

Signaling games: dynamics of evolution and learning

Simon M. Huttegger

Konrad Lorenz Institute for Evolution and Cognition Research
Adolf Lorenz Gasse 2, A-3422 Altenberg, Austria

`simon.huttegger@kli.ac.at`

Kevin J. S. Zollman

Department of Philosophy, Carnegie Mellon University
Baker Hall 135, Pittsburgh, PA 15213-3890

`kzollman@andrew.cmu.edu`

June 11, 2009

1 Introduction

“Let us go down, and there confound their language, that they may not understand one another’s speech” (Genesis 11:1). The state of language confusion described in this passage may be understood as a state of maximal heterogeneity: every possible language is present in a population. It may also be viewed as a state of homogeneity, however; presumably, each possible language is spoken by a very small number of persons, inducing a uniform distribution over the set of languages. Should we expect individuals to stay at such a symmetric state? Or will they rather agree on one language, thereby breaking the symmetry of initial confusion (Skyrms, 1996)?

These questions are basic for the origin of language. When individuals cannot communicate to a sufficiently high degree, how can they decide on signaling conventions? In the philosophical literature, such problems were formulated by Quine (1936), although they have also been considered before Quine. Similar questions have also sparked interest in linguistics (Steels, 2001; Jäger and van Rooij, 2007) and in biology (e.g. communication at the microbiological level to animal signals).

One would be interested to know if coherent signaling evolves under simplified conditions. Perhaps the most simple model one can think about was introduced by David Lewis (see Lewis (1969)). By using some concepts from game theory, Lewis introduced signaling games as a simplified setting to study the emergence of language conventions.

On a larger scale, it should of course be emphasized that the evolution of language is an extremely complex issue where many more factors are involved than are captured in a signaling game. We think that studying very simplified models is nonetheless useful. Both experimental and theoretical approaches are confronted with the complexity of the problem of language evolution (Számádo and Szathmáry, 2006). This makes results from simple mathematical models particularly important. Such simple models sharpen our intuitions as to what might be important features to look for in more complex models. Models like that of signaling games provide a general framework for studying the emergence of communication; making signaling games more complex is a result of giving them more structure. Properties of signaling games will thus reappear at a more structured level. Moreover, simple mathematical models of signaling provide insights into specific processes that play an important role in language evolution. And, lastly, simple and tractable models allow us to identify key components of particular processes.

In this paper we report several results on the dynamics of Lewis signaling games. A dynamical view of signaling games is indispensable since we are interested in the process of the emergence of communication. We spend a considerable part of this paper on the evolutionary dynamics of signaling games as given by the replicator equations and a perturbation thereof. These two models should be viewed as a baseline case with which other studies should be compared. Accordingly, we shall be especially interested in finding differences between these two baseline cases and between them and more sophisticated dynamical models. These include structurally stable games, finite population models, and a number of models of learning in signaling games. We shall argue that the differences between all these models are such that the baseline models do not capture all possible dynamical behaviors. On the other hand, features like persistent and non-decreasing stochastic perturbations of evolutionary or learning dynamics appear to have qualitatively similar effects in a wide range of models. The interplay of various evolutionary and learning models that we describe in this paper may well prove useful in studying more complex models of language evolution or other evolutionary problems.

2 Lewis signaling games

In his book *Convention*, David Lewis describes a situation for the emergence of conventional signaling. One individual, the sender, has some private information about the world and has at her disposal a set of signals. Another individual, the receiver, observes the signal, but not the state, takes some action. Each state has an appropriate action, and both parties are interested in the receiver taking the appropriate action given the state. Because of the common interest, both parties are interested in coordinating on a convention to associate each state with a signal and each signal with the appropriate act. While there are many ways to specify this game, we will consider the easy circumstance where there are n states, n acts, and n signals.

We may thus represent sender strategies and receiver strategies by $n \times n$ matrices M having exactly one 1 in each row, the other entries being 0. If M is a sender matrix, then $m_{ij} = 1$ means that the sender chooses signal j after having observed state i ; if M is a receiver matrix, it means that the receiver chooses action j in response to signal i . Players may also choose what to do in response to the state or signal at random. We can expand the set of admissible matrices to those where for each entry $0 \leq m_{ij} \leq 1$, and each row sums to 1.¹

If a sender and receiver are matched to play the game the both gain a payoff of 1 if the receiver chooses the right action given the state and gain 0 otherwise. If each state is equally likely the sender (with strategy P) and receiver (with strategy Q) will expect to get the following payoff:

$$\pi(P, Q) = \frac{1}{n} \sum_{i,j} p_{ij} q_{ji}. \quad (1)$$

The payoff function given above can of course be modified. The states need not be weighed equally or the interests of the players may not coincide completely. Such modifications lead to interesting games, and we will discuss the first one briefly below.

It is easy to show that one-to-one strategies are of particular importance. A sender strategy P is one-to-one if no two states are mapped to the same signal, i.e. if the matrix P is a permutation matrix. Similarly, a receiver strategy is one-to-one if Q is a permutation matrix. A simple computation shows that if P is a permutation matrix and if Q is the transpose of P ($Q = P^T$ or $q_{ij} = p_{ji}$), then $\pi(P, Q) = 1$, which is the maximal payoff. Such strategy pairs (P, Q) were, for obvious reasons, termed *signaling systems* by Lewis.

Signaling systems can be viewed as simple languages. They are characterized by the property of yielding a maximum payoff to the players; i.e. no other strategy combination earns a payoff of 1. They are also the only strict Nash equilibria of signaling games. There is, however, a number of non-strict Nash equilibria which are part of Nash equilibrium components. If $n = 3$, one such Nash equilibrium component is given by

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \lambda & 1 - \lambda \end{pmatrix}, Q = \begin{pmatrix} \mu & 1 - \mu & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

where $0 \leq \lambda, \mu \leq 1$ (Trapa and Nowak, 2000). At (P, Q) , the players are always able to coordinate state 3 and act 3, but if state 1 or 2 occur they do not always achieve coordination. Here state 1 and 2 are “pooled” onto signal 1 and state 3 is communicated using two different signals (signals 2 and 3). As a result these equilibria are called *partial pooling equilibria*. There are also *total pooling* Nash equilibria. In these equilibria the sender sends the same signal regardless of state and the receiver takes the same action regardless of signal.

¹Representing randomization (or mixed strategies) in this way can either correspond to a single player intentionally randomizing, or alternatively to a population of players that don’t randomize but whose proportions are represented by the probabilities in the matrix.

Partial pooling equilibria like (P, Q) turn out to be particularly important for the emergence of communication in signaling games. Since signaling games have an uncountable number of Nash equilibria, the equilibrium selection problem becomes particularly pressing. Equilibrium refinement concepts like evolutionarily stable strategies and neutrally stable strategies exclude Nash equilibria which are not stable from an evolutionary perspective (Maynard Smith, 1982). In signaling games, signaling systems are the only evolutionarily stable strategies (Wärneryd, 1991). But Nash equilibria such as (P, Q) are neutrally stable. This means that natural selection will not move a population away from a signaling system. (P, Q) is also stable relative to natural selection, but drift may cause a population to move away from a neutrally stable state.

Analysis of signaling games in terms of other equilibrium concepts can also be given (Blume, 1994), but we think that an analysis from an evolutionary perspective is more revealing as to the problem of the emergence of communication. In this case, pinning down the evolutionarily and neutrally stable states does not get us very far. We are still confronted with a large number of possible evolutionary outcomes, and we do not know whether evolution leads to a state of communication. Moreover, concepts like that of an evolutionarily stable strategy appear to have no straightforward connection to models of learning in games. For these reasons, we think without a detailed analysis of various dynamic models our understanding of these games is incomplete at best and misleading at worst.

3 Evolutionary dynamics of signaling games

The basic model of evolutionary game theory is given by the replicator dynamics (Taylor and Jonker, 1978; Schuster and Sigmund, 1983; Hofbauer and Sigmund, 1998). We imagine a population of individuals partitioned into several types. Each type corresponds to a strategy of the underlying game. For signaling games, a type may be characterized by a sender part P and a receiver part Q if we would like to study the evolution of communication within one population. Another possibility consists in analyzing a two-population model, with one sender population and one receiver population. A type in the sender population will, in this case, correspond to a sender strategy P , and a type in the receiver population to a receiver strategy Q .²

The replicator dynamics relates the growth rate of each type of individual to its expected payoff with respect to the average payoff of the population: types with above-average performance increase in relative frequency, while types with below-average performance decrease. In a biological context, payoffs can be interpreted as fitnesses. Thus, we sometimes speak of fitness instead of payoff, or average fitness instead of average payoff.

If x_i is the frequency of type i , $x = (x_1, \dots, x_n)$ is the state of the population (being a probability vector) and $u(x_i, x)$ and $u(x, x)$ are the payoffs to type i and

²Here will restrict ourselves to pure strategies – those where every entry is either 0 or 1.

the average payoff in the population at state x , respectively, then the replicator dynamics is given by

$$\dot{x}_i = x_i(u(x_i, x) - u(x, x)). \quad (3)$$

\dot{x}_i denotes the time-derivative of x_i . Notice that equation (3) is one possibility to formalize the dependency of a type's growth rate to its performance relative to the population average, there are others collectively known as payoff-monotone dynamics. A system similar to (3) can be formulated for a two-population model (see Hofbauer and Sigmund (1998), Sections 10 and 11; in the context of signaling games, see Huttegger (2007b)).

If a population's initial condition is given by x , then (3) defines a unique orbit or solution curve $\phi(t)$ for $t \in \mathbb{R}$ with $\phi(0) = x$. ϕ describes the evolution of the population in the state space of relative frequencies.

If $\dot{x}_i = 0$ for all i , then x is called a rest point of (3). This means that whenever x is the initial condition of a population, it will stay at x for all future times. A rest point x is called Liapunov stable if for all neighborhoods U of x there exists a neighborhood V of x such that $\phi(t) \in U, t \geq 0$ whenever $\phi(0) \in V$. A rest point x is called unstable if it is not stable. A rest point x is asymptotically stable if it is Liapunov stable and if there exists a neighborhood U of x such that $\phi(t)$ converges to x as $t \rightarrow \infty$ whenever $\phi(0) \in U$. The same notions can be defined for a set of points S instead of a rest point x as well. Moreover, we will say that almost all points converge to some set of points S under (3) if the set of points that do not converge to S has Lebesgue measure zero in the state space of relative frequencies.

3.1 Replicator dynamics

Skyrms (1996) simulated the replicator dynamics of a binary Lewis signaling game, and Skyrms (2000) provides a mathematical analysis of a simplified binary Lewis signaling game, which does not include all 16 types (note that this is already a quite formidable number for a mathematical treatment of the dynamics). In simulations, population frequencies always converged to one of the two signaling systems. The same result was shown to hold analytically in the Lewis mini-game.

These results suggested the optimistic conjecture that in every Lewis signaling game almost all initial population states will converge to one of the signaling games under the dynamics (3). Huttegger (2007a) and Pawlowitsch (2008) have shown independently that this is in general not the case, Pawlowitsch by utilizing connections between neutral stability and the replicator equations, and Huttegger by using techniques from center-manifold theory (Carr, 1981). Let us take a closer look at the dynamical properties of signaling games.

Lewis signaling games have interior Nash equilibria. These equilibria represent states where all possible strategies are present. Huttegger (2007a) proves that these states are not stable for any signaling game. Indeed, interior equilibria are linearly unstable for the replicator dynamics (3). This implies that the set of points converging to an interior equilibrium has measure zero. Thus, for

almost all initial populations, symmetry gets broken in the minimal sense that not all signaling strategies will survive under evolutionary dynamics.

Signaling systems are strict Nash equilibria of Lewis signaling games; hence they are the only asymptotically stable states for the replicator dynamics (both for the two-population replicator dynamics and the one-population replicator dynamics of the symmetrized signaling game). At a signaling system s , the strategy pair corresponding to s has a relative frequency of 1. It follows that signaling systems are asymptotically stable for all signaling games.

Asymptotic stability is a local concept: it does not give us global information about the dynamical system. In particular, asymptotic stability does not imply global convergence to one of the signaling systems (global in the almost-all-sense). Indeed, for $n \geq 3$ it turns out that some of the continua of Nash equilibria that were described in Section 2 form sets that attract some nearby points.

Consider the connected set of Nash equilibria N given by (2). If we look at the dynamics (3) close to N we see that population frequencies sufficiently close to N converge to some point in N . When we look at the boundary of the set N , however, some of the Nash equilibria become dynamically unstable; i.e. there exist population frequencies arbitrarily close to such a Nash equilibrium that tend away from it.

This implies that the set N is not asymptotically stable. We cannot find a neighborhood U of N such that any point in U converges to N as time goes to ∞ . But each point x in the interior of N is Liapunov stable. Moreover—and this is the elephant in the kitchen—the interior of N attracts an open set of initial conditions. That is, the set of population frequencies converging to N has non-zero measure.

Components of Nash equilibria such as N exist for all signaling games with $n \geq 3$. This was shown by Huttegger (2007a) and by Pawlowitsch (2008). Pawlowitsch moreover links the existence of components like N to the concept of *neutrally stable strategies*, which was introduced by Maynard Smith (1982) as a generalization of *evolutionarily stable strategies*.

Suppose a whole population adopts a certain strategy s of some game. Then s is neutrally stable if s is a Nash equilibrium and if there exists no strategy s' that yields a higher payoff when played against itself than s yields when played against s' . Thus, neutral stability implies that a strategy is robust against invasion by selection (but it is not robust against drift).

Pawlowitsch (2008) finds an elegant characterization of neutrally stable strategies in Lewis signaling games: if P is a sender matrix and Q is a receiver matrix, then (P, Q) is neutrally stable if and only if (i) P or Q has no zero-column and (ii) neither P nor Q has a column with multiple maximal elements λ such that $0 < \lambda < 1$. Thus, a signal can represent more than one event, but then these events cannot be represented by any other signal. Similarly, an event can be linked to more than one signal; in this case, however, the signals cannot be linked to any other event.

In terms of the replicator dynamics (3), a neutrally stable strategy is a point in a component of strategies such as N ; i.e., if (P, Q) is neutrally stable

and is contained in a component of other neutrally stable strategies, then this component attracts an open set of population frequencies. Whether the reverse statement is also true is an open problem.

Signaling games with $n = 2$ are a special case. In such binary signaling games the existence of a component N that attracts an open set of population frequencies depends on the weights attached to the two events. If both weights are $\frac{1}{2}$, then no such component exists: almost all solution curves converge to one of the signaling systems. Once the weights are asymmetric, however, there exists a component N .

Thus we may conclude that for the replicator dynamics (3) signaling systems do not evolve generically. Numerical simulations show that the size of the basins of attraction of signaling systems is decreasing in n ; moreover, it is already non-negligible for $n = 3$ (Huttegger et al., 2008).

To understand the evolutionary dynamics of signaling games, a complete analysis of the replicator equations (3) is only a first step. The model of evolution as given by (3) can be extended and modified in various directions. Such explorations seem all the more necessary since the situation of having components of Nash equilibria is quite peculiar, as we shall explain now.

3.2 Selection-mutation dynamics

From the point of view of dynamical systems, the continua of rest points corresponding to these Nash equilibrium components are not *structurally stable* (see Guckenheimer and Holmes 1983 or Kuznetsov 2004).³ Structural stability refers to small perturbations of a system of differential equations like (3) (small relative to the functions constituting the differential equations and their partial derivatives). The system is structurally stable if such small perturbations do not change the qualitative properties of the solution trajectories. The solution trajectories of the original and the perturbed system are *topologically equivalent*. A system that is not structurally stable is called *degenerate*.

Systems with continua of rest points are always degenerate. This follows from the fact that continua of rest points are associated with zero-eigenvalues of the Jacobian matrix (the sign of the eigenvalues determines the qualitative nature of the solution trajectories near rest points). Perturbing the system will push zero-eigenvalues into the positive or the negative reals. This implies that the qualitative nature of the flow will change close to continua of rest points. Depending on the perturbation, the dynamics might change in many different ways. Thus, it is essential to choose a plausible perturbation of the dynamical system.

Hofbauer and Huttegger (2007, 2008) argue that the selection-mutation dynamics provides a plausible and (to some extent) tractable perturbation of the replicator equations (3) (for more information on this dynamics see Bürger 2000;

³Notice that continua of Nash equilibria are generic; i.e., if we perturb payoffs in a way that respects the extensive form of the game, Nash equilibrium components persist (cf. Cressman 2003 and Jäger 2008).

Hofbauer 1985; Hofbauer and Sigmund 1998; see also Huttegger et al. 2008). The selection-mutation dynamics is given by

$$\dot{x}_i = x_i(u(x_i, x) - u(x, x)) + \varepsilon(1 - mx_i), \quad (4)$$

where $\varepsilon > 0$ is a uniform mutation rate and $m = n^{2n}$ is the number of strategies for a signaling game with n signals. The first term on the right-hand side of (4) is the selection term, while the second term describes uniform mutation. The mutation term expresses the fact that a type might change into another type at each point in time, at a rate given by ε . If $\varepsilon = 0$, the selection-mutation dynamics coincides with the replicator dynamics.

Hofbauer and Huttegger (2007, 2008) do not study the selection-mutation dynamics (4) directly. They instead focus on the two-population selection-mutation dynamics with a sender population and a receiver population. This enhances the tractability of the model and can be justified by assuming that the roles of sender and receiver are independent. Our remarks below refer to the two-population selection-mutation dynamics.

There are two general results concerning the selection-mutation dynamics of signaling games. Both are statements about the location of rest points of the selection-mutation dynamics in comparison to the location of rest points of the replicator dynamics. First, all rest points of the selection-mutation dynamics are close to Nash equilibria of the signaling game. This rules out rest points that are close to rest points of the replicator dynamics which are not Nash equilibria (Hofbauer and Huttegger, 2008). Second, perturbed signaling systems exist, are unique and asymptotically stable. By a perturbed signaling system we mean a rest point of the selection-mutation dynamics close to a signaling system. Note that the proof of its uniqueness is necessary to define a perturbed signaling system properly. For details of the proof and additional remarks concerning rest points of the selection-mutation dynamics in general consult Hofbauer and Huttegger (2008).

Unfortunately, no general results are available for the existence and stability properties of possible rest points close to the attracting components of Nash equilibria that we described in the previous sections. Indeed, if N is such a component, then there are no general mathematical statements that would allow us to derive conclusions about the behavior of selection-mutation dynamics close to N .

Hofbauer and Huttegger (2007, 2008) analyze the behavior of the selection-mutation dynamics close to N with the help of Taylor expansions in terms of the mutation rates, index or degree theory (Hofbauer and Sigmund, 1998, Section 13.2), and Morse theory (Milnor, 1963). Their results do not give a clear-cut answer to the problem of the evolution of signaling systems. Whether perturbed signaling systems emerge depends the parameters involved, notably the ratio of the mutation rate of the sender population to the mutation rate of the receiver population and the probability distribution over the events.

If all events are equiprobable (the distribution has maximum entropy), then communication is most important (Nowak et al., 2002, Box 2). As the entropy

(the evenness) of the probability distribution decreases, communication becomes less important; always guessing the most probable event and ignoring signals is more attractive in this than in the equiprobable case. Hence, as the distribution becomes less even, the possibility of ending up in a state with suboptimal communication increases. If the receiver population's mutation rate is sufficiently lower than the sender population's mutation rate, then it also becomes more likely to end up in a suboptimal state under the selection-mutation dynamics. This can heuristically be explained by the receivers not being responsive enough to the experiments of the senders. For a precise mathematical formalization of these arguments see Hofbauer and Huttegger (2008).

It is important to notice that these results are specific for the perturbation (4), which is linear. Alternative perturbations could also include non-linear terms, which might create any finite number of perturbed rest points with all kinds of stability properties. Such alternative perturbations might, however, not have an equally clear empirical interpretation like the one given in (4).

3.3 Structurally stable signaling games

Jäger (2008) studies games which he calls *structurally stable*. Structural stability in Jäger's sense does not refer to perturbations of the dynamics, as in the previous subsection, but to perturbations in the payoffs of the players. In particular, he allows for the possibility of an uneven probability distribution for the set of events (like Nowak et al. 2002 and Hofbauer and Huttegger 2007) and requires that different signals incur differential costs.

These features lead to a perturbation of the players' payoffs, which does not destroy the existence of neutrally stable components, however. Jäger (2008) shows that the replicator dynamics still converges to neutrally stable components of Nash equilibria from an open set of initial conditions. Given this result, it seems necessary to approach the problem of degeneracy in signaling games (i.e. the existence of components of Nash equilibria) from dynamical systems theory, as we outlined in the previous paragraph.

3.4 Finite population models

An alternative way to deal with degeneracy in signaling games with techniques from dynamical systems is to consider finite population models. We shall mention this possibility only briefly, since it is the subject of Pawlowitsch's contribution to this volume.

Pawlowitsch (2007) studies signaling games under the frequency-dependent Moran process (cf. Nowak et al. 2004). Her results show that selection never favors a strategy replacing a signaling system, whereas it favors some strategy to replace any strategy other than a signaling system (including neutrally stable strategies). It is important to notice that the model of Pawlowitsch also employs a kind of perturbation (given by weak selection). As is argued in Huttegger et al. (2008), a Moran-process without any kind of perturbation does yield qualitatively the same results as the replicator dynamics.

Some models of finite populations also involve more population structure than is used in either the replicator dynamics or the Moran process. So called cellular automata models use grid structures where individuals are constrained to interact only with their neighbors. Zollman (2005) considers the 2-state/2-signal/2-act signaling game with equiprobable states. He finds that although every individual adopts a signaling system strategy, both type of signaling system strategies persist. On the grid regions form, where individuals are perfectly communicating with those in their region, but are failing with those outside. Without mutation these states are stable, and with mutation they only undergo small persistent changes in the location of the borders. Wagner (2009) extends this model to include 3-state/3-signal/3-act signaling games and also to population structures other than the grid. He finds that population structure of various sorts significantly reduces the probability that either partial or total pooling equilibria will evolve.

4 Learning models

Unlike population models that usually consider a large population of players playing a game against one another, models of individual learning usually consider two players playing against one another repeatedly. Each player chooses a play for each round by following a rule which uses the past plays and payoffs of the game. These models attempt to capture the process by which individuals come to settle on particular behaviors with one another.

The literature is replete with different models of individual learning. In analyzing a wide variety of different learning rules scholars are usually interested in one of three questions. First, how little cognitive ability is needed to learn a signaling system? In the replicator dynamic model we found that at least some of the time a biological process, like natural selection, can result in the emergence of language. Can other simple dynamic systems which are implemented at the individual level result in the same outcome? Second, is the replicator dynamics an appropriate approximation for models of individual learning? If individual learning results in similar outcomes, we have some reason to suppose the replicator dynamics offers a good approximation.⁴ Finally, scholars are interested in determining the relationships between features of the learning rules and their ultimate outcomes. Do all models that have limited memory converge to signaling systems? What about all those that remember the entire history?

With respect to the first question, it appears that very little cognitive ability is needed to result in signaling systems. In fact some very simple learning rules perform better than other more complex counterparts. This later fact also shows that no particular mathematical model (like the replicator dynamics) is likely to capture the range of possibilities presented in individual learning. This suggests that the study of learning in games represents an important avenue of research

⁴Since the replicator dynamics offers a sometimes mathematically simpler model than other learning rules having it represent an adequate approximation can reduce the amount of analysis substantially.

for those interested in the emergence of behavior in games. The last question, regarding the relationship between features of the learning rule and results, is complicated. We will postpone detailed discussion until the end of this section.

In the replicator dynamic models of signaling it is usually supposed that each individual is endowed with a contingency plan over all states or signals. In the one population model every individual had both receiver and sender contingency plans, while in the two population model individuals had only the relevant contingency plan (sender or receiver depending on their population). This model fits well with biological evolution, where individuals' responses are determined by a heritable biological mechanism. A similar model is less plausible in the case of learning. Suppose that state a occurs and a player sends signal x to a counterpart receiver who acts correctly – both receive a reward. It would be unrealistic to suppose that the reward received would influence the sender's propensity to send signal y in state b even though it did not occur. But this would often be the case if we modeled individuals as learning on entire strategies (full contingency plans for each state or signal). Instead, much of the learning literature restricts the learning to particular states or signals and models rewards as effecting only the behavior of the individual with regard to *that state or signal*.⁵

4.1 Minimal memory

We will begin our investigation by turning to the simplest learning rules, those that remember only the most recent round of play. One such learning rule, Win-stay/Lose-switch, was initially considered in a different context by Robbins (1952),⁶ and then later applied to in game theoretic situations by Nowak and Sigmund (1993). As its name suggests, players will remain with their most recent strategy when they “win” and switch to another strategy when they “lose.” For general game theoretic situations, much turns on what is classified as a win or loss, but since signaling games feature only two payoffs this need not concern us here.

Barrett and Zollman (2008) consider Win-stay/Lose-switch and similar Win-stay/Lose-randomize learning rules. They find that Win-stay/Lose-randomize will converge in the limit to perfect signaling both when learning is done on contingency plans and also when learning is done in individual states and signals. Interestingly such a result is not guaranteed for Win-stay/Lose-switch since the forced switch can make players miscoordinate forever.

These learning rules require only limited knowledge of the situation and require no sophisticated reasoning. We might imagine a slightly more cognitively complex learning rule where individuals are capable of counterfactual reason-

⁵It should not be presumed that a strategy learning model is totally implausible, however. For instance, if I am able to observe many plays of a the game before adopting a new strategy, I might be able to observe contingency plans. Similarly, if I recognize the situation as strategic, I may attempt to formulate reasonable contingency plans and adopt them.

⁶Robbins was considering a class of problems known now as bandit problems (cf. Berry and Fristedt, 1985).

ing, but still only consider the previous round. One such learning rule has an individual take the best response to the play of the opponent on the previous round. This requires more knowledge on the part of the player, since she must be capable of calculating what would have happened if she had acted differently.⁷ So-called “myopic best response” or “Cournot adjustment dynamics” has been considered extensively in the economics literature (cf. Fudenberg and Levine, 1998). In the case of 2-state/2-signal/2-act signaling games this learning rule has the same problem faced by Win-stay/Lose-switch, it can cycle forever. Beyond this fact, little is known about this learning rule and how it compares to the other short-memory learning rules.

It is not always appropriate to assume that individuals have only a one period memory. We will now turn to a learning rule which is at the other extreme – it remembers the entire history of play.

4.2 Indefinite memory

We will again return to considering learning rules which only consider their own actions and payoffs without engaging in counterfactual reasoning. So called Herrnstein reinforcement learning is one such learning rule. It was first introduced in the game theoretic literature by Roth and Erev (1995) and Erev and Roth (1998), but the underlying motivation traces to Herrnstein’s (1970) matching law – that the probability of an individual taking an action will be proportional to the sum of the rewards accrued from taking that action. Herrnstein’s matching law is instantiated by defining the probability of an action a using the following formula:

$$\frac{w_a}{\sum_x w_x} \tag{5}$$

w_a is the total rewards from taking action a and the sum in the denominator is the total rewards for taking all actions over past plays. This function for taking past successes and translating them into current propensities for action is known as the “linear response rule.”

As was done with the replicator dynamics, we will first consider the simplest case, two states, signals, and acts, with equiprobable acts. In this case, it has been proven that a separate sender and receiver both employing reinforcement learning on individual actions will converge (almost surely) to signaling systems (Argiento et al., 2007). Unfortunately, the proofs for this case are difficult and generalizations have not been forthcoming. Almost all that is known about other cases is the result of simulation studies.

Barrett (2006) found that for signaling games with more signals, states and acts will often converge to the partial pooling equilibria described above. As the number of states, signals, and acts grow, the proportion that converge to one form of partial pooling or another grow as well, reaching almost 60% for eight state, signal, act games. Barrett did find, however, that those systems always

⁷In signaling games, this learning rule would also require that the receiver be informed of the state after failure, so that she might calculate the best response.

achieved some success at information transfer. He observed no simulation that succeed less than half of the time, and a vast majority achieved relatively high success.⁸ Skyrms (2009) reports that failures similar to the replicator dynamics are observed when states are not equiprobable. In a two state, signal, act game with unequal state distributions total pooling equilibria are sometimes observed. Similar results are reported by Barrett (2006) regarding unequal state distributions for games with more signals, states, and acts.

The story here is interesting. In the replicator dynamics it appears that the introduction of random shocks is sufficient to avoid the pitfalls of partial and total pooling equilibria (at least in some cases). Herrnstein reinforcement learning has persistent randomness, but the magnitude of that randomness decreases over time. Simulation results suggest that this randomness is insufficient to mimic the randomness obtained by the selection-mutation dynamics and thus insufficient to avoid partial pooling equilibria.

Akin to Win-stay/Lose-switch, Herrnstein reinforcement does not use information about one's opponent's actions or about one's alternative responses to those actions. One might modify Herrnstein reinforcement learning to consider such a case, where an individual attempts to "learn" the, possibly mixed, strategy of one's opponent by observing past play.⁹ One assumes that the proportion of past plays represents an opponent's strategy and then takes the best response to that strategy. So called "fictitious play" has been applied in many settings in game theory (cf. Fudenberg and Levine, 1998), but it has not been studied extensively in signaling games. There are also other, more complex, Bayesian learning rules which have likewise not been applied to this game (cf. Young, 2004)

There have, however, been several other modifications to Herrnstein reinforcement that have been considered. They all retain the central idea that one's play is determined only by the rewards one has received in the past and not by strategic considerations like those used in myopic best reply or fictitious play.

4.3 Similar reinforcement models

There are many different ways to modify Herrnstein reinforcement in order to introduce larger persistent randomness. Only a few have actually been studied and there has not been anything close to an exhaustive search of the possibilities.

One might begin by modifying the way by which propensities are updated. It is usually assumed that the game being studied does not have negative payoffs so that propensities cannot become negative (and thus result in incoherent probabilities). Alternatively, one might allow for negative payoffs but truncate the propensities to remain above zero. Barrett (2006) investigates a collection of models where failure receives a payoff of less than zero and thus results in a "punishment" which decreases the probability of taking that action (rather

⁸For instance in a four state, signal, act game he found that, of those that failed, all approached a success rate of 3/4.

⁹The term "learn" may be a bit of a misnomer since, if one is playing against a opponent who is also using this learning rule, there is no stable strategy to learn.

than keeping it the same). Results of simulations involving different amounts of punishment suggests that this substantially decreases the basins of attraction of partial pooling equilibria and results in more efficient languages. Although this depends on the magnitude of the different rewards and punishments. Games with unequal state distributions have not been studied with this model.

In addition, Barrett (2006) considers a model where the propensities are subject to random shocks. Shocks are modeled as a number α which is drawn from some distribution with expectation of 1. On every round the propensities are multiplied by α resulting in random perturbations. Barrett finds that these shocks are sufficient to eliminate partial pooling equilibria in signaling games with more than two states, signals, and acts. Again, however, unequal state distributions have not been studied.

Rather than modifying the updating rules, one can also modify the response rule. Skyrms (2009) considers a model where the probabilities are determined by a logistic (or exponential) response rule:

$$\frac{e^{\lambda w_a}}{\sum_x e^{\lambda w_x}} \quad (6)$$

This exponential response rule alters the way that propensities are translated into probabilities over actions. The structure of this rule allows for small differences in propensities to have very little influence while larger differences have more significant influence. Skyrms (2009) finds that for reasonably small values of λ learners almost always learn to signal both for unequal state distributions and larger number of states signals and acts. This occurs largely because, when λ is small, initial play is more random and later play is more deterministic (than Herrnstein reinforcement) resulting in more early exploration.

4.4 More radical departures

The modifications considered so far preserved the underlying idea that weights are updated by addition (and potentially perturbed). Barrett and Zollman (2008) consider a model where the weights are updated by a weighted average instead of addition and propensities are calculated according to the exponential response in (6). They find that for particular parameter values individuals learn to optimally signal in games with three states, signals, and acts. This occurs largely because this learning rule approximates Win-stay/Lose-switch by continually exploring until it succeeds and then locks into the strategy that produces that success.

Barrett and Zollman (2008) also consider a yet more radical departure from Herrnstein reinforcement, the Adjustable Reference Point (ARP) learning model. ARP was first developed to explain human behavior in games by Bereby-Meyer and Erev (1998). We will avoid specifying the model here, but it is a reinforcement like model meant to capture four features absent in Herrnstein reinforcement: (1) what counts as success and failure can evolve based on past experience, (2) how one responds to “successes” and “failures” can differ, (3)

more distant rewards and punishments have less effect than more recent ones, and (4) rewards in one domain can have effects on other domains as well. Barrett and Zollman find that the ARP model significantly outperforms Herrnstein reinforcement in converging to near-optimal signaling systems.¹⁰ They attribute this success to the persistent randomness introduced by feature (3) – its ability to forget the past. Their conclusion is largely based on the apparent success of other learning rules discussed above which also discard past experience.

5 Conclusions

Overall it does appear that some successful communication can emerge out of initial confusion. Both models of evolution and of individual learning often result in the emergence of somewhat successful communication. Such success is not always guaranteed, however. In signaling games with more than two states, signals, and acts, perfect communication is not guaranteed to emerge. Similarly the emergence of perfect signaling is not certain in games where the states are not equiprobable. These conclusions hold both for evolution and learning models. However, we did find that signaling *can* emerge with very little cognitive sophistication. Communication can emerge from natural selection alone, or from some very simple learning rules like Win-stay/Lose-switch.

Several similarities between the models of learning and evolution are apparent. The results for the replicator dynamics coincided with the results for Herrnstein reinforcement learning. The relationship between these two models is more significant than the similarities mentioned here, and so this result is not entirely surprising (cf. Beggs, 2005; Hopkins and Posch, 2005). The selection-mutation dynamics (for appropriate parameter values) converges to perturbed signaling systems. This coincides with the results obtained for the ARP learning model. However, many of the other learning rules always converge to a (non-perturbed) signaling system – we have no version of the replicator dynamics which models this result.

Many of the learning rules which converged to signaling systems had an interesting feature: they began by exploring the space of possibilities, but then later began playing successful strategies with high probability. This feature is found in Win-stay/Lose-randomize and both reinforcement models with exponential response. Similarly, those that forget the past appeared to perform better than counterparts that did not, as was the case with ARP learning, Herrnstein reinforcement learning with random shocks, Smoothed reinforcement learning, and Win-stay/Lose-randomize.

These learning rules have large persistent randomness (at least early in the process). This feature is partially shared by the selection-mutation dynamics, which has persistent randomness throughout the process of evolution. The results from the extant literature on the evolution of communication suggests that

¹⁰Because there is persistent randomness in ARP learning it will not ever converge to any pure strategy.

this randomness is required in order for populations or individuals to converge on optimal signaling.

References

- Argiento, R., R. Pemantle, B. Skyrms, and S. Volkov (2007). Learning to signal: Analysis of a micro-level reinforcement model. *Manuscript*.
- Barrett, J. A. (2006). Numerical simulations of the Lewis signaling game: Learning strategies, pooling equilibria, and the evolution of grammar. Technical Report MBS 06-09, University of California, Irvine: Institute for Mathematical Behavioral Sciences.
- Barrett, J. A. and K. J. Zollman (2008). The role of forgetting in the evolution and learning of language. *Manuscript*.
- Bürger, R. (2000). *The Mathematical Theory of Selection, Recombination, and Mutation*. New York: John Wiley & Sons.
- Beggs, A. W. (2005). On the convergence of reinforcement learning. *Journal of Economic Theory* 122, 1–36.
- Bereby-Meyer, Y. and I. Erev (1998). On learning to become a successful loser: A comparison of alternative abstractions of learning processes in the loss domain. *Journal of Mathematical Psychology* 42, 266–286.
- Berry, D. A. and B. Fristedt (1985). *Bandit Problems: Sequential Allocation of Experiments*. London: Chapman and Hall.
- Blume, A. (1994). Equilibrium refinements in sender receiver games. *Journal of Economic Theory* 64, 66–77.
- Carr, J. (1981). *Applications of Centre Manifold Theory*. New York: Springer.
- Cressman, R. (2003). *Evolutionary Dynamics and Extensive Form Games*. Cambridge: MIT Press.
- Erev, I. and A. E. Roth (1998, September). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review* 88(4), 848–881.
- Fudenberg, D. and D. K. Levine (1998). *The Theory of Learning in Games*. Cambridge: MIT Press.
- Guckenheimer, J. and P. Holmes (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York: Springer.
- Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior* 15, 245–266.
- Hofbauer, J. (1985). The selection mutation equation. *Journal of Mathematical Biology* 23, 41–53.

- Hofbauer, J. and S. M. Huttegger (2007). Selection-mutation dynamics of signaling games with two signals. In A. Benz, C. Ebert, and R. van Rooij *Proceedings of the ESSLLI 2007 Workshop on Language, Games, and Evolution*.
- Hofbauer, J. and S. M. Huttegger (2008). Feasibility of communication in binary signaling games. *Manuscript, University of Vienna*.
- Hofbauer, J. and K. Sigmund (1998). *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.
- Hopkins, E. and M. Posch (2005). Attainability of boundary points under reinforcement learning. *Games and Economic Behavior* 53, 110–125.
- Huttegger, S. M. (2007a). Evolution and the explanation of meaning. *Philosophy of Science* 74, 1–27.
- Huttegger, S. M. (2007b). Evolutionary explanations of indicatives and imperatives. *Erkenntnis* 66, 409–436.
- Huttegger, S. M., B. Skyrms, R. Smead, and K. J. S. Zollman (2008). Evolutionary dynamics of Lewis signaling games: signaling systems vs. partial pooling. *Synthese, forthcoming*.
- Jäger, G. (2008). Evolutionary stability conditions for signaling games with costly signals. *Journal of Theoretical Biology, forthcoming*.
- Jäger, G. and R. van Rooij (2007). Language structure: biological and social constraints. *Synthese* 159, 99–130.
- Kuznetsov, Y. A. (2004). *Elements of Applied Bifurcation Theory*. New York: Springer.
- Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard: Harvard University Press.
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- Milnor, J. (1982). *Morse Theory*. Princeton: Princeton University Press.
- Nowak, M., N. L. Komarova, and P. Niyogi (2002). Computational and evolutionary aspects of language. *Nature* 364, 56–58.
- Nowak, M., A. Sasaaki, C. Taylor, and D. Fudenberg (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428, 646–650.
- Nowak, M. and K. Sigmund (2002). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner’s Dilemma game. *Nature* 417, 611–617.

- Pawlowitsch, C. (2007). Finite populations choose an efficient language. *Journal of Theoretical Biology* 249, 606–617.
- Pawlowitsch, C. (2008). Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior* 63, 203–226.
- Quine, W. V. (1936). Truth by Convention. In R. F. Gibson: *Quintessence. Basic Readings from the Philosophy of W. V. Quine*. Cambridge: Belknap Press 2004, 3–30.
- Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society* 58, 527–535.
- Roth, A. E. and I. Erev (1995). Learning in extensive-form games: Experimental data and simple dynamics models in the intermediate term. *Games and Economic Behavior* 8, 164–212.
- Schuster, P. and K. Sigmund (1983). Replicator Dynamics. *Journal of Theoretical Biology* 100, 533–538.
- Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge: Cambridge University Press.
- Skyrms, B. (2000). Stability and explanatory significance of some simple evolutionary models. *Philosophy of Science* 67, 94–113.
- Skyrms, B. (2009). *Signals: Evolution, Learning, and the Flow of Information (book manuscript)*.
- Steels, L. (2001). Grounding symbols through evolutionary language games. In: Cangelosi, A. and D. Parisi *Simulating the Evolution of Language*. London: Springer, 211–226.
- Számádo, S. and E. Szathmáry (2006). Selective scenarios for the emergence of natural language. *Trends in Ecology and Evolution* 21, 555–561.
- Taylor, P. D. and L. Jonker (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences* 40, 145–156.
- Trapa, P. E. and M. A. Nowak (2000). Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology* 41, 172–188.
- Wagner, E. (2009). Communication and structured correlation. *Forthcomming in Erkenntnis*
- Wärneryd, K. (1991). Evolutionary stability in unanimity games with cheap talk. *Economics Letters* 36, 375–378.
- Young, H. P. (2004). *Strategic Learning and its Limits* Oxford: Oxford University Press.
- Zollman, K. J. (2005). Talking to neighbors: The evolution of regional meaning. *Philosophy of Science* 72, 69–85.