may be a subset of the entire strategy space which is invulnerable to invasion from outside. We can construct such a set as follows: take a single strategy and add it to the set, then add its set of best reply strategies, then take each of these strategies and add its set of best replies, and so on, stopping when all best replies are already in the set. If this set is not the entire strategy space, then the population may drift neutrally along certain paths within the set, without ever leaving the set. In this sense, it may be considered evolutionarily "stable" though not an ES set.

Example 4. Consider a system with two states, three signals, and two actions. Any binary signalling strategy $\mathbf{P}$ which differentiates the two states will have one unused signal; since this signal is never used, any action would be an appropriate response. On the other hand, any binary receiving strategy must map more than one signal to the same action; since these two signals produce the same response, either can be substituted for the other. This means that starting with a binary signalling strategy that uses two signals, the receiving strategy can wander neutrally until the unused signal has only a single response; then the signalling strategy can wander neutrally until it switches entirely to the previously unused signal, and so on (see Fig. 5 for a specific example).

There are six possible binary signalling strategies which distinguish both states, represented by the upper hexagon in Fig. 5, and there are six possible binary receiving strategies which use both actions, represented by the lower hexagon. All allowable pairs in this set (a corner in one strategy, paired with any point from the neighboring line in the other strategy) have equal, maximal fitness, so after enough evolutionary time, the system will reach some such pair. Once in this set of allowable pairs, the system will wander around neutrally, alternating between changing the signalling strategy and the receiving strategy. Notice that opposite points on the hexagons may be said to give the signals exactly opposite "meanings"; they are produced in the opposite context and induce the opposite action.

The set described in the previous example is not an ES set, because for any strategy pair within the set, most other invading strategy pairs in the set will be selected against. Only those lying along the same segment can neutrally invade. Still, once any point in the set is reached, the communication system can wander neutrally only within the set, and cannot be invaded by any strategy outside the set. This behavior is not unique to this example. We have made no mention of the payoff structure or the probability distribution of states because neither has any effect on the evolutionary behavior of the system. In fact, *any* system which has fewer states than either signals or actions will show similar behavior.

By contrast, the multiple best replies seen in Example 3 arise because two actions give equal payoffs under some pool of states. Because this type of equivalence will disappear when some of the payoffs or probabilities are changed by an arbitrarily small amount, neutral stability of this sort is unlikely to be biologically relevant. Excluding such cases, a system which is limited by the number of signals—which, we have argued, is the most biologically relevant case—will always have an ESS. Limiting the number of signals ensures that the pair of strategies which maximizes payoff will be of the form described by Properties 1–4. Without symmetry in the payoff structure, Properties 5 and 6 will hold in their strict form, so the payoff-maximizing strategy will also be an ESS.

When the number of signals exceeds the number of states and actions, every system will wander neutrally as in Example 4. In this case, a binary receiving strategy has multiple best replies because at least one signal is not used, and can therefore be responded to arbitrarily. However, if errors occur in transmission, every signal will be received with non-zero probability, and this equality in responses will no longer hold. Additionally, a binary signalling strategy has multiple best replies because two of the signals

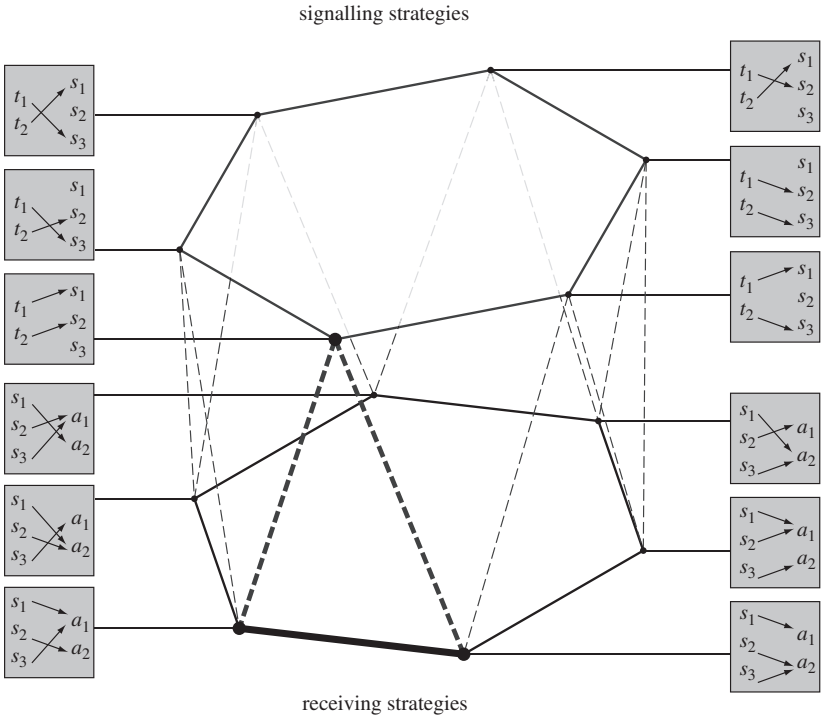


Fig. 5. Meaning can cycle continually when there are more signals than states or actions. Here we have two states, three signals, and two actions. The upper hexagon represents a subset of signalling strategies, where each corner is a binary strategy. Adjacent binary strategies are identical except that they use different signals in response to one of the states. The line between them represents a set of strategies which map that state to the two different signals with positive probability. Similarly, the lower hexagon represents a subset of receiving strategies. When the signalling strategy is fixed at the point marked in bold, where s_3 is unused, the receiving strategy may wander neutrally along the bold line, because both actions are equally good responses to a signal that is never used. Along the interior of this line of receiving strategies, the signalling strategy remains fixed, but once either endpoint is reached (say, the one which maps s_3 to a_2) the signalling strategy will be free to vary, because now two signals (in this case, s_2 and s_3) have the same response and may therefore be used interchangeably.

give rise to the same action, and are therefore equivalent. If the signals have different probabilities of being mistaken for one another, however, this equivalence disappears. When error occurs and particularly when the probability of error varies from signal to signal as in Nowak et al. (1999), we could still have an ESS (that uses only some subset of the available signals) even when there are more signals than states or actions.

4. Conclusions

Previous models of the evolution of communication have suggested that only systems which use a unique signal for every situation can be evolutionarily stable. Under such models, if there were too few signals to distinguish all relevant cases, long-term persistence of a communication system would be impossible. Since real predator alarm call systems tend to employ only a few signals to distinguish between predators, with many types lumped together, a question remained: are these systems evolutionarily unstable, destined to change unpredictably over time, or were the models simply failing to capture some important characteristic of the system? We answer this question by extending those models to allow a more structured representation of the world: certain mistakes in comprehension may carry a higher cost than others. In doing so, we find that stable communication systems are possible under a much broader range of conditions—and thus explain how real predator alarm systems can persist over evolutionary time.

When the ability to discriminate between situations is limited by the number of signals, a communication system must group some situations together by using a common signal. In general, the stability of any particular grouping depends crucially on how important it is to distinguish among the states that are pooled together. Therefore, evolved communication systems of this sort should tend to group states which are similar in a functional sense. Rather than categorizing predators according to morphological characteristics, evolved alarm call systems should group predators which require a similar escape response. For example, Southern lapwings produce an aerial alarm call to several species of hawks, but ignore a similar-looking species which eats only fish (Walters, 1990). The notion that communication and categorization may be intimately linked by the process of evolution has been suggested before (e.g. Allen and Saidel, 1998); our paper demonstrates a mechanism for creating such a linkage.

Alarm call systems lie somewhere along a continuum between the two extremes of predator-specific systems, which distinguish between types of predators, and risk-based systems, which indicate the degree of threat posed by

the predator. It has been suggested that the primarily functionally referential alarm call systems of vervets and ringtailed lemurs evolved because different classes of predators require incompatible escape responses (Macedonia and Evans, 1993). Determining whether this provides a general explanation for the evolution of functional reference, of course, will require detailed study of other alarm calling systems with varying degrees of referential specificity. If the theory holds, however, our model demonstrates *why* this should be so: when categorizing situations by appropriate response yields the same groupings as categorizing them by type, a stable communication system will also show functional reference. On the other hand, when the appropriate response is dictated by the level of urgency, a stable communication system need only specify that level. Whether an alarm calling system evolves to be primarily referential or motivational is determined precisely by what types of situations require different responses.

What about systems that communicate something besides the approach of a predator—like the discovery of food, or agonistic interactions with conspecifics? Although some research has been done in both of these areas which indicates the possibility of referential communication (e.g. Hauser, 1998; Gouzoules et al., 1984) it has been more difficult to demonstrate because the responses to such calls are much less specific. In both cases, individuals react to the calls by orienting towards or approaching the caller, and what is usually measured is the latency to and/or duration of such a reaction. It has therefore been difficult to show that distinct calls really refer functionally to distinct types of food or distinct kinds of interactions. For the same reason, the hypothesis put forth for alarm calls, postulating that mutually incompatible responses to different classes of predators gives rise to referentially specific alarm calls, seems unlikely to hold here, unless there are more specific reactions to different kinds of food and/or agonistic interaction calls which we simply do not observe.

Finally, what implications does our model have for the evolution of referential communication in human language—if any? After all, since we cannot know whether animals associate their signals with some kind of internal representation of external objects, it is still possible that the kind of functional reference we have described here bears only superficial resemblance to the type of referential meaning that words in human language have (Owren and Rendall, 2001). Yet even a superficial resemblance to a referential system could have provided the conditions necessary for a truly referential system to develop. Because we make no assumptions about the "meaning" of signals, our model would provide an appropriate framework for exploring that possibility. It is still a subject of hot debate whether human language evolved from other animal communication systems for the purpose of communication, or is rather an independent outgrowth of selection for enhanced cognitive abilities (Hauser et al., 2002; Pinker and Jackendoff, 2005). Though the results described here cannot contribute directly to this debate, a model based upon ours which demonstrated how symbolic reference, as used in human language, could evolve from functional reference, as seen in other animal communication systems, would provide support at least for the plausibility of the first hypothesis.

Acknowledgments

The authors would like to thank Willem Zuidema, Chris Templeton and several anonymous reviewers for helpful comments. This material is based upon work supported under an NSF Graduate Research Fellowship to M.D. and, in part, by a UW-RRF grant to C.B.

References

Allen, C., Saidel, E., 1998. The evolution of reference. In: Cummins, D.D., Allen, C. (Eds.), The Evolution of Mind. Oxford University Press, New York, pp. 183–203.

Balkenborg, D., 1994. Strictness and evolutionary stability. Center for Rationality and Interactive Decision Theory, Jerusalem, Discussion Paper 52.

Balkenborg, D., Schlag, K.H., 2001. Evolutionarily stable sets. Int. J. Game Theory 29, 571–595.

Bergstrom, C.T., Lachmann, M., 1998. Signaling among relatives. III. Talk is cheap. Proc. Natl. Acad. Sci. USA 95, 5100–5105.

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

Cheney, D.L., Seyfarth, R.M., 1990. How Monkeys See the World. University of Chicago Press, Chicago.

Dittus, W.P., 1984. Toque macaque food calls: semantic communication concerning food distribution in the environment. Anim. Behav. 32, 470–477.

Eshel, I., Feldman, M.W., 1984. Initial increase of new mutants and some continuity properties of ESS in two-locus systems. Am. Nat. 124, 631–640.

Evans, C.S., Evans, L., 1999. Chicken food calls are functionally referential. Anim. Behav. 58, 307–319.

Evans, C.S., Evans, L., Marler, P., 1993. On the meaning of alarm calls: functional reference in an avian vocal system. Anim. Behav. 46, 23–38.

Gouzoules, S., Gouzoules, H., Marler, P., 1984. Rhesus monkey (*Macaca mulatta*) screams: representational signalling in the recruitment of agonistic aid. Anim. Behav. 32, 182–193.

Grant, P.R., Grant, B.R., 1997. Genetics and the origin of bird species. Proc. Natl. Acad. Sci. USA 94, 7768–7775.

Grice, H.P., 1957. Meaning. Philos. Rev. 66, 377–388.

Hammerstein, P., 1996. Darwinian adaptation, population genetics and the streetcar theory of evolution. J. Math. Biol. 34, 511–532.

Hauser, M.D., 1998. Functional referents and acoustic similarity: field playback experiments with rhesus monkeys. Anim. Behav. 55, 1647–1658.

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

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

Lachmann, M., Bergstrom, C.T., 1998. Signalling among relatives. II. Beyond the tower of Babel. Theor. Popul. Biol. 54, 146–160.

Lachmann, M., Számadó, S., Bergstrom, C.T., 2001. Cost and conflict in animal signals and human language. Proc. Natl. Acad. Sci. USA 98, 13189–13194.

Lewis, D.K., 1969. Convention: A Philosophical Study. Harvard University Press, Cambridge, MA.

Macedonia, J.M., 1990. What is communicated in the antipredator calls of lemurs: evidence from playback experiments with ringtailed and ruffed lemurs. Ethology 86, 177–190.

Macedonia, J.M., Evans, C.S., 1993. Variation among mammalian alarm call systems and the problem of meaning in animal signals. Ethology 93, 177–197.

Manser, M.B., 2001. The acoustic structure of suricate alarm calls varies with predator type and the level of response urgency. Proc. R. Soc. London B 268, 2315–2324.

Marler, P., Evans, C.S., Hauser, M.D., 1992. Animal signals: motivational, referential, or both? In: Papousek, H., Jürgens, U., Papousek, M. (Eds.), Nonverbal Vocal Communication: Comparative and Developmental Approaches. Cambridge University Press, Cambridge, pp. 66–86.

Maynard Smith, J., Harper, D., 2003. Animal Signals. Oxford University Press, New York.

Maynard Smith, J., Price, G.R., 1973. The logic of animal conflict. Nature 246, 15–18.

Moynihan, M., 1970. Control, suppression, decay, disappearance and replacement of displays. J. Theor. Biol. 29, 85–112.

Nowak, M.A., Krakauer, D.C., 1999. The evolution of language. Proc. Natl. Acad. Sci. USA 96, 8023–8028.

Nowak, M.A., Plotkin, J.B., Dress, A., 1999. An error limit for the evolution of language. Proc. R. Soc. London B Biol. Sci. 266, 2131–2136.

Owren, M.J., Rendall, D., 2001. Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. Evol. Anthropol. 10, 58–71.

Pinker, S., Jackendoff, R., 2005. The faculty of language: what's special about it? Cognition 95, 201–236.

Quine, W.V., 1960. Word and Object. MIT Press, Cambridge, MA.

Robinson, S.R., 1981. Alarm communication in Belding's ground squirrels. Z. Tierpsychol 56, 150–168.

Selten, R., 1980. A note on evolutionarily stable strategies in asymmetric animal conflicts. J Theor Biol 84, 93–101.

Seyfarth, R.M., Cheney, D.L., Marler, P., 1980. Vervet monkey alarm calls: semantic communication in a free-ranging primate. Anim. Behav. 28, 1070–1094.

Thomas, B., 1985. On evolutionarily stable sets. J. Math. Biol. 22, 105–115.

Trapa, P.E., Nowak, M.A., 2000. Nash equilibria for an evolutionary language game. J. Math. Biol. 41 (2), 172–188.

Waas, J.R., 1991. The risks and benefits of signaling aggressive motivation—a study of cave-dwelling little blue penguins. Behav. Ecol. Sociobiol. 29, 139–146.

Walters, J.R., 1990. Anti-predatory behavior of lapwings: field evidence of discriminative abilities. Wilson Bull. 102 (1), 49–70.

Wärneryd, K., 1993. Cheap talk, coordination, and evolutionary stability. Games Econom. Behav. 5, 532–546.

Zuberbühler, K., Noë, R., Seyfarth, R.M., 1997. Diana monkey long-distance calls: messages for conspecifics and predators. Anim. Behav. 53, 589–604.

Zuidema, W., 2003. Optimal communication in a noisy and heterogeneous environment. In: Proceedings Lecture Notes in Artificial Intelligence, vol. 2801. Springer, Berlin, pp. 553–563.