Synthese (2010) 172:177–191
DOI 10.1007/s11229-009-9477-0

Evolutionary dynamics of Lewis signaling games: signaling systems vs. partial pooling

Simon M. Huttegger · Brian Skyrms · Rory Smead · Kevin J. S. Zollman

Received: 1 October 2007 / Accepted: 7 May 2008 / Published online: 26 February 2009

Abstract Transfer of information between senders and receivers, of one kind or another, is essential to all life. David Lewis introduced a game theoretic model of the simplest case, where one sender and one receiver have pure common interest. How hard or easy is it for evolution to achieve information transfer in Lewis signaling?. The answers involve surprising subtleties. We discuss some if these in terms of evolutionary dynamics in both finite and infinite populations, with and without mutation.

Keywords Signaling · Evolution · Dynamics · Replicator · Replicator-mutator · Moran

1 Introduction

The exchange of information by sending and receiving signals is one of the most fundamental processes in living organisms. It is a largely neglected area of epistemology. (For exceptions see Dretske, Millikan, and Harms). It is a key to teamwork and social structure. This is true not only in human affairs but also at all other levels of biological organization. A web of signals ties social groups together, enabling sharing of information and coordinated activity.

S. M. Huttegger · B. Skyrms (✉) · R. Smead · K. J. S. Zollman
Department of Logic & Philosophy of Science, School of Social Sciences, University of California, 3151 Social Science Plaza A, Irvine, CA, 92697-5100, USA
e-mail: bskyrms@uci.edu

Present Address:
K. J. S. Zollman
Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Successful signaling also has a long history as a philosophical conundrum, although it is often discussed in terms of the origin of language with all the baggage that the term "language" carries. Many of the fundamental questions that have been raised regarding the origin of language from Cratylus to Quine apply to signaling as well. In particular, we are led to ask how *content* or *information* can come to be attached to conventional signals without there being a pre-existing signaling system to set up the convention. What can we say about the possible origin of successful signaling?

To investigate this very philosophical question at a high level of abstraction we need two things: (1) an interactive model—or family of models—of potential signaling situations and (2) an abstract dynamics of evolution or learning that can operate on such a model. Exploration of this territory is a large enterprise. We do not intend to survey the field here. Rather we report analyses of some basic models. We focus on the very simplest signaling games and on dynamics driven by replication or imitation. There are some surprising results. The tools and techniques used here are applicable to richer and more complicated signaling situations.

In his book *Convention,* Lewis (1969) provided the most basic model of sender-receiver games. In Lewis signaling games, nature picks one of N possible states of the world at random and a player, the sender, observes the state and selects one of N signals to send to a receiver. The receiver observes the signal and selects one of N possible acts. There is exactly one act that is "right" for each state, in that both sender and receiver both get a payoff equal to one if the right act is done for the state and both get a payoff of zero otherwise. [Here we take the states to be equiprobable. The more general case is obviously of interest, but the analysis is rather different and it is treated elsewhere; see Hofbauer and Huttegger (2007) and Jäger (2007)]

A sender's strategy is a function from states to signals; a receiver's from signals to acts. The two strategies form a *signaling system equilibrium* if they guarantee that the correct act is always taken. From any signaling system equilibrium, a permutation of signals (the same in both sender and receiver strategies) leads to another signaling system equilibrium with exactly the same payoff. This is why Lewis introduced these games as a model in which meaning of signals is *purely conventional*.

There are other equilibria in these games. There are always *completely pooling* equilibria in which the sender ignores the state and the receiver ignores the signal. For instance, the sender might always send signal 1 and the receiver might have the strategy of always doing act 2. The signals then carry no information. If $N > 2$, there are also *partial pooling* equilibria in which some, but not all, of the information about the state is transmitted. Consider a Lewis signaling game with $N = 3$, where the sender always sends signal 1 in both states 1 and 2, and who in state 3 sometimes sends signal 2 and sometimes sends signal 3. Pair this sender with a receiver, who does act 3 in response to both signals 2 and 3, and who upon receiving signal 1 sometimes does act 1 and sometimes act 2, as shown in Fig. 1. In this equilibrium, information about state 3 is transmitted perfectly, but states 1 and 2 are "pooled".

In Lewis signaling games, signaling system equilibria are distinguished by being *strict*. A player who unilaterally deviates from such an equilibrium is strictly worse off—a fact that plays an important part in Lewis' theory of convention. At a completely pooling equilibrium, a player who unilaterally deviates is equally well off, no matter *what* the deviation. In more general sender-receiver games in which players do not

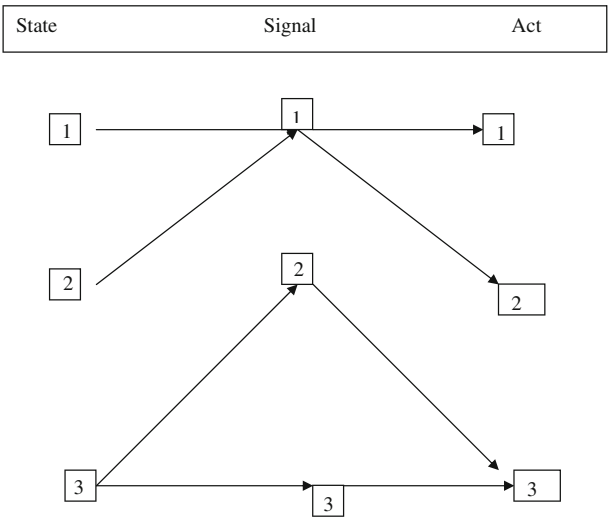

Fig. 1 Partial pooling equilibrium in Lewis signaling game

have common interest, partial pooling equilibria can be strict but in Lewis signaling games this is not so. Consider Fig. 1. Given the sender's strategy, the receiver might as well always choose act 1 on receipt of signal 1, or always choose act 2 or anything in between. Given the receiver's strategy the sender could just as well always send signal 2 in state 3, or always send signal 3, or anything in between.

There is an evolutionary twist to the strictness of signaling system equilibria in Lewis signaling games. In these *games signaling system equilibria are the unique evolutionarily stable strategies*. Other pure strategy equilibria cannot be evolutionarily stable or even neutrally stable. [See Wärneryd (1993)] One might be tempted to conclude that this is the whole story—that signaling systems will always evolve—but this conclusion would be premature. To see this we will need to look at the evolutionary dynamics.

Here, we focus on evolution of strategies in a Lewis signaling game in a two-population context: a population of senders and a population of receivers. We start with the standard large population model of differential reproduction, the replicator dynamics. [See Hofbauer and Sigmund (1998) for a comprehensive treatment.] Our motivating question is: "Will signaling evolve?"

Section 2 presents a positive analysis for the case of $N = 2$. For the replicator dynamics, there is global convergence to an equilibrium. Signaling systems are the only attractors. And all other equilibria are dynamically unstable. However the binary case is special in this regard. In Sect. 3, we see how the picture changes with $N > 2$. The dynamical picture is more complicated than one might expect. Sometimes partial pooling equilibria now spontaneously evolve; sometimes signaling system equilibria. Most partial pooling equilibria are neutrally stable mixed states, in the sense of Maynard Smith. From the point of view of the replicator dynamics, they are Lyapunov-stable but not attractors. But some of the partial pooling equilibria are unstable. The set of partial pooling equilibria is not an attractor, but nevertheless it has a basin of attraction of positive measure.

Section 4, investigates how the dynamical picture for N = 2 and 3 changes when we move from pure replicator dynamics to replicator-mutator (aka selection-mutation) dynamics. Dramatic changes are possible because the previous models are not *structurally stable.* [See Guckenheimer and Holmes (1983) for more information on structural stability and related global concepts for dynamical systems like qualitative equivalence]. Indeed, dramatic changes are what we see. The addition of mutation causes connected components of complete and partial and pooling equilibria to collapse to single points. Analyses of stability properties of these points suggest that positive results for spontaneous evolution of signaling re-emerge.

Section 5 raises the same questions in the context of evolution in finite populations with fixed population size (via the Moran process with and without mutation). The qualitative features of the foregoing analysis using the replicator dynamics are seen to continue to hold except in very small populations. Without mutation, the process can lead to fixation of any profile of pure strategies, but in reasonably sized populations what we see in simulations is fixation of signaling systems and partial pooling equilibria. With mutation, the process is ergodic, and spends most of its time near perturbed signaling system equilibria. Section 6 concludes.

2 The simplest Lewis signaling game

Consider the Lewis game with only 2 states, 2 signals, and 2 acts. The sender has four possible strategies:

Sender 1: State 1 => Signal 1, State 2 => Signal 2
Sender 2: State 1 => Signal 2, State 2 => Signal 1
Sender 3: State 1 => Signal 1, State 2 => Signal 1
Sender 4: State 1 => Signal 2, State 2 => Signal 2

Strategies 3 and 4 are "pooling" since the sender ignores the state and always sends the same signal; the states are "pooled." Strategies 1 and 2 are "separating" since each state elicits a different signal.

Likewise the receiver has four possible strategies:

Receiver 1: Signal 1 => Act 1, Signal 2 => Act 2
Receiver 2: Signal 1 => Act 2, Signal 2 => Act 1
Receiver 3: Signal 1 => Act 1, Signal 2 => Act 1
Receiver 4: Signal 1 => Act 2, Signal 2 => Act 2

Receiver's strategies 3 and 4 act as if they are deaf to the signal, while strategies 1 and 2 act as if the signal contains information about the state, but disagree about what that information is.

Considering combinations of sender and receiver strategies, ⟨S1, R1⟩ and ⟨S2, R2⟩ are signaling system equilibria. They always get things right, for a payoff of 1.

Mismatched separating strategies ⟨S1, R2⟩ and ⟨S2, R1⟩ always get things wrong, for an average payoff of 0. If we assume that the states are equiprobable, and that the population and ensemble of situations faced is large and independent enough, we can fill in the average payoff matrix for senders and receivers:

	R1	R2	R3	R4
S1	1,1	0,0	.5,.5	.5,.5
S2	0,0	1,1	.5,.5	.5,.5
S3	.5,.5	.5,.5	.5,.5	.5,.5
S4	.5,.5	.5,.5	.5,.5	.5,.5

The first entry is the payoff of strategy S_i played against strategy R_j, $W(S_i|R_j)$, and the second entry is the payoff of strategy R_j against S_i, $W(R_j|S_i)$. Note that in every interaction, these are the same. This strong common interest makes this a *partnership game*.

Let x_i be the population proportion of those who use strategy S_i in the population of senders and y_i be the population of those who use strategy R_i in the population of receivers. We assume random matching of senders and receivers, so that:

$$W(S_i) = \sum_j y_j W(S_i|R_j) \text{ and } W(R_j) = \sum_i x_i W(R_j|S_i)$$

The average fitnesses of the sender and receiver population respectively are:

$$W(S) = \sum_i W(S_i)\, x_i \text{ and } W(R) = \sum_j W(R_j)\, y_j$$

We consider the evolution of this two population system using bipartite replicator dynamics [Taylor and Jonker (1978); Hofbauer and Sigmund (1998)]:

$$dx_i/dt = x_i\,[W(S_i) - W(S)]$$
$$dy_j/dt = y_j\,[W(R_j) - W(R)]$$

Because this is a partnership game, average payoff is a Lyapunov function for the system [in fact, it is even a potential function; see Hofbauer and Sigmund (1998)]. Consequently we have global convergence; all trajectories must lead to dynamic equilibria. Analysis reduces to examining the stability properties of these equilibria.

The equilibria can be found algebraically [Mathematica] to be one of the following non-exclusive list of possibilities:

a. $x1 = 0$ & $x2 = 0$
b. $x1 = x2$ & $y1 = y2$
c. $y1 = 0$ & $y2 = 0$
d. $x2 = 1$ & $y2 = 1$
e. $x2 = 1$ & $y1 = 1$
f. $x1 = 1$ & $y2 = 1$
g. $x1 = 1$ & $y1 = 1$

a, b and c define a connected component of rest points. Equilibria d and g are the signaling systems, while e and f are anti-signaling systems.

Linear stability analysis is gotten by calculating the eigenvalues of the Jacobian for the system. These are given in the following table:

a. x1 = 0 & x2 = 0	⟨0,0,0,0, .5(y1-y2), .5(y2-y1)⟩
b. x1 = x2 & y1 = y2	⟨0,0,0,0, –SQRT(x2)SQRT(y2), SQRT(x2)SQRT(y2)⟩
c. y1 = 0 & y2 = 0	⟨0,0,0,0, .5(x1-x2), .5(x2-x1)⟩
d. x2 = 1 & y2 = 1	⟨−1, −1, −.5, −.5, −.5, −.5⟩
(Signaling system)	
e. x2 = 1 & y1 = 1	⟨ .5, .5, .5, .5, 1, 1⟩
f. x1 = 1 & y2 = 1	⟨ 1/2, 1/2, .5, .5, 1, 1⟩
g. x1 = 1 & y1 = 1	⟨−1, −1, −.5, −.5, −.5, −.5⟩
(Signaling system)	

The signaling systems, d and g, have all negative eigenvalues; they are asymptotically stable (sinks). In replicator dynamics all pure strategy combinations are dynamic equilibria (since all alternatives are extinct) and the combinations e and f that always get things wrong qualify. But they have all positive eigenvalues and are repelling (sources). In situation a, senders are pooling. They send the same signal no matter what the state. We have a linearly unstable equilibrium with one negative eigenvalue in all cases where y1 is unequal to y2, indicating that a separating sender could do better against the natives than they do against each other. Where y1 = y2, all eigenvalues are zero, indicating that further analysis is required. Case c is similar, except that here receivers are doing the same thing no matter what message they see. Case b, where separating strategies for both sender and receiver are in equipoise is linearly unstable, except when all separating strategies are extinct [x1 = x2 = y1 = y2 = 0]. In that case we again have all zero eigenvalues.

The equilibria with all zero eigenvalues—special cases of a, b, c—although they are not linearly unstable, are nevertheless unstable. In each of these equilibria, the average population fitness is 1/2. Consider a perturbation that adds an epsilon of a signaling system to the populations, e.g. of S1 to the sender population and R1 to the receiver population. Both S1 and R1 will have average fitness of 1/2 against the natives and of 1 against each other. Consequently their population proportions will grow, leading away from the equilibrium.

Signaling systems are therefore the only stable equilibria in the 2 state, 2 signal, 2 act Lewis signaling game. Because average payoff is a Lyapunov function, almost all points converge to a signaling system. (This is no longer true if we change the game so that states are no longer equiprobable. See Huttegger 2007a, and Hofbauer and Huttegger 2007).

3 Lewis with N = 3

Computer simulations of the N = 2 case discussed in the last section, starting at randomly chosen population proportions, always converge to a signaling system equilibrium. This is no longer the case for N = 3 and greater. Although most simulations converge to a signaling system, a significant number appear to converge to a partial pooling equilibrium of the sort shown in Fig. 1. Using the discrete time version of

the replicator dynamics, approximately 4.7% of the initial starting points converge to an equilibrium with partial pooling. The apparent rest point is different in each case, but each is an example of a partial pooling equilibrium similar to the one pictured in Fig. 1. Are these genuine limiting points of the dynamics, or just points near which motion along the trajectories is extremely slow?

Consider the situation indicated in Fig. 1. Denote the probabilities that the sender sends signals 2 and 3 in state 3 as x, (1−x) and those with which the receiver does acts 1 and 2 upon receiving the ambiguous signal 1 as y, (1−y) respectively. Figure 1 represents a square of partial pooling equilibria. (There are 2 other such squares where sender pools either states 2 and 3 or states 1 and 3, instead of 1 and 2.) At each point, both senders and receivers have an average payoff of 2/3.

In each corner of the square, both sender and receiver are deterministic. The sender only uses 2 signals; the receiver only does 2 acts. The unused signal could be utilized to construct a signaling system. A mutant sender who used the signal to discriminate between the pooled acts paired with a mutant receiver who used that information to do the right act would signal perfectly between each other and do as well against the native as the natives do against each other. For this reason, the relevant set-valued version of evolutionary stability does not apply to the set of partial pooling equilibria. (It is not a strict equilibrium set in the sense of Barrett 2007). The corners of the square are each dynamically unstable in the replicator dynamics. (For an analogous situation in ultimatum bargaining see Gale et al. 1995 and Binmore and Samuelson 1999). Thus the set of partial pooling equilibria is not a dynamical attractor.

Here are the strategies that participate in the partial pooling square together with the strategies used in the four signaling systems that destabilize its corners. (Sender's strategies are shown as maps from states to signals; Receiver' as maps from signals to acts.)

S1: 1=>1, 2=>1, 3=>2 R1: 1=>2, 2=>3, 3=>3 (PPool)
S2: 1=>1, 2=>1, 3=>3 R2: 1=>1, 2=>3, 3=>3 (PPool)
S3: 1=>1, 2=>2, 3=>3 R3: 1=>1, 2=>1, 3=>3 (Sig I)
S4: 1=>2, 2=>1, 3=>3 R4: 1=>2, 2=>1, 3=>3 (Sig II)
S5: 1=>3, 2=>1, 3=>2 R5: 1=>2, 2=>3, 3=>1 (Sig III)
S6: 1=>1, 2=>3, 3=>2 R6: 1=>1, 3=>2, 2=>3 (Sig IV)

Payoffs of one strategy against another are shown in the following table. (There is only one entry because payoff for sender and receiver are the same.)

	S1	S2	S3	S4	S5	S6
R1	2/3	2/3	1/3	2/3	2/3	1/3
R2	2/3	2/3	2/3	1/3	1/3	2/3
R3	1/3	2/3	1	1/3	0	1/3
R4	1/3	2/3	1/3	1	1/3	0
R5	2/3	1/3	0	1/3	1	1/3
R6	2/3	1/3	1/3	0	1/3	1

Consider the corner of the partial pooling square ⟨S2, R2⟩. If a few S3 and R3 types were to enter the population, they would get a payoff of 2/3 against the natives but a payoff of 1 against each other. In like manner, S4 and R4 destabilize ⟨S2, R1⟩, S5 and R5 destabilize ⟨S1, R1⟩ and S6 and R6 destabilize ⟨S1, R2⟩.

But what about the partial pooling equilibria in the interior of the square? A strategy that is part of a signaling system that destabilizes one corner of the square will do worse than the natives, where the native population is in the interior of the pooling square. In populations consisting of these 6 sender strategies and 6 receiver strategies, just off the interior of the partial pooling square, there will be a component of the velocity towards the square. (This would remain true if we included all the possible strategies in the signaling game, since others do worse against this pooling plane than those we are considering.)

If non-poolers are rare, movement with replicator dynamics toward this (partial) pooling plane will be slow. There are two possibilities: (1) as orbits approach the plane they slowly curve around, eventually are attracted to the corners, and then move out toward a signaling system or (2) orbits near the interior of the plane go into the plane.

To see which is the case, we calculate the eigenvalues of the Jacobian at points on the pooling plane with Mathematica. Two zeros are expected, since there is no motion in the pooling plane itself. In the interior of the pooling plane, all the other eigenvalues are negative; at the center they all equal –1/6. At the corners more zeros appear, consistent with the (higher-order) instability caused by signaling systems. We conclude that the possibility of convergence to pooling equilibria is not just an artifact of simulation, but is dynamically genuine asymptotic behavior.

That falls short of a proof for the full game, since we restricted the strategies. But there are two proofs for the full game, using different techniques. Huttegger (2007a,b) uses center-manifold theory. Pawlowitsch (2008) uses the fact that interior points of the partial pooling plane are Lyapunov stable in the replicator dynamics. Their proofs generalize to all Lewis signaling games with states equiprobable and $N > 2$. Some details on this can be found in the Appendix.

4 Mutation N = 2

Sender-receiver games create connected sets of pooling equilibria in the replicator dynamics. The resulting dynamical systems, however, are *structurally unstable* [see Guckenheimer and Holmes (1983)]. This means that a small perturbation in the vector field can yield a completely different dynamical picture, though not every perturbation will do so. Different perturbations may even cause diametrically opposite changes in the qualitative dynamics. If we think of plausible perturbations of the dynamics of replication, the first thing to try is to add a little uniform mutation. (But note that addition of conformist bias, as in Skyrms (2005), would give quite different results than the mutation explored here).

In discrete time, each generation reproduces according to replicator dynamics but (1−e) of the progeny of each type breed true and e of the progeny mutate to all types with equal probability. (Self-mutation is allowed.) Taking the continuous time limit

leads to the selection-mutation equation [Hadeler (1981); Hofbauer (1985)], which we apply to both sender and receiver populations:

$$dx_i/dt = x_i\,[(1-e)\,W(S_i) - W(S)] + (e/n)\,W(S)$$
$$dy_j/dt = y_j\left[(1-e)\,W\left(R_j\right) - W(R)\right] + (e/n)\,W(R)$$

Hofbauer (1985) finds a Lyapunov function for the one population version and, noting that average fitnesses of both populations must be the same, this generalizes to our case as:

$$(1-e)\log W(S) + (e/n)\left[\sum\nolimits_i \log x_i + \sum\nolimits_j \log y_j\right]$$

Therefore, just as in the unmodified replicator dynamics, all orbits must converge to an equilibrium.

The set of equilibria, however, has changed. Let us start by examining the effect on the N = 2 signaling game. The signaling system equilibria are pushed a little bit into the interior by the noise. With a mutation rate of 1%, the ⟨S1, R1⟩ equilibrium moves to a point where pr(S1) = .98743, pr(R1) = .98743. Likewise for the ⟨S2, R2⟩ equilibrium. The plane of pooling equilibria however, dissolves to a single point, and this moves to the point where all strategies are equiprobable. (This makes intuitive sense, for there is no selection pressure on the plane of pooling equilibria, nor when the signaling systems are equiprobable. Only mutation pressure operates, and mutation is uniform.) If we solve for dynamic equilibria (Mathematica), we find that these 3 points are the only remaining equilibria

Here are the eigenvalues of the Jacobian with a mutation rate, e = .01:

Equilibrium	Eigenvalues
Signaling I	⟨−.994949, −.989873, −.497475, −.497475, −.497475,−.492398⟩
Signaling II	⟨−.994949, −.989873, −.497475, −.497475, −.497475, −.492398⟩
Babbling —All equiprobable	⟨−.505, .485, −.01, −.01, −.01, −.01⟩

With e = .01, the eigenvalues of the Jacobian at the perturbed signaling systems are all negative. They are still sinks. The (babbling) point with all strategies equiprobable, however, has changed. It is now linearly unstable. This picture must change at some mutation rate high enough to overwhelm selection and stabilize the babbling equilibrium. This *bifurcation* does not occur until e = 1/3. For $0 < e < 1/3$ the picture remains qualitatively the same, with the perturbed pooling equilibrium unstable and the perturbed signaling systems attracting almost all possible initial points. Hofbauer and Huttegger (2007) provide a deeper mathematical analysis of selection-mutation dynamics for the N = 2 case.

5 Mutation N = 3

Mutation does not change the bottom line for N = 2. Signaling systems will (almost) always evolve. But what will a little mutation do to the partial pooling planes for N > 2? They too must collapse because there is no selection pressure on the partial pooling plane. Consider N = 3 with e = .01. Since there is no selection pressure on the partial pooling plane, mutation tends to push the populations to the center of the plane, but it also pushes the populations off the plane, into the interior of their simplices. Since in this case—unlike the complete pooling plane—there is selection pressure pushing back in, the partial pooling point is located where these pressures come into balance. In the case of the pooling plane discussed in Sect. 3, with mutation e = .01, this happens just a little off the center of the plane. The point was found numerically to high precision using Newton's method. It is at about pr(S1) = pr(S2) = pr(R1) = pr(R2) = .4867146.

At this perturbed partial pooling equilibrium the Jacobian has 2 positive eigenvalues of about .003, with the rest negative: It is an unstable saddle. Mutation has destabilized the whole pooling plane. The perturbed signaling systems remain near the original signaling systems and are sinks.

The foregoing is only an analysis of the effects of mutation on a subsystem of the N = 3 signaling game. The subsystem initially contains a plane of partial pooling equilibria and the components of the four signaling systems with a chance of destabilizing it. This subsystem already strains the resources of Mathematica. We can no longer solve for all equilibria, and the Jacobian fills several pages. Analysis of the full game along these lines does not seem feasible.

However, we have seen in this subsystem how mutation can collapse this partial pooling plane to a single unstable interior pooling point. The same thing will happen in the other subsystems gotten from this one by permutation of signals. The connected component of total pooling equilibria will collapse to a single point in the same way. This suggests the *conjecture* that with small mutation we have a finite number of interior equilibria, all of which are unstable except for the perturbed signaling systems. This conjecture is consistent with the results of computer simulations. Simulations using discrete time replicator-mutator dynamics with both 1% and 0.1% mutation rates found that the system *always* converged to a perturbed signaling system equilibrium.

6 Evolution in finite populations

The replicator dynamics is an infinite population model of differential reproduction. In finite populations the process is stochastic rather than deterministic.

The population may be either (1) varying size or (2) at constant size equal to the "carrying capacity" of the environment. There are simple urn models of each process due, respectively, to Schreiber (2001) and Moran (1962). In Schreiber's model of a variable size finite population, if no strategy goes extinct and the population grows, the process becomes arbitrarily close (with arbitrary high probability) to the replicator dynamics. Thus, if no strategy goes extinct the analysis of long term behavior reduces to that already given.

The Moran model is a finite state Markov chain and all states where the whole population plays the same strategy are absorbing states. One might look here for results most at variance with the replicator dynamics. However, simulations show that, to a large extent, the replicator dynamic analysis carries over.

For investigating evolution of signaling in the Moran Process, we have a fixed, finite population of Senders and another of Receivers, each with M individuals. Each individual is assigned an initial strategy so that the proportions of each type are randomly determined (roughly equivalent to selecting a random point in a simplex). As with the replicator dynamics, we assume random matching of Senders and Receivers and the fitness of each type S_i and R_i is the expected payoff of these interactions:

$$W(S_i) = \sum_j y_j W(S_i|R_j) \text{ and } W(R_j) = \sum_i x_i W(R_j|S_i)$$

Note that x_i and y_j are now positive integers where $\sum_i x_i = M$ and $\sum_i y_i = M$; these values represent the number of individuals with strategy S_i and R_j respectively. Each type in the population is then assigned a probability of reproduction. This probability is a function of the type's fitness and the number of individuals of that type:

$$\text{Rep}(S_i) = x_i W(S_i) / \sum_j x_j W(S_j) \text{ and}$$
$$\text{Rep}(R_i) = y_i W(R_i) / \sum_j y_j W(R_j)$$

Each generation, one sender strategy and one receiver strategy are chosen for reproduction based on these probabilities. Then, one individual in each population is selected at random and adopts the strategy type that is to be replicated.

This process continues until an absorbing state is reached. Any state where some $x_i = M$ and some $y_i = M$ is an absorbing state since $\text{Rep}(S_i) = \text{Rep}(R_i) = 1$. Also, strategies can become extinct since if $x_i = 0$ or $y_i = 0$, then $\text{Rep}(S_i) = \text{Rep}(R_i) = 0$.

For $N = 2$, where signaling always evolves in replicator dynamics, there is a strong tendency for signaling to evolve in finite populations. For instance, simulations for a population of size 1,000 were absorbed into signaling systems more than 98% of the time.

With reasonable size populations, $N = 3$, simulations produce both signaling systems and partial pooling equilibria. The proportion of partial pooling equilibria appears to have some sensitivity to population size. For sender and receiver populations of 10,000 each we get signaling systems about 93% of the time and partial pooling about 7%. With populations of 1,000 the proportion of partial pooling equilibria goes up to 14% and that of signaling systems down to 86%. Although other outcomes are definite theoretical possibilities, they were not observed in these simulations. However, when population size was shrunk to 100, signaling systems evolved in only about 41% of the trials and the other outcomes included not only partial pooling equilibria but other absorbing states as well.

The partial pooling equilibria observed in finite populations are somewhat different in character than the ones seen in the infinite populations. In the infinite population setting, the partial pooling equilibria involve sender and receiver populations that consist

of complementary mixes of strategies as shown in Fig. 1. However, the absorbing states of the Moran process cannot involve mixed populations. In the case of finite populations, a pooling equilibrium can occur when each population is uniform in strategy and *either* the sender population uses one signal for two states *or* the receiver population responds to two different signals with a single act; only one of the two populations needs to be absorbed into a partial pooling state.

Addition of mutation to the Moran process helps to avoid partial pooling and promote (approximate) signaling systems. A mutation parameter is included in the reproduction phase, where, with a small probability, an individual adopts a random strategy instead of replicated strategy. With mutation, because there are no absorbing states and the system is ergodic, no stable state will result. Thus, to gauge the effect of mutation we examine the state of the population in simulations after a large number of generations relative to the population size ($100 \times \mathrm{M}$ generations). Three general results in the populations were observed in simulations: perturbed signaling (average payoff >.85), partial pooling (average payoff between .67 and .60), and transition states. Observing a transition state becomes more likely with smaller populations and with higher-mutation rates but only frequently occurred with $\mathrm{M} = 100$ and a mutation of 5% (10% were in transition). In all other examined settings transition states were <2% of observed cases.

The following table gives the proportion of runs leading to signaling systems (or perturbed signaling systems) for no mutation, a mutation rate of 1%, and a mutation rate of 5%. These are an average of 1,000 trials of $100 \times M$ generations for all but $\mathrm{M} = 10{,}000$ with mutation which are 500 trials.

Population size	No mutation	1% Mutation	5% Mutation
10,000	.934	.942 (average payoff .980)	.980 (average payoff .905)
1,000	.856	.919 (average payoff .981)	.978 (average payoff .904)
100	.414	.629 (average payoff .979)	.795 (average payoff .907)

As in other settings, mutation helps the evolution of signaling systems, but prevents perfect communication. And, the higher the mutation rate is, the stronger these effects are.

Although replicator dynamics and Moran processes with and without mutations are quite different in their asymptotic properties, it appears that for reasonable finite simulations we see an approximation of the effects that we saw in the infinite population model. Both signaling system equilibria and partial pooling equilibria evolve in finite populations. Other possibilities are seen in a significant number of cases only in very small populations. The addition of mutation is, in general, conducive to the evolution of signaling. Even as it prevents perfect signaling it keeps the populations from getting stuck in suboptimal partial pooling equilibria (or worse). [Pawlowitsch (2007) studies a variant of the Moran process with *weak selection*. Using this, she is able to show that signaling system equilibria exhibit better stability properties in finite populations than partial pooling equilibria do.]

7 Conclusion

Analysis of evolution in Lewis signaling games (states equiprobable) using the replicator dynamics leads to the following conclusions:

1. Systems of information transmission spontaneously evolve in Lewis signaling games.
2. Perfect information transmission—signaling system equilibria—always arise in Boolean signaling games with states equiprobable—2 states, 2 signals, 2 acts—which are special in this regard.
3. In Lewis signaling games with $N > 2$, replicator dynamics sometimes leads to perfect information transmission (signaling system equilibria) and sometimes to imperfect information transmission (partial pooling equilibria).
4. Addition of mutation destabilizes total pooling equilibria, and appears to destabilize partial pooling equilibria and to lead to the evolution of signaling systems.

In finite, fixed-size populations in which evolutionary dynamics is modeled as a Moran process, these conclusions remain approximately valid unless the population is very small. In small populations [e.g. 100] without mutation all sorts of absorbing states are seen in simulations, and signaling systems go to fixation less than half the time. However, this is the case in which the addition of mutation makes the most dramatic contribution to the evolution of signaling.

It remains to be seen how far these conclusions generalize. (1) The possibility of the spontaneous emergence of signaling systems should hold under a wide variety of adaptive dynamics because signaling systems are strict equilibria. (2) If we relax the assumption of equiprobable states, total pooling equilibria can also become stable in the replicator dynamics, even in the Boolean case. These models are not structurally stable, and various perturbations remain to be investigated. [See Huttegger (2007a)].

This article has focused on spontaneous emergence of signaling using evolutionary dynamics. There is a parallel investigation of signaling using dynamics of individual learning. For a proof that in Boolean signaling games with states equiprobable, reinforcement learning converges to a signaling system with probability one, see Argiento et al. (forthcoming). For spontaneous emergence of coding under learning dynamics in a richer signaling game see Barrett (2007). Huttegger and Skyrms (forthcoming) investigate the impact of learning dynamics on the emergence of a simple network of signalers.

8 Appendix

In the main text it was claimed that a partial pooling plane attracts an open set of initial conditions. For completeness, we reproduce the proof of this result from Huttegger (2007a, Theorem 9 and proof thereof).

In Fig. 1, two sender strategies, S1 and S2, and two receiver strategies, R1 and R2, are defined. S1 and S2 both map states 1 and 2 to message 1, but S1 maps state 3 to message 2 while S2 maps state 3 to message 3. R1 and R2 map messages 2 and 3 on act 3, but R1 maps message 1 to act 1 whereas R2 maps message 1 to act 2. We denote the corresponding symmetrized strategies by $Z1 = (S1, R1)$, $Z2 = (S1, R2)$, $Z3 = (S2, R1)$

and Z4 = (S2, R2). In addition, let M be the boundary simplex spanned by Z1, . . . , Z4 (i.e. all convex combinations of the Zi's), and let N be the interior of M. It is easy to compute that the average payoff on M is 2/3. Hence M consists entirely of rest points for the replicator equations (since the average payoff is a Liapunov function for the replicator dynamics of partnership games). The corners of M can be destabilized by signaling systems (cf. Wärneryd 1993). Thus, we will study the stability properties of rest points (X,Y) in N.

Let us first prove that (X,Y) is quasi-strict, i.e. every best response to (X,Y) is contained in M. For every sender strategy S which is not identical to S1 or S2, W(S,Rj) <= 2/3 for j = 1,2 since state-act coordination is possible for at most two state-act pairs. The same is true for every alternative receiver strategy R, i.e. W(Si,R) <= 2/3 for i = 1,2. But if W(S,R1) = 2/3, then W(S,R2)<2/3 (analogous results hold if we interchange R1 and R2 and for R). Indeed, if W(S,R1) = 2/3, then S maps state 1 to message 1 and state 3 to message 2 or to message 3. In order to be an alternative to S1 and S2, S maps state 2 to a message other than message 1. But this implies that S and R2 are only able to coordinate state 3 and act 3.

Thus, for all Z' other than Z1, . . . , Z4, there exists a Zi (i = 1, . . . , 4) such that W(Z,Zi) < 2/3. From this it follows that for any Z = (X,Y) in N, W(Z',Z) < 2/3. Therefore, interior points of the set M are quasi-strict. This has important consequences for the replicator dynamics close to Z. The eigenvalues of the Jacobian matrix of the replicator equations at Z can be partitioned into the eigenvalues corresponding to the strategies Z1, . . . , Z4 and the transversal eigenvalues. The former eigenvalues must be zero, since M is a linear manifold of rest points. If Z' is a pure strategy outside the support of M, then the transversal eigenvalue for Z' is given by W(Z',Z) (cf. Cressman 2003). The quasi-strictness of points in M implies that all transversal eigenvalues are negative. Hence, it follows that Z is Liapunov stable since nearby solutions stay nearby. Moreover, the center-manifold theorem (Kelley 1967; Carr 1981) implies that there exists a local center manifold tangent to the zero eigenspace at Z. This center manifold coincides with a region of M that contains Z. To see this, note that the center manifold contains all rest points sufficiently close to Z and that it has two dimensions since Z has two zero-eigenvalues. Center-manifold theory then asserts that trajectories close to Z converge to Z's center manifold. Since this analysis holds for any Z in M, M attracts a set of initial conditions of positive measure. (For more on the use of center-manifold theory in evolutionary games see Cressman 1992, 2003).

References

Akin, E. & Hofbauer, J. (1982). Recurrence of the unfit. *Mathematical Biosciences, 61*, 51–62.

Argiento, R., Pemantle, R., Skyrms, B. & Volkov, S. (forthcoming). Learning to signal: Analysis of a micro-level reinforcement model. *Stochastic Processes and their Applications*.

Balkenborg, K. & Schlag, K. H. (2007). On the evolutionary selection of sets of Nash equilibria. *Journal of Economic Theory, 133*, 295–315.

Barrett, J. (2007). Evolution of coding in signaling games. *Theory and Decision*.

Binmore, K. & Samuelson, L. (1999). Evolutionary drift and equilibrium selection. *Review of Economic Studies, 66*, 363–393.
Carr, J. (1981). *Applications of centre manifold theory*. New York: Springer.
Crawford, V. & Sobel, J. (1982). Strategic information transmission. *Econometrica, 50,* 1431–1451.
Cressman, R. (1992). *The stability concept of evolutionary game theory: A dynamic approach*. New York: Springer.
Cressman, R. (2003). *Evolutionary dynamics and extensive form games*. Cambridge: MIT Press.
Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge: MIT Press.
Fudenberg, D., Imhof, L., Nowak, M. A., & Taylor, C. (2004). Stochastic evolution as a generalized Moran process. Working paper, Harvard University.
Gale, J., Binmore, K. & Samuelson, L. (1995). Learning to be imperfect: The ultimatum game. *Games and Economic Behavior, 8*, 56–90.
Guckenheimer, J. & Holmes, P. (1983). *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. New York: Springer.
Hadeler, K. P. (1981). Stable polymorphisms in a selection model with mutation. *SIAM Journal of Applied Mathematics, 41*, 1–7.
Harms, W. (2004). *Information and meaning in evolutionary processes*. New York: Cambridge University Press.
Hofbauer, J. (1985). The selection-mutation equation. *Journal of Mathematical Biology, 23*, 41–53.
Hofbauer, J. & Huttegger, S. M. (2007). Feasibility of communication in binary signaling games. Working paper, University of Vienna.
Hofbauer, J. & Sigmund, K. (1998) *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.
Huttegger, S. M. (2007a). Evolution and the explanation of meaning. *Philosophy of Science, 74*, 1–27.
Huttegger, S. (2007b). Evolutionary explanations of indicatives and imperatives. *Erkenntnis, 66*, 409–436.
Huttegger, S. M. & Skyrms, B. (forthcoming). Learning to transfer information. *Studia Logica*.
Jäger, G. (2007). Evolutionary stability conditions for signaling games with costly signals. Manuscript, University of Bielefeld.
Kelley, A. (1967). The stable, center-stable, center, center-unstable, unstable manifolds. *Journal of Differential Equations, 3*, 546–570.
Lewis, D. (1969). *Convention*. Cambridge: Harvard University Press.
Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge: MIT Press.
Millikan, R. G. (1993). *White queen psychology and other essays for alice*. Cambridge: MIT Press.
Millikan, R. G. (1996). Pushmi-Pullyu representations. In L. May, M. Friedman & A. Clark (Eds.), *Mind and morals: Essays in cognitive science and ethics* (pp. 145–161). Cambridge: MIT Press.
Moran, P.A.P. (1962). *The statistical processes of evolutionary theory*. Oxford: Clarendon Press.
Pawlowitsch, C. (2008). Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior, 63*, 203–226.
Pawlowitsch, C. (2007). Finite populations choose an optimal language. *Journal of Theoretical Biology, 249*, 606–616.
Schreiber, S. (2001). Urn models, replicator processes and random genetic drift. *Siam Journal of Applied Mathematics, 61*, 2148–2167.
Skyrms, B. (1999). Stability and explanatory significance of some simple evolutionary models. *Philosophy of Science, 67*, 94–113.
Skyrms, B. (2005). Dynamics of conformist bias. *Monist, 88*, 260–269.
Skyrms, B. (forthcoming). "Signals" presidential address of the philosophy of science association. *Philosophy of Science*.
Taylor, P. & Jonker, L. (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences, 40*, 145–156.
Wärneryd, K. (1993). Cheap talk, coordination, and evolutionary stability. *Games and Economic Behavior, 5*, 532–546.