Numerous animal signals are used to communicate, such as calls, postures, and mating patterns. These signals are often simple and specific, conveying information about the sender's identity, sex, or the presence of a predator. However, the evolution of more complex communicative systems, such as language, is a subject of ongoing research. Language is unique in its ability to convey abstract concepts, ideas, and social norms, transcending the physical environment and allowing for the transmission of knowledge across generations. This capability is thought to be a major factor in the evolutionary success of humans, as it facilitates cooperation, socialization, and the development of complex societies.

Evolutionary models of communication are used to shed light on the selective pressures involved in the evolution of simple referential signals, and the constraints hindering the emergence of signs. Error-prone communication results from errors in transmission (in which individuals learn the wrong associations) and communication (in which signs are mistaken for one another). We demonstrate how both classes of errors are required to generate diversity and subsequently impose limits on the sign repertoire within a population. We then explore the influence of geographic structuring of a population on the evolution of a shared sign system and the importance of such structure for the maintenance of sign diversity. Deceit tends to erode conventional sign systems thereby reducing signal diversity, we demonstrate that population structure can act as a hedge against deceit, thereby ensuring the persistence of sign systems.

1. Introduction

The apparent absence of animal signals comparable to human language, even among apes (Cheney & Seyfarth 1982; Savage-Rumbaugh 1992; Savage-Rumbaugh et al., 1994), suggests the existence of constraints on signal evolution. Some see human language as a fortunate byproduct of large brains evolved for non-linguistic purposes combined with some as yet unknown laws of structure and growth (Piatelli-Palmarini, 1989). However, the complexity of language, its evident design for the efficient transmission of propositional structures through a serial interface (the speech apparatus) (Pinker & Bloom, 1990), and its “universal” nature (sensu Chomsky, 1972) make a signalling-based hypothesis appealing. Furthermore, these design elements in language make it possible that it evolved under the pressures of Darwinian natural selection (Pinker & Bloom, 1990; Szathmáry & Smith, 1995).

Adaptive evolution by natural selection can proceed at wildly variable rates, but tends towards small increments. How could a gradualistic process have gone from animal-style signals to the discrete and complex syntactic network that constitutes human language? This puzzle has led some to look for clues in the recapitulation of evolutionary history in language development (Bates et al., 1989; Studdert-Kennedy, 1998), or in the process of creolization (Bickerton, 1990). However, such approaches are always at risk of not distinguishing between current function and “adaptiveness” in our ancestors. Thus, modern infants are already in possession of an advanced language device and thereby might acquire language in a fundamentally different way.
(Pinker, 1996). Nevertheless, the analogy between learning and early evolution has proved to be very illuminating regarding necessary expectations and constraints in language acquisition.

In this paper, we investigate what is likely to have been one of the first, essential steps in the evolution of a sign system, namely the formation of discrete referential signals constituting a rudimentary “lexicon”. A lexicon is simply the set of correlations that is formed between objects and signal responses, or sounds. Thus, we explore the emergence of a simple form of reference, which it is assumed constitutes one of the first steps towards a more complex mode of communication. In essence, reference makes coordination between signallers and receivers more difficult by expanding the set of states about which information is transmitted. Reference is hardly unique to human sign systems, but the magnitude of reference found among the human species is far in excess of that found among other species. Here we attempt to characterize those factors influencing this disparity.

Following Nowak & Krakauer (1999), we employ game theoretic models that assign payoffs to communication (see also Oliphant, 1996; DiPaolo, 1997). In the first part of the paper, the payoff from communication is interpreted as an “imitative fitness”, such that an individual with a high payoff is more likely to be imitated. Under such a scenario it has been shown that the population will rapidly develop a single lexicon allowing mutual comprehension (Nowak & Krakauer, 1999). This paper explores the influence of errors during communication, and errors that arise during transmission of the lexicon. We stress that both types of errors play a very important role in both facilitating the emergence of a sign system and in constraining the number and diversity of signs employed. We then consider the influence of geographic structuring of the population on the development of a shared lexicon and its role in facilitating lexicon evolution and how it protects against the risks of deception.

2. Basic Language Game

In the basic language game an individual communicates the presence of an object to another individual by making a sound. The sound produced is determined by the speaker’s lexicon $L$. In the simple model introduced here the lexicon consists of a single matrix $P$, where $p_{ij}$ determines the probability that sound $j$ is associated with object $i$. Sound comprehension is, in turn, determined by the lexicon of the listener. Due to its probabilistic nature the rows of $P$ must sum to one, which implies that all sounds are heard and interpreted. When the entries in $P$ are constrained to be either 0 or 1 then this is defined as a consistent lexicon (as opposed to inconsistent). Each object will always be associated with the same sound by the speaker (the speaker is consistent), and each sound will be uniquely interpreted by the listener.

Consider two individuals A and B, with $L_A$ and $L_B$, respectively, each referring to $n$ objects with $m$ sounds. We will consider $m = n$. The comprehension of A by B is given by the matrix product $P_A P_B^T$, where $T$ indicates the transpose of $P_B$. The diagonals of this matrix product give the probability that A correctly informs B about object $i$, whilst the off-diagonals give the probabilities of misunderstanding (probability that object $i$ is understood as object $j$). A measure of the ability of A to convey information to B, assuming that all objects arise with equal probability, is given by

$$F(L_A, L_B) = \text{tr}\{P_A P_B^T\},$$

(1)

where $\text{tr}\{M\}$ indicates the trace of a matrix $M$ (sum of the diagonal elements). Note the provisional assumption that $F(L_A, L_B) = F(L_B, L_A)$, and if $P_A$ is consistent $F(L_A, L_B) = n$. If each object $i$ has value $w_i$ then

$$F(L_A, L_B) = \text{tr}\{WP_A P_B^T\},$$

(2)

where $W$ is a diagonal matrix with entries $\{w_1, w_2, \ldots, w_n\}$. An object could be a food resource or a predator. Thus, comprehension of A by B can be beneficial to B, such as when an individual attends to an honest alarm call. In the evolutionary language game $F(L_A, L_B)$ is interpreted as the payoff gained by B listening to A. The greater the payoff to an individual the more likely its lexicon will spread through imitation to other members of the population.
3. Communication Error

Communication may of course be subject to error, where sounds are mistaken for another. We term this sort of error communication error. Define $E$ as an $m \times m$ communication error matrix, where $e_{ij}$ is the probability that sound $i$ is heard as sound $j$. The fitness function is now given by

$$F(L_A, L_B) = \text{tr}\{WP_AP_B^T\}. \quad (3)$$

If $s_{ij}$ measures the similarity between sound $i$ and sound $j$ where $0 < s_{ij} < 1$, then let $\bar{e}_{ij} = s_{ij}/\sum_{k=1}^{m}s_{ik}$. Thus, communication error increases with the number of sounds $m$. Let $s_{ij} = 1$ if $i = j$ and $\sigma$ otherwise, such that the probability of correctly hearing a sound $\bar{e}_{ii} = 1/[1 + (m - 1)\sigma]$. If there is a single lexicon $L_A$ in the population then fitness is given by

$$F(L_A, L_A) = \text{tr}\{WE\} = [1 + (m - 1)\sigma]^{-1}\sum_{i=1}^{m}w_i \quad (4)$$

for $m = n$. If we allow $m$ to vary and rank objects according to their value such that $w_1 > w_2 > w_3 \ldots$ and $w_i = 0$ for $i > m$, then for non-zero $\sigma$, the fitness $F$ is found to have a maximum for some $m$ where $m < n$ [Fig. 1(a)]. In other words, there is a limit on the number of sounds in a lexicon, above which having additional sounds causes confusion and reduces overall fitness. This has been referred to as a “pre-linguistic error limit” (Nowak & Krakauer, 1999). Consider the following example, using a simulation scheme as described in detail by Nowak et al. (1999). A population of $N = 900$ individuals is initiated, each with the same lexicon of $m = n = 50$ sounds. The probability that an individual in the next generation imitates the complete lexicon from a given individual in the preceding generation is then determined by that individual’s relative payoff determined during the communication round. As is shown in Nowak et al. (1999) this is not qualitatively different to imitation of single signs. The payoff of an individual with lexicon $L_v$ is given by

$$F_v = \sum_{i=1}^{x}F(L_i, L_v) \quad (5)$$

summing over each of the $x$ randomly chosen individuals who communicate to $v$, and where $L_i$ is the lexicon of the $i$-th individual. The relative imitative fitness is simply an individual’s payoff divided by the summed individual payoffs for the entire population. Thus, if an individual has a relative imitative fitness of $0.1$, then in the next generation each individual will imitate that individual’s lexicon with probability $0.1$. In these simulations, $\sigma = 0.4$ (the reason for this is made clear in Section 5). Signs are transmitted with a small probability that a sound is lost (through not being heard) or gained (through being mis-heard). After each loss or gain the $P$ matrix is re-normalized. Thus, the number of sounds in the lexicon can change through time as a result of errors in transmission, and is expected to
approach the optimum predicted by eqn (4). This is shown for different values of communication error ($\sigma$) in Fig. 1(b).

4. Transmission Error

As we have noted above, propagation of a lexicon is subject to transmission errors as sound can be added to the lexicon or existing sound–object association can be modified through mis-hearing a signal. These are errors in learning signs, which play a constructive role in establishing a conventional sign system. In this section, we provide an analytical framework for understanding transmission error. In the presence of transmission error each entry in the $P$ matrix of a lexicon, with a small probability ($\mu$), is imitated with error. This error can involve the re-referencing of a sound in a consistent lexicon, or the altering of the matrix entry by a small, uniformly distributed random number in the case of an inconsistent lexicon. One can see that transmission error is essential to the evolution of shared sign system. In the presence of transmission error a population of individuals with random, inconsistent signs can evolve towards a population with a single shared, consistent lexicon (Nowk & Krakauer, 1999).

Although transmission error is essential for the coordination of a shared lexicon, it also acts to restrict the number of sounds in the lexicon. A balance is reached between selection amongst individuals favouring a lexicon, and transmission error eroding the lexicon. This balance is analogous to the mutation–selection balance seen in genetic evolution (Burger, 1998; Haldane, 1937).

Consider a large population where each individual possess the same consistent lexicon $L_1$. We introduce transmission error such that $\mu$ is the probability, assumed to be small, that an entry in $P$ is transmitted with error, thus generating a new lexicon, $L_i$. In this case, we can assume that the inequality $F_1 > F_i$ for all $i \neq 1$ will hold at any time (the population is in transmission error–selection balance). Let $f = f_1, f_2, \ldots, f_k$ be the frequency of each lexicon in the population, where there are $k$ possible lexica. In the case, $\mu = 0$ the population will be in a globally stable equilibrium $f^0 = \pm 1, 0, \ldots, 0$. A population with small $\mu$ will be in a locally stable equilibrium $f^\mu$ which satisfies $f^0 - f^\mu = O(\mu)$ (cf. Burger, 1998). Therefore, to the first order the expected fitness of a randomly chosen member of the population is given by

$$E(F_0) \approx (1 - mn\mu) \times F_1$$

since there are $mn$ opportunities for mutation away from $L_1$ at transmission. Thus, fitness is reduced by transmission error pushing lexica away from $L_1$. This further limit imposed on the number of sounds in a lexicon by transmission error is shown in Fig. 2.
5. Population Structure

So far we have only considered the evolution of shared signs in a population where every individual has an equal probability of communicating with every other individual. The social structure of a population and its distribution over a landscape make this assumption unrealistic. In this section, we introduce population structure into the game theoretic models.

Imagine a structured population in which the N individuals occupy the vertices of an X × Y grid, where X × Y = N and opposing edges of the grid are joined (it is a torus). Each individual plays a language game with its eight neighbours, the “Moore” neighbourhood (Durrett & Levin, 1994). Neighbourhood can reflect space in the conventional sense, or social structure such that “neighbours” are those individuals who are in positions within a social group such that they regularly and exclusively communicate. After each round of communication, if an individual has a greater payoff than its neighbour, then its vertex becomes occupied by the lexicon of this neighbour.

We firstly examine the dynamics in a well-mixed system with a coordination number equivalent to that of the Moore neighbourhood. Consider a population with two sign systems, L_A and L_B, and no transmission error. In a non-structured (“panmictic”) population, the fitness of a chosen individual v is given by eqn (5). Let x = 8 (as for structured population), and define the two lexica, L_A and L_B, such that $F(L_A, L_A) = \psi_{AA}$, $F(L_A, L_B) = \psi_{AB}$, etc. The expected fitness of an individual given that it has lexicon L_A is given by

$$E(F_A) = 8\psi_{AA} f_A^8 + \binom{8}{7} (7\psi_{AA} + \psi_{BA}) f_A^7 f_B + \binom{8}{6} (6\psi_{AA} + 2\psi_{BA}) f_A^6 f_B^2 \cdots + 8\psi_{BA} f_B^8,$$

where $f_A$ is the frequency of lexicon L_A in the population and $f_B$ the frequency of lexicon L_B.

In the special case, where $\psi_{AA} = \psi_{BB} = 1$ and $\psi_{AB} = \psi_{BA} = 0$ then this expected fitness is binomially distributed with expectation $E(F_A) = 8f_A$, and variance $V(F_A) = 8f_A(1 - f_A)$. If we assume N is large then the average fitness of a member of the population is given by

$$\bar{F} = f_A E(F_A) + f_B E(F_B) = 8(f_A^2 + (1 - f_A)^2).$$

Thus, the expected fitness of an average member of the population increases as $f_A$ tends from 0 to 1. The expected frequency of lexicon $L_A$ in the next generation is much harder to calculate, as it involves summing over all possible combinations of eight individuals for each communication event to calculate the expected overall “fitness” of lexicon $L_A$. However, if we ignore variation in the communication process by allowing the number of individuals each individual communicates with to tend to $\infty$, then the expected frequency of lexicon A in the next generation is

$$E(f_A(t+1)) = \frac{f_A(t) E(F_A)}{f_A(t) E(F_A) + f_B(t) E(F_B)}$$

$$= \frac{f_A(t)^2}{f_A(t)^2 + (1 - f_A(t))^2},$$

where $f_A(t)$ is the frequency of lexicon $L_A$ at time t. This can be plotted recursively for several generations (Fig. 3), and clearly demonstrates the rapid evolution of a bi-lexical population to one with a single lexicon (where $f_A = 0$ or 1) when the starting frequency $f_A(0)$ deviates from 0.5. In other words, once a lexicon begins to dominate in a non-structured population, positive frequency-dependent selection causes the population to very quickly become mono-lexical for that lexicon. Thus, the numerically dominant sign system acts as a “language steamroller” flattening out the signal diversity in the population.

In an explicitly structured population this loss of lexical diversity does not occur since it is only necessary for an individual to have more lexically identical neighbours than the neighbours themselves to form a stable (un-invadable) linguistic community. Indefinite coexistence of independent sign systems becomes possible. The argument appears to apply to populations where there are at least 2 or more alternative sign systems. However, this is analytically difficult to prove. Instead, we simulated the evolutionary language game
FIG. 3. The expected frequency of lexicon A after 0, 1, ..., 4 generations vs. the starting frequency $f_A(0)$ of that lexicon in the case where the number of individuals communicated with is $N$ and $N \rightarrow \infty$.

100 times with and without a structured population of size $N = 900$ initiated with individuals possessing random consistent sign systems. After 50 generations the number of sign systems typically reached an equilibrium and was noted. In the non-structured case at equilibrium the population was always mono-lexical, whereas for the structured population the average number of sign systems was 20.92 (S.D. = 1.55). In the presence of transmission error ($k = 10^{-3}$) overall lexical diversity was also higher in the structured population with a mean number of sign systems given by 33.45 (S.D. = 3.96), compared to the non-structured population where the mean was 9.64 (S.D. = 1.62). Thus, error creates transient diversity even in the non-structured case, analogous to a quasispecies distribution of mutant genes around a single wild type.

6. Deception

Natural selection results in the differential survival of sign systems according to the relative payoffs of the individuals possessing them. Thus, any strategy that can be acquired along with the lexicon and which augments an individual’s payoffs will spread. Deception is common to almost all natural communication systems, from warning colouration to primate vocalizations. A long-standing question in behavioural ecology has been how honesty in natural populations can be maintained. Under some restricted conditions, cost can ensure honesty (Zahavi, 1977; Grafen, 1990). One property of arbitrary sign systems is that costs of production are very low, and hence conventional handicap style arguments do not seem pertinent. We consider how deception will spread in the evolutionary language game, where deception implies transmitting false information in order to profit from comprehension. We then explore more plausible, cost-free mechanisms for stabilizing honesty.

To incorporate deceit, the payoff from comprehension received by listener B [eqn (3)] is reduced by a factor $0 < t_A < 1$, the truthfulness of speaker A, which determines the probability that A signals honestly. The payoff received by B listening to speaker A is now given by

$$G(L_A, L_B) = t_A \times F(L_A, L_B).$$

We define an “honest” individual as one where $t_A = 1$, and all other individuals “deceivers”, where deceivers lie in all encounters.

Consider a non-structured population with a single, consistent lexicon $L$ in the absence of transmission error and where $F(L, L) = 1$. This lexicon may be associated with honesty $L_H$ or deceive $L_D$. If we assume, as before, that the population size $N$ is large then the expected fitness of an individual is given by

$$E(G_{H or D}) = 8 f_H^8 + \binom{8}{7} (7 + t) f_H^7 (1 - f_H)$$

$$+ \binom{8}{6} (6 + 2t) f_H^6 (1 - f_H)^2$$

$$+ \cdots + 8t (1 - f_H)^8,$$

where all deceivers have truthfulness $t$ and $f_H$ is the frequency of honesty in the population [cf. eqn (7)]. Since an individual’s truthfulness does not have an impact on its own payoff, there is no pressure for honesty in communication. Truthfulness values can drift and the strength of selection for a shared lexicon is reduced. Perhaps more realistically deceit may be rewarded with a fitness pay-off proportional to the frequency of deceitful utterances $(1 - t)$. For example, communication may be about a food resource that the speaker could utilize were the listener to be
misinformed. If $b$ is the coefficient of proportionality determining the relative benefit from deceit then

$$E(G_D) = E(G_H) + 8b(1 - t) \quad (12)$$

since each individual listens to eight others. Clearly, deceit will always spread in the case that $1/8 > b > 0$. Furthermore, the fitness received by a deceitful individual is maximized when $t = 0$. If the value of $t$ is transmitted with error then it will evolve in the population towards this optimal value and there will no longer be selection for a shared lexicon. Therefore, in a non-structured population deceit, even if only marginally beneficial to the speaker, undermines lexicon formation.

In a structured population the spread of deception or honesty is influenced by the fact that individuals are more likely to communicate with their imitative “offspring”. Honesty therefore has an inherent advantage (Krakauer & Pagel, 1995; Nowak & May, 1992). Space has the effect of creating local pockets of high relatedness (Van Ballen & Rand, 1998). The stability of honesty at a given frequency is difficult to determine since it depends on the exact distribution of honest individuals in the structured population. Instead, we determine whether honesty will invade a population of deceivers ($f_H \approx 0$), by considering whether the neighbours of an honest individual become honest, given that they are currently deceitful. In the case that only one individual is honest, honesty will not spread since the honest individual will encounter only deceit, gaining a fitness payoff of $8t$, whilst its deceitful neighbours will gain $1 + 7t + 8b(1 - t)$. The latter payoff is always greater than $8t$ except when $t = 1$, in which case the neighbours are honest as defined.

If a $2 \times 2$ group of honest individuals arises through chance, the increase in the number of honest neighbours allows honesty to spread under certain conditions. The fitness of a member of the honest community is simply $G_H = 3 + 5t$, since each individual has three honest and five deceitful neighbours (see Fig. 4). The fitness of a deceiver is largest when it has the maximum possible number of honest neighbours, which is 2. The fitness payoff for such a deceiver is $G_D = 2 + 6t + 8b(1 - t)$. For honesty to invade we require $G_H > G_D$, which results in the constraint that $b < 0.125$. If honesty expands beyond a $2 \times 2$ community, then for further expansion the same constraint must hold for expansion to the boundary corners (position $a$ in Fig. 4), whilst at the edges (position $b$) the constraint is relaxed such that $b < 0.25$.

Therefore, in contrast to a non-structured population honesty can invade and remain stable in a structured population, as long as the reward from deceitful behaviour is not too great ($0 \leq b < 0.125$). In this way, a structured population can maintain a shared lexicon where it would be lost by a non-structured population. This result will hold for different arrangements of individuals within a structured population, although with different constraints on $b$, as long as individuals are more likely to communicate with their imitative descendents.

A further buffer against deceit has been explored by Nettle & Dunbar (1997), where high rates of change within the sign system give cheats a moving target. Thus, “mutability” becomes in effect a strategy for escaping from the costs of dishonesty. Such a strategy is evidently constrained by the need to maintain coordination, spatial structure promotes such coordination.
7. Discussion

Using simple game theoretic models of referential signalling we have explored the influence of two sources of error on the “evolution” of an imitation-propagated sign system. Transmission errors (mistakes during sign acquisition) and communication errors (mistakes in the perception of signs) both limit the maximum number of sounds in a lexicon. Larger sign systems (greater $m$) are unlikely to be acquired without error, and an increased number of sounds promotes the probability of mis-identifying an individual sound. It is therefore expected that organisms with error prone communication mechanisms (communication error) or poor learning ability (transmission error) will have sign systems of a restricted size with the ability to refer to only a few objects. Furthermore, within a population, all individuals are observed to employ the same conventional lexicon. If the rates of transmission error are large, then transient lexical diversity is observed, reflecting non-competent sign usage. Thus, one might argue that these two sources of error have acted historically to both facilitate the emergence of referential sign systems and subsequently imposed upper bounds on their diversity. Given the near inevitability of such errors it remains somewhat mysterious that discrete, referential signs should be so rarely reported. It is certainly not correct to argue that signal usage is restricted to those species “smart” enough to use them. Dogs, parrots (Farrabaugh & Dooling, 1996), dolphins (Reiss & McCowan, 1993) and apes can all be taught new arbitrary signals relatively easily up until some small number. And yet these species do not appear to use this capability naturally. It is instructive to note that primate sign use has relied on an artificial technology (computer-assisted, iconic sign systems) in order to teach arbitrary symbol–object associations (Savage–Rumbaugh, 1992). The explanation does appear to be the need to keep the signal clean and thereby easy to discriminate. The X-ray analysis of the vocal tracts of apes shows that the larynx is positioned towards the back of the mouth, high in comparison to humans. This arrangement limits the range of vowel sounds that can be produced. Lieberman (1992) has suggested that animal signals have been largely constrained by the inability to produce clear signals, and that an expanded vocal repertoire was the essential pre-adaptation for the evolution of articulate signals. Of course, it could be some independent factor that promoted communication, and the vocal tract evolved as an adaptation in response to this pressure. Nonetheless, the modern supralaryngeal vocal tract is the central engineering tool for producing language. Thus, Liberman’s theory makes perfecting of human speech an integral element of human language. A position consistent with the importance errors play in constraining sign systems in our models.

In contemporary humans there are specialized neuro-anatomical structures serving language production. Broca’s area, involved in the construction of sentences in H. sapiens, appears in endocasts of fossil australopithecine crania 2.5 Myr ago, and is clearly present together with Wernicke’s area in fossil H. habilis from 2 Myr ago (Deacon, 1992). The neural mechanisms and structures involved in the production of articulate speech are thought to have arisen about 1.6 Myr (Lieberman, 1992). It is as recent as 50 000 years ago that reasonable archaeological evidence of language arises, in the form of complex tool technology and signs of conceptualization and planning, trade, art and ritual (Holden, 1998). Thus, biological traits correlated with speech production have arisen and undergone systematic change. This would suggest that speech itself has undergone systematic change. Hence, signals and anatomy appear to have “co-evolved” in order to improve the production and perception of signals.

Using our models we find that social or geographic structuring of populations promotes the stability of small sign communities and consequently fosters lexical diversity. Each sign community can refer to different objects using a common set of sounds. Thus, if objects are distributed heterogeneously through the environment a structured, poly-lexical population is able to communicate about a larger number of objects than a mono-lexical, panmictic one. This is analogous to the phenotypic specializations seen within populations responding to a diversity of ecological niches. The explanation for the increased diversity is that no single sign community...
exceeds the linguistic error threshold. An examination of human language, and indeed animal signals, testifies to the important role of spatial segregation in allowing for signal diversity. The average community size in our models is determined exclusively by the payoffs derived from the lexicon. Within contemporary human society there are many other factors at play, including at the very least economic and military power (Cavalli-Sforza & Cavalli-Sforza, 1995).

Deceit erodes shared sign systems by selecting in favour of receivers who choose not to attend to signals. Thus, in panmictic populations, the prevailing lexicon will only persist if signallers and receiver cooperate (as is assumed in Nowak & Krakauer, 1999). This is unlikely to always be the case. A reliable source of profit is the gullibility of receivers. Hence, “crying wolf” is expected to evolve in order to manipulate the “beliefs” of receivers (Dawkins & Krebs, 1978). Where we observe sign usage naturally, among vervets, bees or humans, there is evidence of persistence of signals. How are these maintained in the face of fundamental instability? Non-reference, affective/emotional animal signals can be maintained by cost. Hence, reliability is ensured by making deception pay a disproportionate penalty (Zahavi, 1977; Grafen, 1990). There is little evidence of a high cost of “words” in human language. Indeed, it is very difficult to derive the veracity of propositions based on signal content alone. There must be an additional means of penalizing dishonesty. Population structure can ensure the stability of sign systems even in the face of deceit. This is because payoffs are determined by local interactions thereby promoting the emergence of cooperative or honest sign systems through a mechanism of repeated encounters. In our model these payoffs are the immediate spatial neighbours on a lattice. Some memory of interlocutors can replace space in a panmictic population, through a strategy of image scoring (Nowak & Sigmund, 1998). Both mechanisms will allow a sign system to evolve through repeated interactions with the same individuals.

Many thanks to the two anonymous referees for their constructive comments, Martin Nowak, Muna Dawoodi, Lolo Burschel, Roland Fleißner and Hind Khalifeh for stimulating discussion. N.G. and A.v.H were supported by the Deutsche Forschungsgemeinschaft. D.C.K. gratefully acknowledges support from The Alfred P Sloan Foundation, The Florence Gould Foundation, The Ambrose Monell Foundation and the J. Seward Johnson Trust.

REFERENCES


