THE ROYAL SOCIETY

Received 24 April 2002
Accepted 6 August 2002
Published online 20 November 2002

Common language or Tower of Babel? On the evolutionary dynamics of signals and their meanings

Minus van Baalen[1*] and Vincent A. A. Jansen[2]

[1]*Institut d'Écologie, UMR 7625, Université Pierre et Marie Curie, Bât. A, 7ème Étage CC237, 7 quai St Bernard, 75252 Paris Cedex 05, France*
[2]*School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK* (*vincent.jansen@rhul.ac.uk*)

We investigate how the evolution of communication strategies affects signal credibility when there is common interest as well as a conflict between communicating individuals. Taking alarm calls as an example, we show that if the temptation to cheat is low, a single signal is used in the population. If the temptation increases cheaters will erode the credibility of a signal, and an honest mutant using a different signal ('a private code') will be very successful until this, in turn, is cracked by cheaters. In such a system, signal use fluctuates in time and space and hence the meaning of a given signal is not constant. When the temptation to cheat is too large, no honest communication can maintain itself in a Tower of Babel of many signals. We discuss our analysis in the light of the Green Beard mechanism for the evolution of altruism.

Keywords: evolution; communication; language; cooperation; conflict; Green Beards

1. THE EVOLUTION OF SYMBOLIC COMMUNICATION

Symbolic communication requires the establishment of an agreed vocabulary, as the transfer of information only works if there is a clear set of associations between (external) objects and the symbols that represent them. A central aspect in the evolution of symbolic communication, or language, is how symbols acquire their meaning. In a series of articles (Nowak & Krakauer 1999; Nowak *et al.* 1999*a*,*b*, 2000), Krakauer, Nowak and co-workers developed an evolutionary approach to study how initially arbitrary symbols may get associated to particular objects.

Underlying this approach is the hypothesis that the benefit associated with successful communication is fully and equally shared by the communicators (Grassly *et al.* 2000; Lachmann *et al.* 2001). When communicators have an identical stake in the success of information transfer, natural selection will favour the evolution of communication channels that convey messages as efficiently and unambiguously as possible (Krebs & Dawkins 1984; Noble 1999). However, it is less clear what will happen when the interest of sender and receiver do not fully overlap, because then sender and receiver need not necessarily 'agree' about what the signal means (Godfray 1991; Johnstone & Grafen 1992; Grafen & Johnstone 1993; Maynard Smith 1994; Noble 1999, 2000; Grassly *et al.* 2000; Lachmann *et al.* 2000, 2001; Noble *et al.* 2001).

To understand how signal/meaning associations (i.e. symbols or words) arise, one needs to separate intention from interpretation (Grafen & Johnstone 1993). If receivers do not interpret the symbol the way intended by the sender, then it no longer pays the sender to use this particular signal. Because individuals in any interaction almost inevitably have diverging interests (Van Baalen & Jansen 2001) selfish usage of signals may therefore lead to the erosion of the meaning of signals (Grassly *et al.* 2000).

But the potential advantage of successful communication remains. This implies that communicators that have agreed upon a different signal (a 'private code') might accrue a considerable benefit. This advantage may be important, in particular if cheating has reduced the ecological performance of the resident population. When the new signal becomes sufficiently common, cheaters that exploit this signal ('the private code gets cracked') have a selective advantage and can invade in turn.

This process could thus lead to the unstable coexistence of signalling strategies. Then, the dynamics of the meaning of a suite of signals will fluctuate, and symbols may continuously acquire and lose meaning, reflecting the dynamics of honest use and cheating in the system. Most theories being based on the assumption of evolutionary stability, this possibility has not received much attention (but see Noble *et al.* 2001). Many studies on signal use have assumed that information is transferred by varying some continuous character (Grafen 1990; Zahavi 1993; Krakauer & Pagel 1995). In this case the 'intensity' of the signal rather than the choice of signal conveys the meaning. Here, we study the evolution of communication when information can be carried by qualitatively discrete (rather than continuous) sets of possible signals (Grassly *et al.* 2000).

We will explore how the resulting pattern of signal use and misuse depends on the benefits associated with efficient communication as well as on the temptation to cheat. We will formulate a model based on the analysis by Krakauer & Pagel (1995) that allows signals to be used dishonestly. In contrast with previous approaches for the study of communication under partial alignment of interest (Krakauer & Pagel 1995; Noble 1999, 2000; Lachmann *et al.* 2001), we investigate the case where the

*Author for correspondence (mvbaalen@snv.jussieu.fr).

Table 1. Payoffs per communication event in the Cry Wolf game (payoff to sender payoff to receiver).

	receiver	
sender	gullible	suspicious
honest	$(0,R)$	$(0,-P)$
dishonest	$(T,-S)$	$(0,-P)$

problem of the sender is not so much choosing the *level* of the signal ('how loud to speak') but rather *which* symbol from a given set to use ('in which language'). We thus assume that signals are essentially cost-free, so that honest communication (if it arises) is not maintained by Zahavi's handicap principle (Zahavi 1977; Grafen 1990).

Communication has been frequently observed to evolve in simulations involving various mixtures of conflict and cooperation (see Ackley & Littman 1994; Oliphant 1996; Grim *et al.* 1999; Wagner 2000; Reggia *et al.* 2001). Oliphant (1996) observed that whether the population settled at a single common communication system or whether it flipped back and forth between two different systems depended on whether or not successful communication affected the sender's fitness. This shows that the conflict of interest between sender and receiver is important, but gives little insight as to when one outcome gives way to the other. Here, we will extend these studies by formulating an explicitly spatial, individual-based model for the population dynamics of communicating individuals. We will show the results of some simulations of the corresponding probabilistic cellular automaton (PCA) to illustrate the possible spatio-temporal outcomes. The novel aspect of our study is that we will derive and analyse a correlation dynamics version of the model (Matsuda *et al.* 1992; Van Baalen & Rand 1998; Van Baalen 2000), which allows us to carry out a bifurcation analysis to assess the robustness of our results and obtain insight into how the dynamics of signal credibilities depends on selfish and common interest.

2. THE 'CRY WOLF' GAME

Suppose that an individual can warn another individual whenever it perceives that the latter is in danger. Sending a warning signal comes at no cost to the sender, but permits the receiver (if it reacts timely) to accrue a benefit (which we will denote by R) instead of paying the cost of not being warned (P). The sender can also cheat by giving false warnings; whenever the receiver heeds such a false warning, the sender will gain a benefit (T) whereas the receiver will pay a cost (S). Thus, it pays to heed the signal if it is given honestly, but the only direct benefit to the sender is associated with cheating (table 1). Whether it pays to warn depends on the common interest of sender and receiver and on their degree of relatedness.

We have adopted the well-known Prisoner's Dilemma game naming scheme, where R is 'reward', P is 'punishment', T 'is temptation' and S is the 'sucker's payoff' (Axelrod & Hamilton 1981; Axelrod 1984). However, note that in some aspects the Cry Wolf game is fundamentally different. First, in the Cry Wolf game there is an inherent asymmetry (first the sender emits the signal, then the receiver interprets and acts), whereas the roles in the Prisoner's Dilemma game are symmetric where both players make their moves simultaneously (but see Nowak & Sigmund 1994). Note also that, in deviation of what is customary in game theory, we indicate costs by minus signs (so that P and S themselves are positive parameters).

Table 2. Average payoffs in the Cry Wolf game if occasions for honest signalling arise with frequency τ, and dishonest signals are sent with frequency ϕ (average payoff to sender, average payoff to receiver).

	receiver	
sender	gullible	suspicious
honest	$(0,\tau R)$	$(0,-\tau P)$
dishonest	$(\phi T, -\tau P - \phi S)$	$(0,-\tau P)$

Second, the Cry Wolf game requires variation in the external world such that information is known to one player but not to the other. In our model this information is the presence of danger to the receiver. We denote the frequency with which this danger occurs as τ. Occasions where the sender can cheat occur with a frequency ϕ. The receiver perceives the signal but does not know whether it is honest or not. Assuming that cheaters never signal honestly, the average payoffs to sender and receiver can be calculated easily and are given in table 2. Note that a cheated gullible receiver pays the double cost of the punishment *and* the sucker's payoff when it is cheated. Underlying all of our analyses is the assumption that a cheater signals with a frequency that is equal to that of an honest signaller (i.e. $\phi = \tau$), so that the rate of signalling alone does not give information about the signaller's intentions. If cheaters signal more often, receivers could try to measure signalling frequencies of the individuals they are dealing with and use these estimations as an indication.

In principle, individuals can cheat or be honest independent of how they interpret information. This could give rise to the strategies gullible/cheat and suspicious/honest. The latter combination can never invade. The former can invade through kin selection only if the combined gains of a gullible/cheat couple ($T - S$) exceeds the gains of a gullible/honest couple (R). This strategy could therefore exploit a relation based on an alternation of cheating and being cheated. We did not include the analysis of these paradoxical strategies in our model.

In the basic game in neither role will a player let its decisions depend on what is gained by its partner; there is no common interest shared by sender and receiver. When individuals interact with individuals taken at random from the entire population cheating is the only evolutionarily stable strategy.

3. COMMON INTEREST

Selfish interests always favour cheating. However, if sender and receiver are related, their relatedness implies a common interest which profoundly changes the game as the sender has to take the consequences of its actions for

Table 3. Symbols and their default values (if applicable).

class	symbol	description	default value
interaction parameters	R	reward	1
	P	punishment	0
	T	temptation	2
	S	sucker's payoff	0.1
	τ	opportunities for honest signalling	0.5
	ϕ	(potential) rate of cheating	0.5
demography	b	birth rate	1.5
	d_0	background mortality rate	1
	m	movement rate	5
	μ_{int}	mutation rate (intention)	0.001
	μ_{sig}	mutation rate (signal use)	0.001
space		lattice size	4900
	n	nearest neighbours	6
dynamic variables	p_i	global densities	
	$q_{i\|j}$	local densities	
	F	credibility	

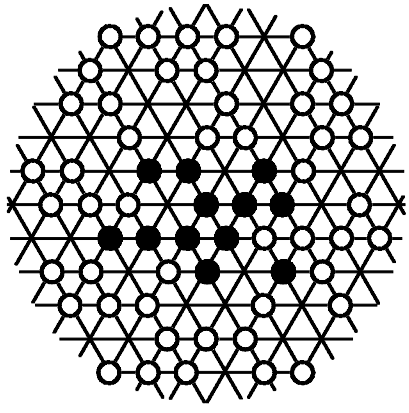

Figure 1. Example of a cluster of mutants (black) invading a resident population (white) on a triangular lattice. The lattice that we use for the simulations is a triangular lattice like this one, of 70 × 70 sites with periodic boundary conditions.

the receiver into account. Prolonged or repeated interaction between the same individuals also creates a common interest with potentially similar effects (Axelrod & Hamilton 1981; Van Baalen & Jansen 2001).

To work out the invasion chances of a new communication strategy one needs to take into account the environment that users of such a new strategy find themselves in. In a large well-mixed population such a new strategy is unlikely to encounter others with the same strategy, but in a spatially structured population individuals tend to be surrounded by relatives having the same strategy. In well-mixed populations, the probability of encountering a particular strategy is proportional to its frequency in the population, but in spatially or socially structured populations, local densities may be quite different from global densities (Matsuda *et al.* 1992; Van Baalen 2000). In our analysis we will assume that individuals live on a triangular lattice where they interact with at most six nearest neighbours, and that all demographic processes are local (figure 1).

We assume that initially all individuals use the same signal, but some (denoted H) are honest and gullible, whereas others are dishonest and suspicious (cheaters, denoted C). We will denote the probability that a site next to an individual of type i ($i = H, C$) is empty by $q_{\emptyset|i}$, and the probability that it is occupied by an individual of type j by $q_{j|i}$. Global densities, i.e. the probabilities to find a randomly chosen site occupied by an individual of type i, are denoted by p_i. Both global and local densities are dynamic variables (Matsuda *et al.* 1992; Van Baalen 2000), so that the social structure experienced by the different strategies is not imposed by the modeller, but is a direct consequence of the dynamics of the system.

We assume overlapping generations so that births, deaths and movement occur in continuous time. For simplicity, we will assume that costs and benefits affect mortality only. The mortality rate of the average honest individual ($d(H)$) is then given by

$$d(H) = d_0 - (-q_{\emptyset|H}\tau P + q_{H|H}\tau R - q_{C|H}(\tau P + \phi S)),$$

where d_0 is the background mortality rate. Because $q_{\emptyset|i} + q_{H|i} + q_{C|i} = 1$ for any i, this is equivalent to

$$d(H) = d_0 + \tau P - q_{H|H}\tau(P + R) + q_{C|H}\phi S.$$

Note that costs increase mortality while benefits decrease it. Note also that there is a price to pay for being located next to empty sites because there is no one there to issue warnings.

Similarly, the mortality rate of the average cheater, $d(C)$, is

$$d(C) = d_0 + \tau P - q_{H|C}\phi T.$$

Together with assumptions about birth rates (a fixed birth rate b is assumed) and dispersal across the lattice (individuals are assumed to move from site to site with a constant rate m which is the same for both strategies) this defines a so-called asynchronously updated PCA or, more precisely, a 'spatial ecology' as it includes movement of individuals (Rand 1999; Van Baalen 2000).

Before analysing the consequences of social structure it is insightful to consider what will happen in a well-mixed system without social structure. Then all local densities $q_{i|j}$ are equal to the global density p_i, the equations for mortality rate become

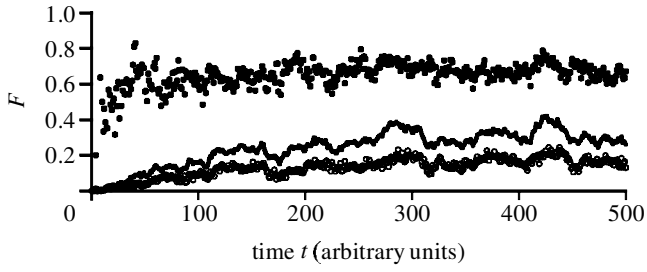

Figure 2. Absolute (drawn curve) and experienced credibilities of the signal (filled circles, by honest users; open circles, by cheaters) in a typical spatially explicit simulation with a single signal (with parameters as in table 3).

$$d(H) = d_0 + \tau P - p_H\tau(P + R) + p_C\phi S$$

and

$$d(C) = d_0 + \tau P - p_H\phi T.$$

As there are no other differences between the strategies, honest signallers will invade if $d(H) < d(C)$, from which we can derive the invasion threshold

$$p_H > \frac{\phi S}{\tau(P + R) - \phi T} p_C,$$

which is positive if potential benefits are sufficiently large (if $\tau(P + R) < \phi T$ honest signallers can never invade). This threshold increases with increases in S and T and decreases with increases in R and, perhaps less intuitively, increases in P (but remember that honest signallers prevent each other from being punished which is a relative benefit as cheaters are always punished).

Rare honest signallers cannot invade because there is nobody to signal with. However, the fact that there is an invasion threshold indicates that invasion conditions are less stringent in spatially or socially structured populations. In spatially explicit simulations of this system we found that honest signallers can invade and coexist with the cheating signallers where they cannot in homogeneously mixed populations. When clusters of honest signallers are small and tightly connected they can support each other but when these clusters grow too big they can be exploited by cheaters (see Van Baalen & Rand 1998).

When honest signallers occur in clusters, the average honest signaller has more honest signallers in its environment than does the average cheater. Thus, the stream of signals received by the average honest signaller will contain more honest information. This can be formalized by defining the credibility F of the signal as the proportion of all signals that have been sent with honest intentions,

$$F = \frac{\tau p_H}{\tau p_H + \phi p_C},$$

or, as received by a specific receiver i,

$$F_i = \frac{\tau q_{H|i}}{\tau q_{H|i} + \phi q_{C|i}}.$$

Running explicit simulations of our system, we found that under certain conditions honest signallers can persist, as long as the system remains sufficiently clustered. Data obtained from a typical run (figure 2) show why: to the average honest signaller the signal can be much more credible (and hence more profitable) than it is to the average cheater.

4. DISTINCT SIGNALS

Simulations (with parameter values as given in table 3) show that if having accurate information is important, honest signallers can withstand widespread cheating but the information content is severely compromised by the surrounding cheaters (figure 2). Under those conditions, a mutant honest signaller that uses a *different* signal (say signal B instead of the signal that is common usage, say signal A) might thus have a selective advantage since it is immune to the A-using cheaters. However, it also lacks support of the honest A-signallers, so it is not at all immediately obvious that it can always invade.

Of course it is hard to see how new signals can evolve *de novo*. Here, we imagine that there exists a repertoire of (observable) actions that in principle can be used as signals. The emergence of a new signal then involves the use of that signal as an alarm call.

We denote honest signallers by a capital letter (A,B) and cheaters of a signal by a lowercase letter (a,b). If signal B is rare, the population dynamical equilibrium of the resident A-users will be given by

$$\frac{\mathrm{d}p_A}{\mathrm{d}t} = \frac{\mathrm{d}p_a}{\mathrm{d}t} = 0.$$

In this mixture of honest and cheating users, for both types birth rates will equal death rates. For the honest signallers net birth rate equalling net mortality rate implies

$$bq_{\emptyset|A} = d_0 + \tau P - q_{A|A}\tau(P + R) + q_{a|A}\phi S,$$

where $q_{\emptyset|A}$, $q_{A|A}$ and $q_{a|A}$ reflect the local densities they experience under equilibrium (more about this below). A strain of honest B-users can invade if its birth rate exceeds death rate

$$bq_{\emptyset|B} > d_0 + \tau P - q_{B|B}\tau(P + R).$$

If they experience the same density of empty sites to reproduce in (a reasonable assumption if their background demographic rates are not very different) so that $q_{\emptyset|A} = q_{\emptyset|B}$, then B can invade if

$$q_{B|B}\tau(P + R) > q_{A|A}\tau(P + R) - q_{a|A}\phi S.$$

The mutant has the advantage being free from cheaters. The invasions chances of a new signal therefore strongly depend on the probability that a mutant signaller is surrounded by kin, denoted $q_{B|B}$. This quantity, which is the spatial equivalent of relatedness (Hamilton 1964; Van Baalen & Rand 1998) will vanish in a well mixed population but in a structured population it may be considerable and give an honest mutant the edge over the resident strategies.

For two signals a cycle of invasion events can be found: a first signal (say A) is dominant →A-cheaters invade and degrade its credibility → honest B invades the A–a mixture →B-cheaters invade and degrade its credibility →honest A re-invades. In a simulation with the same parameter combination as in figure 2 but with two distinct signals, cycle-like behaviour (figure 3*a*) can be observed. The cycles correspond to the growth and demise of clusters of

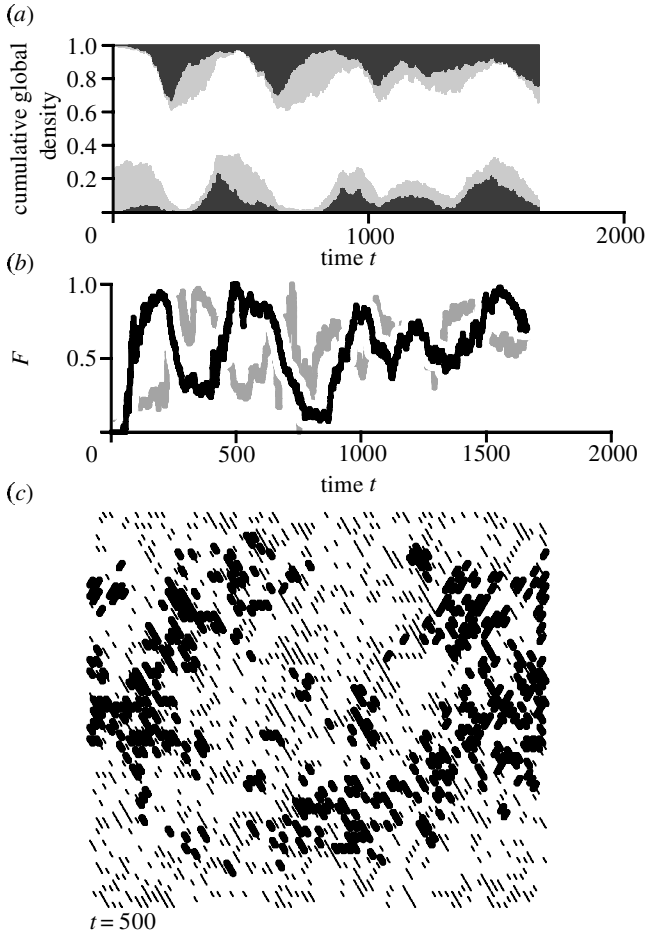

Figure 3. Spatio-temporal dynamics of honest and cheating signallers when two distinct signals are possible. (*a*) Cumulative global densities of *A* (lower two bands) and *B* (upper two bands) signallers (where light shading indicates the densities of cheaters and dark shading the densities of honest signallers) as a function of time. The white band in between represents the density of empty sites. (*b*) How the credibility of signals *A* (grey) and *B* (black) changes through time. (*c*) Snapshot of the lattice showing the distribution of the signalling strategies at $t = 500$. Honest signallers are represented by thick line segments, cheaters by thin line segments; users of signal *A* are indicated by downward slopes (\), users of signal *B* upward slopes (/). Parameter values are as in table 3; time in arbitrary units.

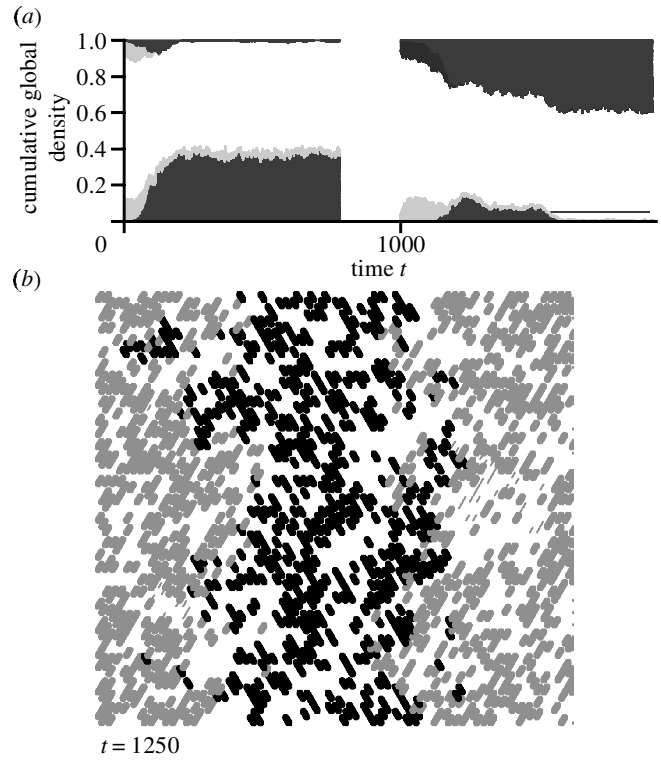

Figure 4. Spatio-temporal dynamics of honest and cheating signallers when the temptation to cheat is much reduced ($T = 1$). Note that the simulation was re-initialized at $t = 1000$. Other parameters are as in table 3; legend as for figure 3 (time in arbitrary units). The simulation lasted for 1664 time units.

A- and *B*-users (figure 3*c*). Clusters grow when honest users invade by virtue of the high credibility of their signal when rare, but then its credibility decreases through the invasion of cheaters (figure 3*b*). If the system were enlarged with many more sites the global densities of the two signals would equalize but the local interaction is likely to remain unstable, with *A*- and *B*-patches continually shifting in space (Jansen & de Roos 2000). Similar cycles and patterns have been observed in simulations of varying complexity, but similar in the sense that they all permitted multiple signals and cheating (Oliphant 1996; Grim *et al.* 1999; Di Paolo 2000).

Simulations of a system with a lower temptation to cheat ($T = 1$) give very different results. Most conspicuously, the system exhibits a much more stable pattern. Initially the system settles at a mixture of two honest populations (with the odd cheater thrown in by mutation). But soon large patches of same-sign use develop (figure 4*b*). Within such patches selection favours conformation to the majority so that the patches only move through random drift effects at the boundaries. Eventually, only one of the signals will disappear (figure 4*a*). However, time to extinction increases rapidly with lattice size so that large systems will tend to settle at large 'frozen' patches of same-signal use. This phenomenon of the formation of large patches is similar to the regional majorities that result in the 'voter' model where individuals cast their vote according to the local majority (Durrett 1988). These patches are much more constant in time than the dynamic pattern that results for higher levels of the temptation for cheating.

5. CORRELATION DYNAMICS

The simulations presented here convey only limited information as they are run only for short periods; by occurring on a small lattice; by being repeated only once and by being run for a single combination of parameters. To find out how robust the conclusions drawn from our simulations are we analysed a corresponding correlation model. Such models are efficient tools for approximating spatial ecologies, especially when running stochastic simulations becomes prohibitive (Matsuda *et al.* 1992; Harada *et al.* 1995; Nakamaru *et al.* 1997; Van Baalen & Rand 1998; Van Baalen 2000).

The principle of this approach is that from the densities of pairs of neighbours (rather than of single individuals)

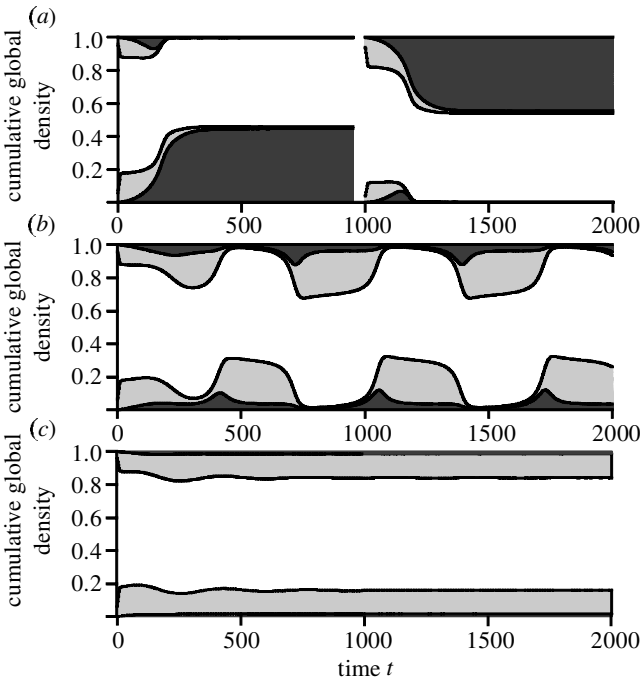

Figure 5. Temporal dynamics of honest signallers (dark shading) and cheaters (light shading), as predicted by the correlation dynamics model for a number of values of the '*T*' parameter: (*a*) $T = 1$ (compare with figure 4*a*), (*b*) $T = 2$ (compare with figure 3*a*) and (*c*) $T = 4$. Legend and parameters as in figure 3 (time in arbitrary units).

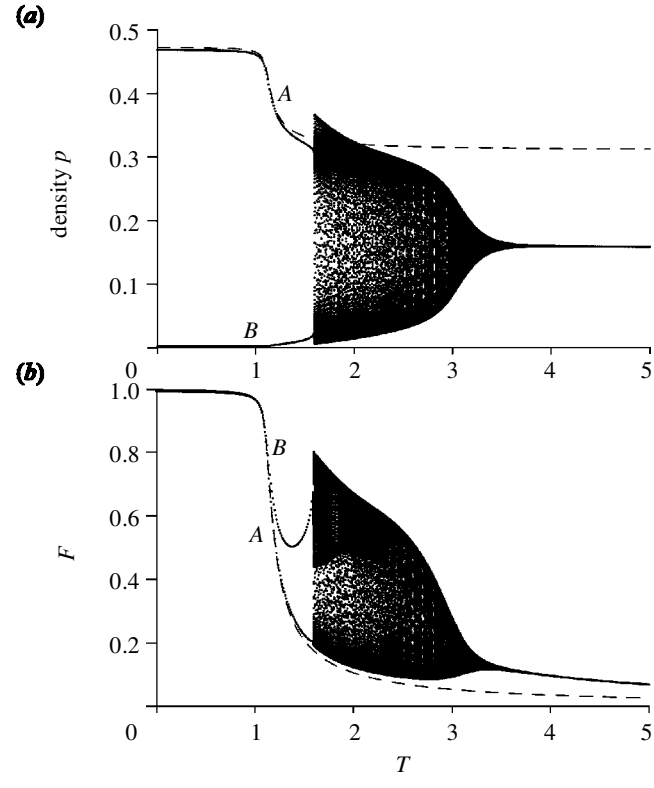

Figure 6. Numerical bifurcation analysis of the correlation dynamics model as a function of the temptation parameter *T*, under initial conditions that favour signal *A*. (*a*) Overall densities of *A*- and *B*-users (dots). (*b*) Overall credibilities of signals *A* and *B*. For low values of *T* the system settles on an equilibrium in which a single signal is used; for intermediate values the system oscillates, whereas for high values an equilibrium consisting of cheaters of the two signals results. The dashed lines represent the equivalent situation without *B*-users. The figures show the results of a numerical analysis in which *T* is varied from 0 to 5 in steps of 0.01, plotting datapoints at $t \in \{2000,2010,...3000\}$. All other parameters as in table 3.

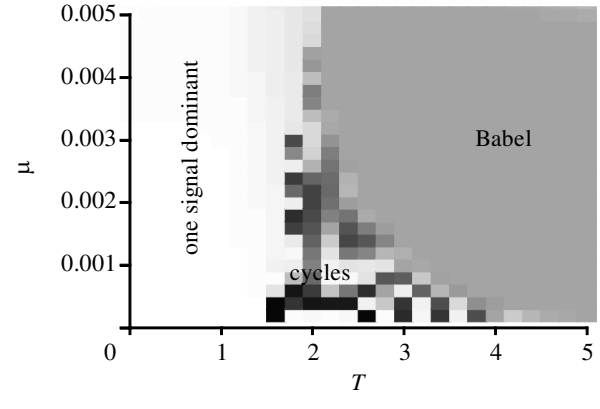

Figure 7. Dynamic outcome of the correlation dynamics as a function of temptation *T* and mutation rate μ. Shading indicates the proportion of *B*-users at time $t = 4000$ for a number of combinations of temptation *T* and mutation rate $\mu_{int} = \mu_{sig} = \mu$. All other parameters as in table 3.

one can obtain information about local environments. The correlation dynamics version of our model (equivalent to the model that we described) is composed of 15 differential equations that keep track of the densities of all possible combinations of neighbours (i.e. the densities of ØØ, Ø*A*, Ø*a*, Ø*B*, Ø*b*, *AA*, *Aa*, *AB*, *Ab*, *aa*, *aB*, *ab*, *BB*, *Bb* and *bb* pairs). Because of space limitations we cannot present the full model here (details available upon request), but it suffices to know that correlation dynamics equations essentially perform a book-keeping of how the densities of these combinations change through the demographic events of births, deaths and movement (Van Baalen 2000). Depending on parameter values, we found that the correlation dynamics version of our model exhibits the same patterns as we observed in the stochastic simulations, from competitive exclusion (figure 5*a*) via limit-cycle-like behaviour (figure 5*b*) to stable coexistence of all strategies (figure 5*c*).

This pattern emerges even more conspicuously in the numerical bifurcation analysis presented in figure 6. If the temptation to cheat is low, the system settles at a uniform communication scheme, even if cheaters are present (figure 6*a*). Which scheme this is depends on the initial conditions, a phenomenon called 'symmetry breaking' in physics (Di Paolo 2000); in figure 6 initial conditions always favour signal *A*. If the temptation to cheat increases, the frequency of honest use and hence signal credibility decreases (figure 6*b*). A rare signal (*B*) has a higher credibility (figure 6*b*), which means that at some point it will invade and coexist with *A*. The resulting dynamics are cyclic until the temptation to cheat is so large that honest users of neither signal can persist. However, they are maintained at a low frequency of honest signallers (figure 5*c*), which indicates that our results will be sensitive to the level of mutation. Indeed, the range where fluctuations occur depends on the mutation rate, as is indicated by the bifurcation analysis presented in figure 7. None the less the result is qualitatively robust:

whatever the mutation rate (and for other ecological parameters, results not shown) increasing the temptation to cheat produces a Tower of Babel: the number of signals increases, negatively affecting overall mutual intelligibility.

6. DISCUSSION

(a) *Honest communication and cheating*

Depending on the selfish advantage (the temptation to cheat) different outcomes are possible when individuals may convey either honest or false information. If the temptation to cheat is low, the advantages of having a common communication system prevail, thus corroborating Krakauer, Nowak and co-workers' results (Nowak & Krakauer 1999; Nowak *et al.* 1999*a*,*b*, 2000). Indeed, some apparently honest signalling systems are widely used in animal communities, even among different species. For example, even though the subtler points of bird song probably escape us, we all immediately recognize a bird's alarm call. Birds also often respond to other species' hawk alarm calls (Skelton 1993).

When the temptation is minimal, cheaters cannot maintain themselves at all in the population because the gains received by exploiting their gullible neighbours are not sufficient to make up for the loss they suffer by ignoring useful information. After all, cheaters are suspicious and ignore other signals (Krakauer & Pagel 1995)! When the temptation increases to moderate values, however, cheaters can invade but this does not necessarily mean that they will replace the honest signal users. Thus, the mere fact that cheaters invade does not imply that honest signalling cannot persist. Our analysis corroborates the finding in earlier studies that in structured populations honest signallers can withstand a certain level of cheating (Ackley & Littman 1994; Krakauer & Pagel 1995; Oliphant 1996; Grim *et al.* 1999; Grassly *et al.* 2000; Wagner 2000; Reggia *et al.* 2001), just as altruists that maintain themselves in a selfish population as long as they can cluster (Van Baalen & Rand 1998).

When the temptation to cheat increases even further the cheaters effectively take over. However, then honest signallers that employ a different signal, a 'private code' not understood by the rest of the population, will have an advantage and can thus increase in numbers (this can happen after a single mutation: when a mutant arises it has a certain probability to form a cluster through demographic stochasticity). The mutant's increase continues until the code is 'cracked' by cheaters, after which the credibility of the signal will drop again and another honest signal can invade. If this is the case, the meaning of symbols used in the population is not fixed but variable and will continually change over time. In a spatial setting this will correspond to a shifting mosaic of signals. The characteristics of this mosaic will depend on many details: background demography of the population, dispersal strategies, the possible number of distinct signals, the rates of mutation of the different signals, the geometry of the environment, and so on.

Eventually, when the temptation to cheat is too high, no honest signaller can maintain itself and a babel of mutually unintelligible signals results. A question of obvious interest is how many different signals can be maintained in the population. For the frozen pattern that results for low temptations to deceive, in principal large numbers of signals can coexist, provided there is sufficient space for them not to overlap. For an increased temptation we found the two signals could coexist in the same region, forming a shifting mosaic. For simplicity, the number of possible signals was limited to two, but nothing prevents in principle the arrival (through mutation) of other signals. Our reasoning is likely to hold for these new signals as well. If temptation is increased further the fraction of dishonest signallers decreases and at some point it will be possible for a rare third signal to invade. Once abundant, this third code will get cracked and in this way a cyclic pattern of three signals can ensue. This pattern leads us to suggest that the number of signals will increase with the temptation to deceive.

(b) *Green Beards*

Our results may give new life to what is called the 'Green Beard' hypothesis for the evolution of altruism. Hamilton (1964) suggested that kin recognition would favour the evolution of altruism, but Dawkins (1976) argued that this would not work on the basis of a genetically determined trait ('green beard') as it would be vulnerable to cheating. First, the fact that honest communicators can maintain themselves in a spatial world shows that communication can be used to function to preferentially help relatives and thus function as a green beard. Moreover, our results suggest an outcome that seems not to have been considered by Dawkins, but if the benefit of kin recognition is sufficient, the process could result in fluctuating hair colour polymorphisms ('beard chromodynamics'). In reality hair colour might not mutate fast enough for this mechanism to operate (although it is noteworthy that in humans, facial hair *is* among the most variable traits), but this constraint does not apply to the use of language or words as a recognition system where individuals are free to adopt new communication systems.

(c) *Human language*

For human languages many patterns of word use are very constant over time (Pinker 1994; Crystal 1997), which seems to corroborate the idea that language is used for efficient communication (Livingstone 2001). However, it is well known that word usage can change over time. Our analysis predicts that under divergent interests language patterns may be unstable and thus it provides an explanatory mechanism for certain types of linguistic change. Of course, there are other mechanisms that explain linguistic change (Kirby 1998; Livingstone 2001) or even Towers of Babel (Lachmann & Bergstrom 1998), but our model could give insight into the conditions that favour the use of codes that are associated with various subcultures (e.g. 'cockney rhyming slang' in English or '*verlan*' in French): such communication systems either increase the efficiency of communication within a group or permit the exploitation of gullibles in the population at large. Either case implies that to better understand the evolution of human language we may need to appreciate the fact that human communication mediates not only cooperation but also conflict.

Willem Iemhoff, Jean-François Le Gaillard, Sara Magalhaes, Kees Nagelkerke and Robert Payne are thanked for discussion and comments on previous versions of the manuscript. The comments of two anonymous referees were of great help in improving this paper. Part of the research reported in this paper was carried out when M.v.B. was a research fellow of the

Royal Dutch Academy of Arts and Sciences (KNAW) whose support is gratefully acknowledged. M.v.B. also gratefully acknowledges support by the French Centre National de la Recherche Scientifique (CNRS).

REFERENCES

Ackley, D. H. & Littman, M. L. 1994 Altruism in the evolution of communication. In *Artificial Life IV: Proc. 4th Int. Workshop on the Synthesis and Simulation of Living Systems* (ed. R. A. Brooks & P. Maes), pp. 40–49. Cambridge, MA: MIT Press.

Axelrod, R. R. 1984 *The evolution of cooperation*. New York: Basic Books.

Axelrod, R. R. & Hamilton, W. D. 1981 The evolution of cooperation. *Science* **211**, 1390–1396.

Crystal, D. 1997 *The Cambridge encyclopedia of language*. Cambridge University Press.

Dawkins, R. 1976 *The selfish gene*. Oxford University Press.

Di Paolo, E. A. 2000 Ecological symmetry breaking can favour the evolution of altruism in an action–response game. *J. Theor. Biol.* **203**, 135–152.

Durrett, A. 1988 *Lecture notes on particle systems and percolation*. Pacific Grove, CA: Wadsworth & Brooks.

Godfray, H.-C. J. 1991 Signalling of need by offspring to their parents. *Nature* **352**, 328–330.

Grafen, A. 1990 Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546.

Grafen, A. & Johnstone, R. A. 1993 Why we need ESS signalling theory. *Phil. Trans. R. Soc. Lond.* B **340**, 245–251.

Grassly, N. C., von Haeseler, A. & Krakauer, D. C. 2000 Error, population structure and the origin of diverse sign systems. *J. Theor. Biol.* **206**, 369–378.

Grim, P., Kokalis, T., Tafti, A. & Kilb, N. 1999 Evolution of communication with a spatialized genetical algorithm. *Evol. Commun.* **3**, 105–134.

Hamilton, W. D. 1964 The genetical evolution of social behaviour. I and II. *J. Theor. Biol.* **7**, 1–16 and 17–52.

Harada, Y., Ezoe, H., Iwasa, Y., Matsuda, H. & Sato, K. 1995 Population persistence and spatially limited social interaction. *Theor. Popul. Biol.* **48**, 65–91.

Jansen, V. A. A. & de Roos, A. M. 2000 Predator–prey dynamics across spatial scales: the role of space in reducing population cycles. In *The geometry of ecological interactions: simplifying spatial complexity* (ed. U. Dieckmann, R. Law & J. A. J. Metz), pp. 183–201. Cambridge University Press.

Johnstone, R. A. & Grafen, A. 1992 The continuous Sir Philip Sidney game: a simple model of biological signaling. *J. Theor. Biol.* **156**, 215–234.

Kirby, S. 1998 Fitness and the selective adaptation of language. In *Approaches to the evolution of language: social and cognitive basis* (ed. J. Hurford, C. Knight & M. Studdert-Kennedy), pp. 359–383. Cambridge University Press.

Krakauer, D. C. & Pagel, M. 1995 Spatial structure and the evolution of honest cost-free signalling. *Proc. R. Soc. Lond.* B **260**, 365–372.

Krebs, J. R. & Dawkins, R. 1984 Animal signals: mind-reading and manipulation. In *Behavioural ecology. An evolutionary approach*, 2nd edn (ed. J. R. Krebs & N. B. Davies), pp. 380–402. Oxford: Blackwell.

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

Lachmann, M., Sella, G. & Jablonka, E. 2000 On the advantages of information sharing. *Proc. R. Soc. Lond.* B **267**, 1287–1293. (DOI 10.1098/rspb.2000.1140.)

Lachmann, M., Szamado, S. & Bergstrom, C. T. 2001 Cost and conflict in animal signals and human language. *Proc. Natl Acad. Sci. USA* **98**, 13 189–13 194.

Livingstone, D. 2001 The evolution of dialect diversity. In *Simulating the evolution of language* (ed. A. Cangelosi & D. Parisi), pp. 99–118. London: Springer.

Matsuda, H., Ogita, N., Sasaki, A. & Sato, K. 1992 Statistical mechanics of population: the lattice Lotka–Volterra model. *Prog. Theor. Phys.* **88**, 1035–1049.

Maynard Smith, J. 1994 Must reliable signals always be costly? *Anim. Behav.* **47**, 1115–1120.

Nakamaru, M., Matsuda, H. & Iwasa, Y. 1997 The evolution of cooperation in a lattice-structured population. *J. Theor. Biol.* **184**, 65–81.

Noble, J. 1999 Evolved signals: expensive hype vs. conspirational whispers. In *Artificial Life VI: Proc. 6th Int. Workshop on the Synthesis and Simulation of Living Systems* (ed. C. Adami, R. Belew, H. Kitano & C. Taylor), pp. 358–367. Berlin: Springer.

Noble, J. 2000 Cooperation, conflict and the evolution of communication. In *The emergence of language* (ed. C. Knight, M. Studdert-Kennedy & J. Huford), pp. 40–61. Cambridge University Press.

Noble, J., Di Paolo, E. A. & Bullock, S. 2001 Adaptive factors in the evolution of signaling systems. In *Simulating the evolution of language* (ed. A. Cangelosi & D. Parisi), pp. 53–78. London: Springer.

Nowak, M. A. & Krakauer, D. C. 1999 The evolution of language. *Proc. Natl Acad. Sci. USA* **96**, 8028–8033.

Nowak, M. A. & Sigmund, K. 1994 The alternating Prisoner's Dilemma. *J. Theor. Biol.* **168**, 219–226.

Nowak, M. A., Krakauer, D. C. & Dress, A. 1999*a* An error limit for the evolution of language. *Proc. R. Soc. Lond.* B **266**, 2131–2136. (DOI 10.1098/rspb.1999.0898.)

Nowak, M. A., Plotkin, J. B. & Krakauer, D. C. 1999*b* The evolutionary language game. *J. Theor. Biol.* **200**, 147–162.

Nowak, M. A., Plotkin, J. B. & Jansen, V. A. A. 2000 The evolution of syntactic communication. *Nature* **498**, 495–498.

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

Pinker, S. 1994 *The language instinct: the new science of language and mind*. Harmondsworth, UK: Penguin.

Rand, D. A. 1999 Correlation equations and pair approximations for spatial ecologies. In *Advanced ecological theory* (ed. J. McGlade), pp. 100–142. Oxford: Blackwell.

Reggia, J., Schulz, R., Wilkinson, G. & Uriagereka, J. 2001 Conditions enabling the emergence of inter-agent signalling in an artificial world. *Artificial Life* **7**, 3–32.

Skelton, P. 1993 *Evolution: a biological and palaeontological approach*. Harlow, UK: Prentice-Hall.

Van Baalen, M. 2000 Pair approximations for different geometries. In *The geometry of ecological interactions: simplifying spatial complexity* (ed. U. Dieckmann, R. Law & J. A. J. Metz), pp. 359–387. Cambridge University Press.

Van Baalen, M. & Jansen, V. A. A. 2001 Dangerous liaisons: the ecology of private interest and common good. *Oikos* **95**, 211–224.

Van Baalen, M. & Rand, D. A. 1998 The unit of selection in viscous populations and the evolution of altruism. *J. Theor. Biol.* **191**, 631–648.

Wagner, K. 2000 Cooperative strategies and the evolution of communication. In *Artificial Life VI. Proc. 6th Int. Conference on Artificial Life* (ed. C. Adami, R. K. Belew, H. Kitano & C. E. Taylor), pp. 149–179. Cambridge, MA: MIT Press.

Zahavi, A. 1977 The cost of honesty (further remarks on the handicap principle). *J. Theor. Biol.* **67**, 603–605.

Zahavi, A. 1993 The fallacy of conventional signalling. *Phil. Trans. R. Soc. Lond.* B **340**, 227–230.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.