Some dynamics of signaling games

Simon M. Huttegger, B. Skyrms, Pierre Tarrès, and Elliott O. Wagner

and

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Information transfer is a basic feature of life that includes signaling within and between organisms. Due to its interactive nature, signaling can be investigated by using game theory. Game theoretic models of signaling have a long tradition in biology, economics and philosophy. For a long time the analyses of these games has mostly relied on using static equilibrium concepts such as Pareto optimal Nash equilibria or evolutionarily stable strategies. More recently signaling games of various types have been investigated with the help of game dynamics, which includes dynamical models of evolution and individual learning. A dynamical analysis leads to more nuanced conclusions as to the outcomes of signaling interactions. Here we explore different kinds of signaling games that range from interactions without conflicts of interest between the players to interactions where their interests are seriously misaligned. We consider these games within the context of evolutionary dynamics (both infinite and finite population models) and learning dynamics (reinforcement learning). Some results are specific features of a particular dynamical model, while others turn out to be quite robust across different models. This suggests that there are certain qualitative aspects that are common to many real world signaling interactions.

signaling games | costly signaling games | evolutionary dynamics | learning dynamics

Introduction

The flow of information is a central issue across biological and social science. In both of those domains, entities have information that can be communicated, wholly or partly, to other entities by means of signals. Signaling games are abstractions that are useful for studying general aspects of such interactions. The simplest signaling games model interactions between two individuals: a sender and a receiver. The sender acquires private information about the state of the world, and contingent on that information selects a signal to send to the receiver. The receiver observes the signal, and contingent on the signal observed chooses an action. Payoffs for sender and receiver are functions of state of the world, action chosen, and (possibly) signal sent. Where payoffs only depend on state and act, interests of sender and receiver may be coincident, partially aligned, or totally opposed. Another layer of complexity is added when payoffs reflect differential costs to the sender, of different signals in different states.

The baseline case is given by signaling when the interests of the sender and the receiver are fully aligned. This scenario was introduced by the philosopher David Lewis in 1969 in order to analyze conventional meaning (25). We thus call them *Lewis signaling games*. In the simplest case the action chosen by the receiver is appropriate for exactly one state of the world. If the appropriate action is chosen, then the sender and the receiver get the same payoff of, say, 1; otherwise they get a payoff of 0.

By varying these payoffs, the sender's and the receiver's incentives may change quite radically (6). Sometimes the sender might have an incentive to not inform the receiver about which of several states is the true one because the action that would be best for the sender does not coincide with the action that would be best for the receiver in a certain state. An extreme form that will be discussed briefly below is arrived at when this is always the case so that the signaling interaction is essentially zero-sum.

Misaligned interests lead to the question of how reliable or honest signaling is possible in such cases (45; 10; 26). The resulting games are known as *costly signaling games* because in these games each signal may have an associated cost. Costly signaling games are studied in economics, starting with the Spence game (37), and in biology (e.g. 10). The Spence game is a model of job market signaling. In a job market, employers would like to hire qualified job candidates, but the level of a job candidate's qualification is not directly observable. Instead, a job candidate can send a signal about her qualification to the employer. However, if signals are costless, then job candidates will choose to signal that they are highly qualified regardless of their qualification. If, on the other hand, signals are costly so that only job candidates of high qualification can afford to send it, signaling can be honest in equilibrium whenever the cost of signals is sufficiently high. In this case employers get reliable information about a job candidate's quality.

The models of costly signaling in theoretical biology have a similar structure. They model situations as diverse as predator-prey signaling, sexual signaling, or parent-offspring interactions (35). A simple example of the latter kind of situation is the so-called Sir Philipp Sidney game, which was introduced by John Maynard Smith in order to capture the basic structure of costly signaling interactions (26). In the Sir Philipp Sidney game there are two players, a child (sender) and a parent (receiver). The sender can be in one of two states, needy or not needy, and would like to be fed in either state. The receiver can choose between feeding the sender or abstaining from doing so. She would like to feed the sender provided that the sender is needy. Otherwise the receiver would rather eat herself. This creates a partial conflict of interest between the sender and the receiver. If the sender is needy, then the interactions between the two players is similar to matching a states and an act in the Lewis signaling game. However, if the sender is not needy, then the payoffs of sender and receiver diverge.

Now, it is assumed that the a needy sender profits more from being fed than a sender that is not needy. In addition, the sender is allowed to send a costly signal. If the cost of the signal is sufficiently high, then there is again the possibility that in equilibrium the sender signals need honestly and the receiver feeds the sender upon receipt of the signal.

In costly signaling games, signaling costs realign the interests of sender and receiver by making it disadvantageous for certain types of senders to use signals. This captures quite precisely the so-called *handicap principle* (45), which states that there must be a significant differential cost in order for honest signaling to be possible.

Much of the analysis in the literature on these games has focused on the most mutually beneficial (Pareto optimal) equilibria of the games under consideration. This would lead one to expect perfect information transfer in partnership games, partial information transfer in games of partially aligned interests, and no information trans-

Reserved for Publication Footnotes

fer in games of totally opposed interests. Perfect information transfer might be restored in problematic cases by the right differential signaling costs.

But we do not want to simply rely on faith that Pareto optimal equilibria will be reached. What is required is an investigation of an adaptive dynamic that may plausibly be operative. Many dynamic processes deserve consideration. Here we focus on some dynamics of evolution and of reinforcement learning, where sharp results are available.

Replicator dynamics

The replicator dynamics is the fundamental dynamical model of evolutionary game theory (17). It describes evolutionary change in terms of the difference between a strategy's average payoff and the overall average payoff in the population. If the difference is positive, the strategy's share will increase; if it is negative, it will decrease. This is one way to capture a basic feature of any selection dynamics. Not surprisingly, the replicator dynamics can be derived from various first principles which describe selection more directly (43).

The two most common varieties of replicator dynamics are the *one population* and the *two population* replicator dynamics. The one population replicator dynamics can be used for symmetric two player games (i.e. two-player games where the player roles are indistinguishable). Let $s_1, \ldots, s_n$ denote the n pure strategies that are available to each player. Let $\pi_i(s_j, s_k)$ be the payoff that player i receives from choosing s_j when the other player is choosing s_k. Since the game is symmetric, $\pi_1(s_j, s_k) = \pi_2(s_j, s_k)$. This allows us to drop the indices referring to a player's payoff when considering symmetric games. Let $\mathbf{x} = (x_1, \ldots, x_n)$ denote a mixed strategy. Then $\pi(s_i, \mathbf{x})$ is the expected payoff that a player gets when choosing s_1 against $\mathbf{x}$:

$$\pi(s_i, \mathbf{x}) = \sum_j \pi(s_i, s_j) x_j$$

Furthermore, $\pi(\mathbf{x}, \mathbf{x})$ is the expected payoff from choosing $\mathbf{x}$ against itself:

$$\pi(\mathbf{x}, \mathbf{x}) = \sum_j \pi(s_j, \mathbf{x}) x_j$$

Suppose now that there is a population consisting of n types, one for each strategy. Then a mixed strategy $\mathbf{x}$ describes the relative frequency of strategies in that population. The state space of the population is the $n-1$ dimensional unit simplex. The population evolves according to the replicator dynamics if its instantaneous change is given by

$$\dot{x}_i = x_i(\pi(s_i, \mathbf{x}) - \pi(\mathbf{x}, \mathbf{x})) \quad \text{for } 1 \leq i \leq n. \qquad \textbf{[1]}$$

Here, the payoff $\pi(s_i, \mathbf{x})$ is interpreted as the fitness of an i strategist in the population state $\mathbf{x}$, and $\pi(\mathbf{x}, \mathbf{x})$ is the average fitness of that population. Since these fitnesses are expected payoffs, the replicator dynamics requires the population to be essentially infinite.

The two population replicator dynamics can be applied to asymmetric two-player games. Let $s_1, \ldots, s_n$ be player one's pure strategies and $t_1, \ldots, t_m$ player two's pure strategies. The mixed strategies $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_m)$ can be identified with the states of two populations, one corresponding to player one and the other to player two. The state space for an evolutionary dynamics of the two populations is the product space of the $n-1$-dimensional unit simplex and the $m-1$-dimensional unit simplex. The two population replicator dynamics defined on this state space is given by

$$\dot{x}_i = x_i(\pi_1(s_i, \mathbf{y}) - \pi_1(\mathbf{x}, \mathbf{y})) \quad \text{for } 1 \leq i \leq n \text{ and} \qquad \textbf{[2a]}$$

$$\dot{y}_j = y_j(\pi_2(t_j, \mathbf{x}) - \pi_2(\mathbf{y}, \mathbf{x})) \quad \text{for } 1 \leq j \leq m. \qquad \textbf{[2b]}$$

Here, $\pi_1(s_i, \mathbf{y})$ is the fitness (expected payoff) of choosing strategy s_i against population state (mixed strategy) $\mathbf{y}$, and $\pi_1(\mathbf{x}, \mathbf{y})$ is the average fitness in population one; likewise for $\pi_2(t_j, \mathbf{x})$ and $\pi_2(\mathbf{y}, \mathbf{x})$.

Both the two population and the one population replicator dynamics are driven by the difference between a strategy's fitness and the average fitness in its population. This captures the mean field effects of natural selection, but it disregards other factors such as mutation or drift. In many games these factors will only play a minor role compared to selection. But as we shall see, the evolutionary dynamics of signaling games often crucially depends on these other factors. The reason is that the replicator dynamics of signaling games is generally not *structurally stable* (11). This means that small changes in the underlying dynamics can lead to qualitative changes in the solution trajectories.

This makes it important to study the effect of perturbations of the replicator dynamics. One plausible deterministic perturbation that has been studied is the *selection mutation dynamics* (14). We shall consider this dynamics in the context of two population models. If mutations within each population are uniform, the selection mutation dynamics is given by

$$\dot{x}_i = x_i(\pi_1(s_i, \mathbf{y}) - \pi_1(\mathbf{x}, \mathbf{y})) + \varepsilon(1 - nx_i) \quad \text{for } 1 \leq i \leq n \text{ and} \qquad \textbf{[3a]}$$

$$\dot{y}_j = y_j(\pi_2(t_j, \mathbf{x}) - \pi_2(\mathbf{y}, \mathbf{x})) + \delta(1 - my_j) \quad \text{for } 1 \leq j \leq m. \qquad \textbf{[3b]}$$

The non-negative parameters ε and δ are the (uniform) mutation rates in population one and two, respectively. Instantaneously, every strategy in a population is equally likely to mutate into any other strategy at a presumably small rate. As ε and δ go to zero, the two population selection mutation dynamics approaches the two population replicator dynamics. If the replicator dynamics is structurally stable, there will be no essential difference between the replicator dynamics and the selection mutation dynamics as long as ε, δ are small. However, the introduction of deterministic mutation terms can significantly alter the replicator dynamics of signaling games.

Lewis signaling games. From the point of view of static game theory, the analysis of Lewis signaling games seems to be straightforward. If the number of signals, states and acts coincide, then signaling systems are the only strict Nash equilibria. It can also be shown that they are the only evolutionarily stable states (42). However, other Nash equilibria, despite being non-strict, are neutrally stable states (a generalization of evolutionary stability)(30). This suggests that an analysis of the evolutionary dynamics will reveal a more fine-grained picture, as indeed it does.

Consider the one population replicator dynamics first. The Lewis signaling game as given in the preceding section is not a symmetric game. It can, however, be symmetrized by assuming that a player assumes the roles of a sender and a receiver with equal probability and receives the corresponding expected payoffs (7). If there are n signals, states and acts, the symmetrized signaling game will have n^{2n} strategies. This results in a formidable number of dimensions for the state space of the corresponding dynamics **[1]** even for relatively small n. A fairly complete analysis of this dynamical system is nevertheless possible because the Lewis signaling game exhibits certain symmetries. The assumption that both players get the same payoff in every outcome makes it a *partnership game*, a class of games for which it is known that the average payoff $\pi(\mathbf{x}, \mathbf{x})$ is a *potential function* (17). This implies that every solution trajectory converges to a rest point, which need to be Nash equilibria. The stable rest points are the local maximizers of $\pi(\mathbf{x}, \mathbf{x})$.

It is clear that signaling systems are locally asymptotically stable since they are strict Nash equilibria. The question is whether there are any other locally asymptotically stable rest points. It can be proved that this is essentially not the case for signaling games with two states, signals and acts, where both states are equally probably (20): every open set in the strategy simplex contains an $\mathbf{x}$ such that the replicator dynamics with this initial condition converges to a signaling system. This rather special case does not generalize, however. If the states are not equi-probable in this signaling game, there is an

open set of trajectories that does not converge to a signaling system. Instead, they converge to states where receivers always choose the act corresponding to the high-probability state (20).

If there are more than two signals, states and acts, there always is an open set of trajectories that does not converge to a signaling system (20; 30). The rest points to which they converge are often referred to as 'partial pooling equilibria' (21). Partial pooling equilibria share three features: (i) some, but not all, signals are unequivocally used for states; (ii) some, but not all acts are unequivocally chosen in response to a signal; and (iii) no signal is unused. The last feature makes it impossible for mutants who use a signaling system strategy to invade by exploiting an unused signal. A set of partial pooling equilibria P is not an attractor since for any neighborhood N of P there exists a solution trajectory that leaves N. A partial pooling equilibrium is Liapunov stable, though. In addition, in any sufficiently small neighborhood of a partial pooling equilibrium all solution trajectories that do not start at a pooling equilibrium converge to one. As is shown in (30), partial pooling equilibria coincide with those neutrally stable states that are not also evolutionarily stable states.

A result that holds for all Lewis signaling games concerns the instability of interior rest points. At an interior rest point, every strategy is present, creating a 'tower of Babel' situation. It can be shown that any such rest point is linearly unstable. This implies that the unstable manifold of these rest points has a dimension of at least one. Hence there is no open set of trajectories converging to the set of interior rest points (20).

These results carry over to the case of two populations. The game is again a partnership game. It follows that $\pi_1(\mathbf{x}, \mathbf{y}) = \pi_2(\mathbf{y}, \mathbf{x})$ is a potential function of **[2]** and that every trajectory must converge. Signaling systems are asymptotically stable by virtue of being strict Nash equilibria. For signaling games with two signals, states and acts where the states are equiprobable essentially all trajectories converge to a signaling system. This result fails to hold when the states are not equiprobable (15). Furthermore, partial pooling equilibria are stable for **[2]** as in the one population case.

The selection mutation dynamics **[3]** of Lewis signaling games was studied computationally in (21) and analytically in (15; 16). The main reason for studying selection mutation dynamics **[3]** is that partial pooling equilibria are not isolated. They constitute linear manifolds of rest points. It is well known that this situation is not structurally stable. Introducing mutation terms as in **[3]** will destroy the linear manifolds of rest points and create a topologically different dynamics in that region of state space.

For other games this topic was studied in (5). In (15) it is shown that the function

$$\pi_1(\mathbf{x}, \mathbf{y}) + \varepsilon \sum_i \log x_i + \delta \sum_j \log y_j$$

is a potential function for the selection mutation dynamics of the Lewis signaling game. Hence all trajectories converge. There are two additional general results. The first says that rest points of the perturbed dynamics **[3]** must be close to Nash equilibria of the signaling game. There are thus no 'anomalous' rest points that are far away from any Nash equilibrium. Second, there is a unique rest point close to any signaling system that is asymptotically stable. Signaling systems remain evolutionarily significant.

There are no further general results. In (15) the case of two states, two signals and two acts is explored in more detail. If the states are equally probable, the overall conclusions are similar to the results of the replicator dynamics. There are three rest points: two are close to signaling systems (perturbed signaling systems), and the third is the barycenter of the state space. The latter is linearly unstable while the perturbed signaling systems are linearly stable. So, although there are only finitely many rest points for the selection mutation dynamics (as opposed to the replicator dynamics) of this game, the basic conclusion is that for every open set in the state space there is an $\mathbf{x}$ such that the selection mutation dynamics of **[3]** with this initial condition converges to a signaling system.

Things are more nuanced if the two states are not equiprobable. In this case the dynamic behavior depends on the ratio of the mutation parameters δ/ε. If δ/ε is above a certain threshold, which includes the important case $\delta = \varepsilon$, then almost all trajectories converge to one of the signaling systems. If δ/ε is below the threshold, then there exists an asymptotically stable rest point where nearly all members of the receiver population choose the act that corresponds to the more probable signal. Thus outcomes with basically no communication are robust under the introduction of mutation into the replicator equations **[2]**,

It is very difficult to analyze the selection mutation dynamics **[3]** of Lewis signaling games for the case of three states, acts and signals. The main reason is the rapidly increasing dimensionality of the state space. In a two-population model, there are 27 types of individuals in each population, resulting in the product of two simplexes that has 52-dimensions. By exploiting the underlying symmetry of the Lewis signaling game and by introducing certain simplifications, it is nonetheless possible to prove some results (16). The main results concern the existence of rest points close to Nash equilibria other than the signaling systems. Most notably, at least one rest point exists close to each component of partially pooling Nash equilibria. It can be shown that for all mutation parameters this rest point is linearly unstable.

Costly signaling games. Several costly signaling games were studied with the help of the replicator dynamics: the Spence game and the Sir Philipp Sidney game. Additionally, simplified versions of the Sir Philip Sidney game and related games were analyzed recently (46). For all these games, the replicator dynamics leads very similar results, which can differ quite markedly to those obtained in finite population models (see the subsequent section 'finite population dynamics').

The dynamics of Spence's model for job market signaling is explored in (28) for a dynamic process of belief and strategy revision that is different from the dynamical models considered here, although the analysis leads to somewhat similar results. In (39) the two-population replicator dynamics **[2]** of Spence's game is investigated. (Strictly speaking, it is a discretized variant of Spence's original game which has a continuum of strategies.) For all parameter settings there exists a pooling equilibrium where senders don't send a signal and receivers ignore the signal. If signaling cost is sufficiently high, then a separating equilibrium exists; the separating equilibrium can be viewed as the analogue to a Lewis signaling system where signals are used for revealing information about the sender. If signaling cost is not high enough, a so-called hybrid equilibrium exists where senders mix between using signals reliably and unreliably, and where receivers sometimes respond to a signal and sometimes ignore it. As is pointed out in (39), the hybrid equilibrium has been almost completely ignored in the large literature on costly signaling games, although it provides an interesting low-cost alternative to the standard separating equilibrium (46).

It can be shown that the pooling equilibrium in this version of Spence's game is always asymptotically stable for the replicator dynamics **[2]**. The same is true for the separating equilibrium, provided that it exists. The dynamic behavior of the hybrid equilibrium is particularly interesting. It is Liapunov stable. More precisely, it is a spiraling center. It lies on a plane on the boundary of state space. With respect to the plane, the eigenvalues of the Jacobian matrix evaluated at the hybrid equilibrium have zero real part, which makes it into a center when we restrict the dynamics to the plane. More precisely, on the plane the trajectories off the rest point are periodic cycles. With respect to the interior of the state space, the eigenvalues of the Jacobian evaluated at the hybrid equilibrium are negative. Hence from the interior trajectories approach the hybrid equilibrium in a spiraling movement.

The same is true for other costly signaling models such as the well known Sir Philipp Sidney game (23; 46). In particular, hybrid equilibria exist and are dynamically stable for **[2]** in the same way as for the job market signaling game. This suggests that hybrid equilibria can be evolutionarily significant outcomes. Both (39; 23) present numerical simulations that reinforce this conclusion in terms of the relative sizes of basins of attraction. According to these simulations, the basin of attraction of hybrid equilibria is quite significant, while the basin of attraction for separating equilibria is surprisingly small.

A question that has only recently been investigated is whether the hybrid equilibrium continues to be dynamically stable under perturbations of the dynamics **[2]**. The question of structural stability is important here as well because a spiraling center is not structurally stable. Small perturbations of the dynamics will push the eigenvalues with zero real part to having positive or negative real part. So it seems possible that by introducing mutation as in **[3]** the hybrid equilibrium might cease to be dynamically stable.

That this is not so for sufficiently small mutation parameters is shown in (22). First, it follows from the implicit function theorem that there exists a unique rest point of **[3]** close to the hybrid equilibrium (46). Second, it can be proved that all eigenvalues of the Jacobian matrix of **[3]** evaluated at this perturbed rest point are negative. Thus the rest point corresponding to the hybrid equilibrium is not a spiraling center anymore but is asymptotically stable instead. This actually reinforces the qualitative point made above, namely, that hybrid equilibria should be considered as theoretically significant evolutionary outcomes.

Another question is whether the hybrid equilibrium is also empirically significant. One of the most robust findings in costly signaling experiments and field studies is that observed costs are generally too low in order to validate the hypothesis that the handicap principle is at work (35). The hybrid equilibrium is in certain ways an attractive alternative to the handicap principle. It allows for partial information transfer at low costs. For this reason it was suggested that hybrid equilibria could be detected and distinguished from separating equilibria in real world signaling interactions (46).

Opposed Interests. The possibility of signaling when interests conflict has also been studied in a rather extreme setting where the interests of senders and receivers are completely opposed (40). There is no signaling equilibrium possible in this case. However, in the two-population replicator dynamics **[2]** there is information transfer off equilibrium to varying degrees since there exists a strange attractor in the interior of the state space. By information transfer we understand that signals have information in the sense of Kullback-Leibler entropy: Conditioning on the signal changes the probability of states so that on average information is gained (for details on applying information theory to signaling games see 36).This result reinforces the diagnosis that a dynamical analysis is unavoidable if one wants to understand the evolutionary significance of signaling phenomena.

Finite population dynamics

Consider a small finite population of fixed size. Each step of the dynamics works as follows. First, everyone plays the base game with everyone else in a round robin fashion. Each individual's fitness is given by a combination of her background fitness and her average payoff from the round robin tournament. Following (27) we will take the fitness to be

$$1 - w + w \times u(s_i, \mathbf{x})$$

where $w \in [0, 1]$ is a parameter that measures the intensity of selection and $u(s_i, \mathbf{x})$ is the expected payoff of strategy i against the population $\mathbf{x}$, just as in the case of the replicator dynamics.

The background fitness w is the same for everyone. If $w = 0$ the game's payoffs do not matter to an individual's fitness. If $w = 1$ the game's payoffs are all that matter. Next, one individual is selected at random to die (or to leave the group) and a new individual is born (or enters the group). The new individual adopts the strategy of an individual chosen from the population with probability proportional to its fitness. Successful strategies are more likely to be adopted and will therefore spread through the population. This dynamics, known as the frequency-dependent Moran process, is a Markov chain with the state being the number of individuals playing each strategy. Due to the absence of mutation or experimentation all monomorphic population compositions are absorbing states of this process.

Lewis signaling games. Pawlowitsch (29) studies symmetrized Lewis signaling games under this dynamics. Let N be the number of individuals all playing strategy s_j. Imagine that one spontaneously mutates to s_i. The probability that strategy s_i goes on to take over the entire population is given by the fixation probability

$$\rho_{s_j s_i} = \frac{1}{1 + \sum_{k=1}^{N-1} \prod_{l=1}^{k} \frac{g_l(s_i, s_j)}{f_l(s_i, s_j)}}$$

where $f_l(s_i, s_j)$ is the fitness of an s_i agent in a population of l individuals playing strategy s_i and $N - l$ individuals playing s_j and $g_l(s_i, s_j)$ is the fitness of an s_j individual in that same population. If the mutation is neutral then its probability of fixation is $1/N$. Pawlowitsch (29) uses this neutral threshold to assess the evolutionary stability of monomorphic population compositions. In Lewis signaling games under the Moran process with weak selection ($wN \ll N$) Pareto optimal strategies—i.e., perfectly informative signaling strategies—are the only strategies for which there is no mutant type that has a fixation probability greater than this neutral threshold. For this reason Pawlowitsch argues that finite populations will choose an optimal language in Lewis signaling games. This highlights an important difference between the behaviors of infinite and finite populations in these games.

Costly signaling games. But what about signaling games where interests conflict? To address this question we can consider a slight variation of the frequency dependent Moran process described above. In particular, suppose that with small probability, ε, the new individual mutates or decides to experiment and chooses a strategy at random from all the possible strategies–including those not represented in the population–with equal probability. The presence of this mutation makes the resulting Markov process ergodic. Fudenberg and Imhof (9) show that it is possible to use a so-called embedded Markov chain to calculate the proportion of time that the population spends in each state in the limiting case as ε goes to zero. The states in this embedded Markov chain are the monomorphic population compositions, and the transition probability from the monomorphic population in which all individuals play s_j to the monomorphic population in which all individuals play s_i is given by the probability that a type s_i mutant arises (ε) multiplied by the probability that this mutant fixes in the population ($\rho_{s_j s_i}$). The stationary distribution of this embedded chain gives the proportion of time that the population spends in each monomorphic state in the full Moran process as ε goes to zero. Intuitively, this is because when ε is very small the system will spend almost all of the time in a monomorphic state waiting for the next mutation event, and after an event the Moran process will return the population to a monomorphic state before the next mutant arises.

Consider the 2 state, 2 signal, 2 act signaling game with payoffs given in figure 1. The receiver prefers the act high in the high

	Act High	Act Low
State High	$1, 1$	$0, 0$
State Low	$1, 0$	$.8, 1$

Fig. 1. The payoff structure underlying a signaling game. If the state is High, then the sender's and the receiver's interest coincide. If the state is Low, then the sender prefers the receiver to choose High, while the receiver would want to choose Low.

state and low in the low state, whereas the sender always prefers the act high. This game is structurally similar to the Sir Philip Sidney game, but here we will assume that both signals are costless. In this game there is no Nash equilibrium in which the signals discriminate between the states. The only Nash equilibria are pooling, where the sender sends signals with probabilities independent of the states and the receiver acts low.

The Moran process is a one-population setting, so we will consider the symmetrized version of this game (7). We then let our round robin phase match each pair both as (sender, receiver) and (receiver, sender). Suppose that selection is strong (i.e., $w = 1$), and that the probability of state high is .4. Then in the small mutation limit the process spends 57% of its time in states with perfect signaling and 19% of its time in Nash equilibria (41). In other words, this small population spends most of its time communicating perfectly even though such information transfer is not a Nash equilibrium of the underlying signaling game. This phenomenon is robust over a wide range of selection intensities and state probabilities, but as the population size is increased the proportion of time spent signaling diminishes. This is to be expected because as the population size tends to infinity the behavior of the Moran process tends toward the behavior of the payoff-adjusted replicator dynamic (38), which does not lead to information transfer in this game.

The importance of non-Nash play in the rare mutation limit is also evident in the case of cost-free pre-play signaling. Two players play a base game, but before they do so each sends the other a cost-free signal from some set of available signals, with no pre-existing meaning. The small population, rare mutation limit for a related dynamics is investigated in (34) for the cases where the base game is (i) Stag Hunt and (ii) Prisoner's Dilemma. Like the Moran process, this related dynamics is composed of two steps. First, all individuals play the base game with each other in a round robin fashion to establish fitness. Second, an individual is randomly selected to update her strategy by imitation. This agent randomly selects another individual and imitates that other individual with a probability that increases with an increase in the fitness difference between the two agents. In particular, this probability is given by the Fermi distribution from statistical physics so that the probability that an individual using strategy s_i will imitate an individual using strategy s_j is given by the function

$$[1 + e^{-\beta[\pi(s_i,\mathbf{x})-\pi(s_j,\mathbf{x})]}]^{-1}.$$

where β should be interpreted as noise in the imitation process. For high values of β a small payoff difference translates into a high probability of imitation, whereas when β tends to zero selection is weak and the process is dominated by random drift (38).

In the case where the base game is the Prisoner's Dilemma, the only Nash equilibrium is non-cooperation. But if the population is small and the set of signals is large, the population may spend most of its time cooperating. If the base game is the Stag Hunt, where there are both co-operative and non-cooperative equilibria, pre-play signals enlarge the amount of time spent cooperating, with the more signals the better.

Why is it that dynamics in finite populations with rare mutations can favor informative signaling even when such behavior is not a Nash equilibrium? Consider again the game in figure 1. In a pooling equilibrium the sender's expected payoff is .48 and the receiver's is .6. Jointly separating, however, garners the sender an expected payoff of .88 and the receiver an expected payoff of 1. Signaling Pareto dominates pooling, and this means that a small population is likely to transition from a monomorphic pooling state to a monomorphic signaling state. Of course, if the receiver discriminates, then the sender can gain by always sending whichever signal induces act high. Such behavior will net the sender an expected payoff of 1. But note that there is a smaller difference in payoff for the sender between this profile and the separating profile than there is between the separating profile and the pooling equilibrium. Consequently, the probability of transitioning away from a monomorphic separating population to the monomorphic population in which the senders always induce the receivers to perform act high is less than the probability of transitioning from a monomorphic pooling population to a monomorphic separating population. For this reason, in the long run the population will spend more time signaling informatively than it will spend pooling. A similar story explains why finite population dynamics can favor informative signaling and cooperation in prisoner's dilemma games with cost-free pre-play signaling (34).

Reinforcement learning

In models from evolutionary game theory—such as the replicator dynamics—"learning" occurs globally since the size of populations with more fitness increases faster. However, in the reinforcement learning model it is the individuals' behavior that evolves iteratively: the players tend to put more weight on strategies that have enjoyed past success, as measured by the cumulative payoffs they have achieved. This linear response rule corresponds to Herrnstein's "matching law" (13).

Reinforcement learning is one of a variety of models of strategic learning in games, where players adapt their strategies with the aim to eventually maximize their payoffs: no-regret learning, fictitious play and its variants, and hypothesis testing, are other examples of such procedures analyzed in game theory (44).

However, reinforcement learning is a particularly attractive and simple model of players with bounded rationality. The amount of information used in the procedure is small: players need only observe their realized payoffs, and may not even be aware that they are playing a game with or against others. It accumulates inertia, since the relative increase in payoff decreases with time. On the one hand this might be expected from a learning procedure, although on the other it could be exploited by other players in certain games.

In Erev and Roth (33; 8) reinforcement learning is proposed and tested experimentally as a realistic model for the behavior of agents in games—see Harley (12) for a similar study in a biological context.

Formally, in a game played repeatedly by N players, each having M strategies, each individual i is assumed at a time step n to have a propensity q^i_{jn} for each strategy j, and plays the strategy with probability proportional to its propensity,

$$\frac{q^i_{jn}}{\sum_{k=1}^{m} q^i_{kn}}. \qquad [4]$$

Each individual i is endowed with an initial vector of positive weights $(q^i_{j0})_j$ at time 0. At each iteration of the learning process, the strategy j taken by any player i results in a non-negative payoff, $U_j(i)$, and the weights are updated by adding that payoff to the weight of the act taken:

$$q^i_{j,t+1} = q^i_{j,t} + U_j(i), \qquad [5]$$

with the weights of strategies not taken remaining the same.

This process can be exemplified by an urn model. Each individual starts with an urn containing some balls of different colors, one for each potential strategy. Drawing a ball from the urn (and then replacing it) determines the choice of strategy. After receiving a payoff, the number of balls of the same color equal to the payoff achieved are added to the urn.

As balls pile up in the urn, jumps in probabilities become smaller and smaller in such a way that the stochastic process approximates a deterministic mean field dynamics, which is known as the adjusted or Maynard-Smith version of the replicator dynamics (2; 18). However, classical stochastic approximation theory (32; 3; 4) does not allow to deduce much in general, since this ordinary differential equation takes place in an unbounded domain.

Beggs shows in (2) that, if all players apply this rule, then iteratively strictly dominated strategies are eliminated; and the long-run

average payoff of a player who applies it cannot be forced permanently below its minmax payoff. He also studies two person constant sum games, where precise results can be obtained. Hopkins and Posch (18) show convergence with probability 0 towards unstable fixed points of the Maynard-Smith replicator dynamics, even if they are on the boundary, which solves earlier questions raised in particular in (24). This second result is however not relevant in signaling games, where the unstable fixed points are not isolated and consist of manifolds of finite dimension (1; 19).

Consider the simplest Lewis signaling game. The reinforcement learning model proposed in (1) considers, in equations **[4]**-**[5]**, each state possibly transmitted by the sender as a player whose strategies are the signals and, similarly, each signal as a player whose strategies are the (guessed) states. Now a state i which "plays" signal j gets a payoff of 1 if, conversely, j "plays" i and Nature chooses i. In practice, at each time step, only the state chosen by Nature will play along with its chosen signal, so that this reinforcement procedure can be simply explained from a Sender-Receiver perspective.

Let us first consider the case of two states 1 and 2: Nature flips a fair coin and chooses one of them. The sender has an urn for state 1 and a different urn for state 2. Each has balls for signal A and signal B. The sender draws from the urn corresponding to the state, and sends the indicated signal. The receiver has an urn for each signal, each containing balls for state 1 and for state 2. The receiver draws from the urn corresponding to the signal and guesses the indicated state. If correct, both sender and receiver are reinforced and each adds a duplicate ball to the urn just exercised. If incorrect, there is no reinforcement and the urns are unchanged for the next iteration of the process.

There are now four interacting urns contributing to this reinforcement process. However, the dimensionality of the process can be reduced because of the symmetry resulting from the strong common interest assumption. Since the receiver is reinforced if and only if the sender is, the numbers of balls in the receiver's urns are determined by the numbers of balls in the sender's urns. Consider the four numbers of balls in the sender's urns: $1A$, $1B$, $2A$, $2B$. Normalizing these, $1A/(1A + 1B + 2A + 2B)$ etc., gives four quantities that live on a tetrahedron. The mean field dynamics may be written in terms of these. There is a Lyapunov function that rules out cycles. The stochastic process must then converge to one of the zeros of the mean field dynamics. These consist of the two signaling systems and a surface composed of pooling equilibria. It is possible to show that the probability of converging to a pooling equilibrium is zero. Thus, reinforcement learning converges to a signaling system with probability one (1).

This raises several questions. Does the same result hold for N states, N signals, and N acts? What happens if there are too few signals to identify all the states or if there is an excess of signals—that is, N states, M signals, N acts? Nature now rolls a fair die to choose the state, the sender has N urns with balls of M colors and the receiver has M urns with N colors.

This reinforcement process is analysed in (19). Common interest allows a reduction of dimensionality, as before. The sender is reinforced for sending signal m in state s just in case the receiver is reinforced for guessing state s when presented with signal m. Again, for a state i and a signal j, consider the number ij of balls of color j in the sender's urn i. As in the 2×2 case, the dynamics of the normalised vector of sender-receiver connections $ij/\sum ij$ is studied, as stochastic approximation of a noncontinuous dynamics on the simplex.

The expected payoff can be shown to be a Lyapunov function for the mean field dynamics; convergence to the set of rest points is deduced, with a technically involved argument: this is required because of the discontinuities of the dynamics, and since not all Nash equilibria are rest points for this dynamics, contrary to what we have in the standard replicator dynamics.

The stability properties of the equilibria of this mean field dynamics can be linked to static equilibrium properties of the game, which are described by Pawlowitsch in (30): the zeros of the gradient of the payoff, the linearly stable equilibria as well as the asymptotically stable equilibria of the mean field dynamics correspond, respectively, to the Nash equilibria, neutrally stable strategies and evolutionarily stable strategies of the signaling game.

Finally, the following result can be stated in terms of a bipartite graph between states and signals, such that there is an edge between a state and a signal if and only if that signal is chosen infinitely often in that state. It is shown that any such graph with the following property, **P**, has a positive probability of being the limiting result of reinforcement learning:

> **P**: *(i) Every connected component contains a single state or a single signal and (ii) each vertex has an edge.*

In the event where property **P** holds, if there is an edge between a state and a signal, then the limiting probability of sending that signal in that state is positive.

When $M = N$, property **P** is exemplified by signaling systems, where each state is mapped with probability one to a unique signal. But even in this case it is also exemplified by configurations that contain both synonyms and information bottlenecks as in figure 2. Evolution of optimal signaling has positive probability, but so does evolution of this kind of suboptimal equilibrium. The case of $M = N = 2$ is very special. This corresponds closely to the replicator dynamics of signaling games, where partially pooling equilibria (which contain synonyms and bottlenecks) can be reached by the replicator dynamics.

Discussion

The results on the replicator dynamics suggest that for large populations the emergence of signaling systems in Lewis signaling games with perfect common interest between sender and receiver is guaranteed only under special circumstances. The dynamics also converges to states with imperfect information transfer. Introducing mutation can have the effect of making the emergence of perfect signaling more likely, though this statement should be taken with a grain of salt since the precise outcomes may depend on the mutation rates.

When interests are diametrically opposed in signaling games, there is no information transmission in equilibrium. But the equilibrium may never be reached. Instead senders and receivers may engage in a mad "Red Queen" chase, generating cycles or chaotic dynamics. In well-known costly signaling games, where interests are mixed, this "Red Queen" chase is a real possibility. Along the trajectories describing such a chase there are periods with significant information transfer from senders to receivers. Those interactions are

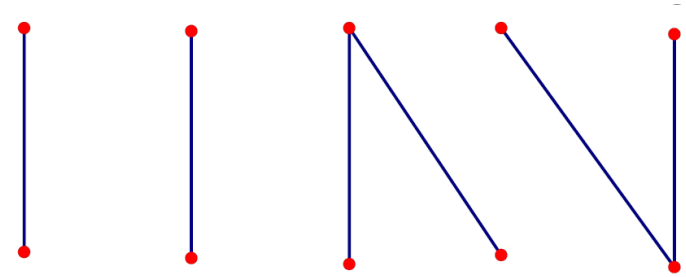

Fig. 2. Synonyms and information bottlenecks in a signing game with four states and four signals. The third states is represented by two signals (synonym), while the fourth signal represents two states (bottleneck). Under reinforcement learning such a configuration is reached asymptotically with positive probability.

undermined because of the underlying conflicts of interest, resulting in periods of low information transfer, from which a new period of higher information transfer can start.

Unlike large populations, a small population may spend most of its time efficiently signaling, even when the only Nash equilibrium does not support any information transfer. A small population engaged in "cheap talk" costless pre-play signaling may spend most of its time cooperating even when the only Nash equilibrium does not support cooperation.

A similar difference between small and large populations may be at work in costly signaling games. In the small population, small mutation limit costless signaling is possible even in games where the handicap principle claims that it should not be. In large populations this is not true, but there are alternatives to the costly signaling equilibria where signaling cost can be low while in equilibrium there is partial information transfer.

We encounter a similarly nuanced picture for models of individual learning. Herrnstein-Roth-Erev reinforcement learning leads to perfect signaling with probability one in Lewis signaling games only in the special case of 2 equiprobable states, 2 signals, 2 acts. In more general Lewis signaling games, the situation is much more complicated. To our knowledge, nothing is known about reinforcement learning for games with conflicts of interest and costly signaling games. This would be a fruitful area for future research.

We conclude that the explanatory significance of signaling equilibria depends on the underlying dynamics. Signaling games have multiple Nash equilibria. One might hope that natural dynamics always selects a Pareto Optimal Nash equilibrium, but this is not always so. On a closer examination of dynamics, in some cases, Nash equilibrium recedes in importance and other phenomena are to be expected.

ACKNOWLEDGMENTS. The work of S. H. was supported by the National Science Foundation under Grant No. EF 1038456. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. R. Argiento, R. Pemantle, B. Skyrms and S. Volkov. Learning to signal: Analysis of a micro-level reinforcement model. *Stochastic Processes and their Applications*, 119:373–90, 2009.
2. A. W. Beggs. On the convergence of reinforcement learning. *Journal of Economic Theory*, 122:1–36, 2005.
3. M. Bena‟m and M. W. Hirsch Asymptotic pseudotrajectories and chain recurrent flows, with applications. *Journal of Dynamics and Differential Equations*, 8(1):141–176, 1996.
4. M. Bena‟m Dynamics of stochastic approximation algorithms. *Sï£¡minaire de probabilitï£¡s*, XXXIII:1–68, 1999.
5. K. Binmore and L. Samuelson. Evolutionary drift and equilibrium selection. *Review of Economic Studies*, 66:363–394, 1999.
6. V. Crawford and J. Sobel. Strategic information transmission. *Econometrica*, 50:1431–1451, 1982.
7. R. Cressman. *Evolutionary Dynamics and Extensive Form Games*. MIT Press, Cambridge, MA, 2003.
8. I. Erev and A. E. Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88:848-81, 1998.
9. D. Fudenberg and L. Imhof. Imitation processes with small mutations. *Journal of Economic Theory*, 131:251–62, 2006.
10. A. Grafen. Biological signals as handicaps. *Journal of Theoretical Biology*, 144:517–546, 1990.
11. J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, New York, 1983.
12. C. Harley. Learning the evolutionary stable strategy. *J. Theor. Biol.* , 89:611–633, 1981.
13. R. J. Herrnstein. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 13:243–66, 1970.
14. J. Hofbauer. The selection mutation equation. *Journal of Mathematical Biology*, 23:41–53, 1985.
15. J. Hofbauer and S. M. Huttegger. Feasibility of communication in binary signaling games. *Journal of Theoretical Biology*, 254:843–849, 2008.
16. J. Hofbauer and S. M. Huttegger. Selection-mutation dynamics of lewis signaling games. Unpublished Manuscript, University of Vienna and UC Irvine, 2013.
17. J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, 1998.
18. E. Hopkins and M. Posch. Attainability of boundary points under reinforcement learning. *Games and Economic Behavior*, 53::110–25, 2005.
19. Y. Hu, B. Skyrms and P. Tarrẹs. Reinforcement learning in a signaling game. ArXiv, 2011.
20. S. M. Huttegger. Evolution and the Explanation of Meaning. *Philosophy of Science*, 74:1–27, 2007.
21. S. M. Huttegger, B. Skyrms, R. Smead, and K. Zollman. Evolutionary dynamics of lewis signaling games: Signaling systems vs. partial pooling. *Synthese*, 177:177–191, 2010.
22. S. M. Huttegger and K. J. S. Zollman. Robustness of hybrid equilibria in costly signaling games. Working paper UC Irvine, 2013.
23. S. M. Huttegger and Kevin J. S. Zollman. Dynamic stability and basins of attraction in the sir philip sidney game. *Proceedings of the Royal Society London, B*, 277:1915–1922, 2010.
24. J. F. Laslier, R. Topol and B. Walliser A behavioral learning process in games. *Games and Economic Behavior*, 37:340–366, 2001.
25. D. K. Lewis. *Convention. A Philosophical Study*. Harvard University Press, Harvard, 1969.
26. J. Maynard Smith. Honest signalling: the philip sidney game. *Animal Behavior*, 42:1034–1035, 1991.
27. M.A.. Nowak, A. Sasaaki, C. Taylor, D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428:246–650, 2004.
28. G. Nøldeke and L Samuelson. A dynamic model of equilibrium selection in signaling games. *Journal of Economic Theory*, 73:118–156, 1997.
29. C. Pawlowitsch. Finite populations choose and optimal language. *Journal of Theoretical Biology*, 249:606–616, 2007.
30. C. Pawlowitsch. Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior*, 63:203–226, 2008.
31. R. Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4:1–79, 2007.
32. H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400-7, 1951.
33. A. E. Roth and I. Erev. Learning in extensive form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8:164–212, 1995.
34. F. Santos, J. Pacheco and B. Skyrms. Co-evolution of pre-play signaling and cooperation. *Journal of Theoretical Biology*, 274:30–35, 2011.
35. W. A. Searcy and S. Nowicki. *The Evolution of Animal Communication*. Princeton University Press, Princeton, 2005.
36. B. Skyrms. *Signals: Evolution, Learning, and Information*. Oxford University Press, Oxford, 2010.
37. M. Spence. Job market signaling. *The Quarterly Journal of Economics*, 87:355–374, 1973.
38. A. Traulsen, J.C. Claussen, C. Hauert. Coevolutionary Dynamics: from finite to infinite populations. *Physical Review Letters*, 95:238701, 2005.

39. E. O. Wagner. The dynamics of costly signaling. *Games*, 4:161–183, 2013.
40. E. O. Wagner. Deterministic chaos and the evolution of meaning. *The British Journal for the Philosophy of Science*, 63:547–575, 2012.
41. E. O. Wagner. Semantic meaning in finite populations with conflicting interests. *The British Journal for the Philosophy of Science*, forthcoming.
42. K. Warneryd. Cheap talk, coordination and evolutionary stability. *Games and Economic Behavior*, 5:532–546, 1993.
43. J. Weibull. *Evolutionary Game Theory*. MIT Press, Cambridge, Mass., 1995.
44. P. Young. *Strategic learning and its limits*. Oxford University Press, 2005.
45. A. Zahavi. Mate selection – the selection of a handicap. *Journal of Theoretical Biology*, 53:205–214, 1975.
46. K. J. S. Zollman, C. T. Bergstrom, and S. M. Huttegger. Between cheap and costly signals: The evolution of partially honest communication. *Proceedings of the Royal Society London, B*, 280:20121878, 2013.