Signals: Evolution, Learning, and Information

Brian Skyrms

https://doi.org/10.1093/acprof:oso/9780199580828.001.0001

Published: 08 April 2010 **Online ISBN:** 9780191722769 **Print ISBN:** 9780199580828

CHAPTER

8 8 Learning in Lewis Signaling Games

Brian Skyrms

https://doi.org/10.1093/acprof:oso/9780199580828.003.0009 Pages 93–105

Published: April 2010

Abstract

This chapter argues that we can and do learn to signal. We are not the only species able to do this, although others may not do it so well. The real question is what is required to be able to learn to signal. Or, better, what kind of learning is capable of spontaneously generating signaling? If the learning somehow has the signaling system preprogramed in, then learning to signal is not very interesting. If the learning mechanism is general purpose and low level, learning to signal is quite interesting.

Keywords: signals, signaling, learning

Subject: Philosophy of Science, Epistemology, Philosophy of Language

Collection: Oxford Scholarship Online

Can we learn to signal? Obviously we can and do. We are not the only species able to do this, although others may not do it so well. The real question is what is required to be able to learn to signal. Or, better, *what kind of learning is capable of spontaneously generating signaling?* If the learning somehow has the signaling system preprogramed in, then learning to signal is not very interesting. If the learning mechanism is general purpose and low level, learning to signal is quite interesting. In Chapter 1, we saw that for one kind of signaling game, low level reinforcement learning could learn to signal. If many kinds of low level learning allow the spontaneous emergence of signaling in many situations, we are on the way to a robust explanation.

Roth–Erev reinforcement

We return to two-state, two-signal, and two-act games with states equiprobable, and put in all possible strategies. There are now an infinite number of pooling equilibria, as well as the signaling systems. We would
p. 94 most like an analysis of this case where reinforcement operates not on whole strategies, but rather on ↳ individual acts. Then agents would not even need to see the situations they find themselves in as part of a single game.

Suppose that the sender has a separate set of inclination weights—of accumulated past reinforcements—for each state of the world. You can think of each state as coming equipped with its own urn, with balls of different colors for different signals to send. The receiver has a separate set of accumulated reinforcements for each signal. You can think of the receiver as having a different urn for each signal received, with balls of different colors for different acts to choose.

Spontaneous emergence of signaling in this more challenging set-up would be fully consonant with the spirit of Democritus, "who sets the world at chance."[1] It requires no strategic reasoning, just chance and reinforcement. This is, in fact, just what happens. Individuals *always* learn to signal in the long run. This is not only confirmed by extensive simulations, it is also a theorem.[2] In this situation individuals converge to a signaling system with probability one, with the two possible signaling systems being equally likely. Spontaneous emergence of signaling is virtually guaranteed.

That is limiting behavior, but what of the short run? Figure 8.1 shows the results of simulations starting with initial weights all equal to 1. Learning is fast. On average, after 100 trials individuals have an 80% success rate.
p. 95 After 300 trials they are right 90% of the time. ↳

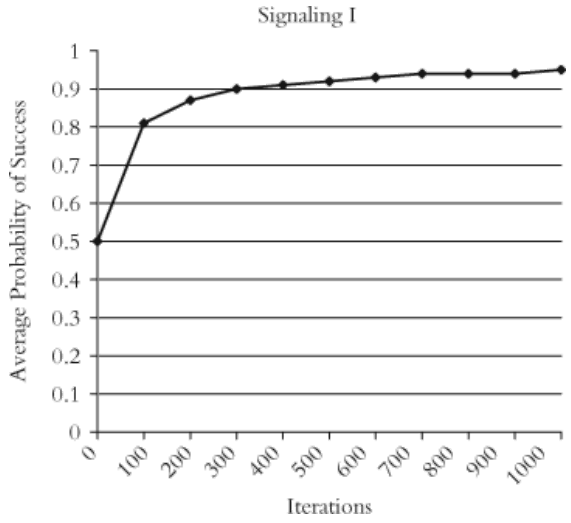

Figure 8.1: Learning to signal with 2 states, 2 signals, 2 acts with states equiprobable. Initial weights =1. Reinforcements for success=1.

Harder cases

Does the bad news about the replicator dynamics carry over as well as the good news? Does reinforcement learning sometimes learn partial-pooling (with only partial information transfer) in Lewis games with three states, three signals, and three acts? And does it sometimes end up in total pooling (with no information transfer) where there are only two states, signals and acts, and the states have unequal probabilities?

A full analytic treatment of these questions is not available. But they can be investigated by simulation. We will concentrate on reinforcing acts. There is only one parameter of the reinforcement, the initial weights with which we start the process. For the purpose of initial simulations, we start each player with an *initial weight of one* for each possible choice, players choose with probability proportional to their weights, and we augment
p. 96 weights by *adding a payoff of one* for a success. In Lewis signaling games with three equiprobable ↳ states, three signals and three acts, reinforcement learning learns to signal in a little more than 90% of trials, but lands on partial pooling in the rest. As the number of states, signals and acts increases the success rate goes

down. If the number is 4, simulations hit signaling a little less than 80%; if the number is 8, perfect signaling emerges less than half the time.[3]

And even in the basic game where the number of states, signals and acts is 2, unequal probability of the states can sometimes lead to signals that contain no information at all. How often depends on the magnitude of the inequality. When one state has probability .6, suboptimal outcomes hardly ever happen, at probability .7 they happen 5% of the time. This number rises to 22% for probability .8, and 44% for probability .9.[4] Suboptimal equilibria are still there.

Roth and Erev found their learning relatively insensitive to initial choice of weights, but they were considering a different class of games. So we should try varying the weight parameter. We set the probabilities of states quite unequal, at 90%–10% and run reinforcement dynamics with initial weights of different orders of magnitude. The probability of ending up in pooling equilibrium instead of a signaling system is shown in the figure 8.2.

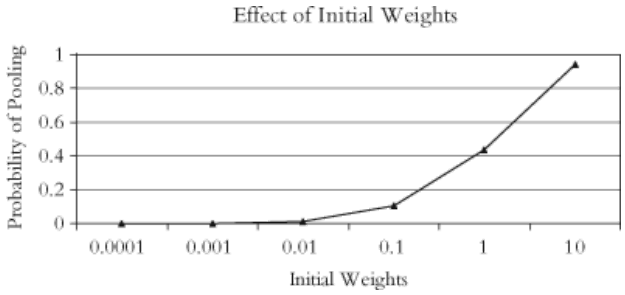

Figure 8.2: Effect of initial weights where state probabilities are 90%–10%.

Initial weights make an enormous difference! If we raise them to 10, then the probability of getting trapped in a pooling equilibrium goes up to 94%. If we lower them to .01 probability of pooling goes down to 1%. And at the minuscule initial weights of .0001, we saw no pooling at all; each trial led to a signaling system.[5] The one innocuous parameter of Roth–Erev learning becomes crucial. Small initial weights also lead to signaling in larger Lewis signaling games.

How are they performing their magic? The explanation cannot come near the end of the learning process.
p. 97 There the initial ↳ weights, whether great or small, have been swamped by reinforcement. Rather, small initial weights must have their impact at the beginning of the learning process, where they make the initial probabilities easy to modify. Perhaps the explanation is that they both facilitate initial exploration and enhance sensitivity to success.

Bush–Mosteller reinforcement

In the simplest Lewis signaling game with equiprobable states, it was proved that Roth–Erev learners would learn to signal with probability one. In the proof, it is crucial that Roth–Erev learners do not learn too fast or too slowly. They are neither too hot nor too cold. This is no longer true for reinforcement learners who learn according to the basic dynamics of Bush and Mosteller. The basic Bush–Mosteller learning dynamics is too cold. Sometimes it freezes into suboptimal states.[6] This is not to say, however, that Bush–Mosteller learners never learn to signal. To get an indication of how often they learn successfully, and how fast, we turn to simulations.

p. 98 The surprising result is that, despite the theoretical possibilities for unhappy outcomes, Bush–Mosteller learners are very successful indeed. The only parameter of the learning dynamics is the learning rate, which is

between zero and one. In our basic signaling game, for a wide range of learning rates between .05 and .5, individuals learned to signal in at least 99.9% of the trials. These results are from running simulations out to 10,000 iterations of the learning process. For the short run, consider just 300 iterations of learning. With the learning parameter at .1, then in 95% of the trials individuals had already learned to signal with a success rate of more than 98%. Learning to signal is no longer guaranteed, but it is still to be strongly expected.

What of the more problematic cases, in which states have unequal probabilities?

Here, variations in the learning parameter can make a big difference, just as variations in the magnitudes of the initial weights did in Roth–Erev reinforcement. For comparison with figure 8.2, we reconsider the case in which the state probabilities are 90%–10%. using Bush–Mosteller.[7]

With a high enough learning parameter, we reliably learn to signal even with highly unequal state probabilities.
p. 99 If we concentrate on the short and medium run, the situation with ↳ Bush–Mosteller reinforcement doesn't look much different from that of Roth–Erev reinforcement.

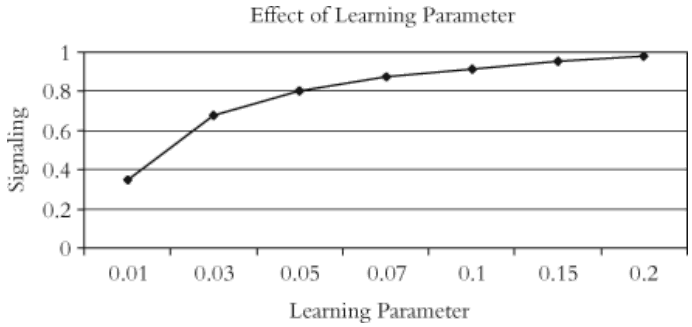

Figure 8.3: Effect of learning parameter with state probabilities 90%–10%.

Exponential response

Consider Roth–Erev modified by using an exponential response rule. Choice probabilities are no longer simply proportional to weights, but rather to:

$$\mathrm{Exp}[\lambda * \mathrm{weight}].$$

The constant λ controls the noise in the response probabilities. When λ=0, noise washes out all other considerations, and all possible choices are equally probable. When λ is high, the rule almost always picks the alternative with the highest weight.

The exponential response rule interacts with cumulative reinforcements of propensities in an interesting way. As propensities grow, the learner moves more and more towards a sure pick of the alternative with the highest propensity. If we start with a very small λ, we start with lots of random exploration that gradually moves towards deterministic choice.[8]

In two-state Lewis signaling games with unequal state probabilities (probability of state one at .6, .7, .8, .9), simulations of this learning model with small λ (.0001 to .01) always converge to signaling systems.[9] Similar
p. 100 results are gotten for three-state Lewis games.[10] The range of values is not implausible for human learning. ↳ This form of reinforcement learning allows individuals to avoid suboptimal equilibria and to arrive at efficient information transfer.

More Complex Reinforcement

There are all sorts of refinements and variations of the foregoing models. Yoella Bereby-Meyer and Ido Erev[11] compare models and conclude that a modification of Roth–Erev learning, which they call the ARP model, best fits the data. This model incorporates negative payoffs, which result in balls being taken out of the urn, and a floating reference point, which determines which payoffs are positive or negative. Payoffs above the reference point are positive, those below are negative, and the reference point itself adjusts in response to past experience. In good times your reference floats up, in bad times it settles down. What you are used to eventually tends to become your reference point. Negative payoffs are subtracted from weights until they are almost equal to zero.

The point where we stop subtracting is called the truncation point. There is a little discounting of the past. There are errors. All of these modifications have psychological currency. This is a model with a lot of parameters. Erev and Bereby-Meyer fit the parameters to the data.

Jeffrey Barrett has taken the ARP model, together with the parameter values gotten from the data by Erev and Bereby-Meyer, and shown how this type of learning allows one to learn to signal.[12] Barrett finds that the basic modifications introduced into the model for psychological reasons tend to make it easier to learn to signal.[13] I
p. 101 believe that at this point we can conclude that the ↳ possibility of learning to signal by simple reinforcement is a reasonably robust finding.

Neural Nets

Patrick Grim, Paul St. Denis, and Trina Kokalis[14] consider spontaneous emergence of signaling in neural nets. There is a spatial array of agents, each equipped with a neural net. Both food sources and predators migrate through space. There are different optimal actions—feed or hide—in the presence of a food source or a predator. Individuals can utter potential signals that are received (taken as inputs) by their neighbors.

Periodically, individuals have their neural nets "trained up" by their most successful neighbors. Simulations show spontaneous emergence of successful signaling in which individuals "warn" of predators in the neighborhood and "advertise" wandering food sources.

Imitating neighbors

Kevin Zollman also investigates learning to signal by interaction with neighbors on a spatial grid, using imitation dynamics.[15] Each individual looks at eight neighbors, to the N, NE, E, SE, S, SW, W, NW, and imitates the most successful neighbor if that neighbor does better than she does. Ties are broken at random. He considers two games. The first is a Lewis signaling game with two states, acts, and signals. Signaling evolves. In 10,000 simulations, starting with a random assignment of strategies, signaling systems always emerged.
p. 102 However, alternative signaling systems coexisted, each occupying ↳ different areas. We see spontaneous generation of regional signaling dialects.

His second game is even more interesting. Signaling is possible prior to playing another game—the Stag Hunt—with neighbors. Play in the Stag Hunt can be conditional on the signal received. A strategy now consists of a signal to send and an act in the Stag Hunt for each possible signal received. Just as before, signals have no preexisting meaning. Meaning now must co-evolve with behavior in the Stag Hunt.

In the Stag Hunt game, each player has two possible acts: Hunt Stag, Hunt Hare. Payoffs in one canonical Stag Hunt are:

	Stag	Hare
Stag	4,4	0,3
Hare	3,0	3,3

There are two equilibria, one in which both players hunt Stag and one in which they both hunt Hare. The former is better for both players, but each runs a risk by hunting Stag. If the other hunts Hare, the Stag hunter gets nothing. Hare hunters run no such risk. For this reason, conventional evolutionary dynamics favors the Hare hunting equilibrium.

Zollman finds that with interactions with neighbors on a spatial grid and imitate-the-best learning, pre-play signaling evolves such that all players end up hunting Stag. This happens even though the signaling systems are not all the same. We end up with a heterogeneous population that has spontaneously learned both to signal and to use those signals to cooperate. (Grim, Kokalis, Tafti, and Kilb had already used imitation dynamics on a spatial grid in their signaling game with food sources and predators.[16])

The foregoing papers are confined to interactions with neighbors on a special kind of structure—a spatial grid
p. 103 with edges wrapped to ↳ form a torus. Elliott Wagner extends the analysis to arbitrary interaction networks.[17] He also considers not only the nice case of two states, signals and acts, states equiprobable, but also our problem cases with unequal probabilities and bigger games.

He finds that the network structure is very important for whether individuals learn to signal, and for whether they learn the same signaling systems or evolve regional dialects. Small world networks are highly conducive to arriving at uniform signaling across the population, and they are remarkably effective in promoting efficient signaling even in our problem cases.

Belief learning

Let us move up to the simplest form of belief learning, and see what difference it makes. We now assume that the agents involved know the payoff structure of the game, but do not directly observe what the other player did. What happens with best response dynamics? In general, players *may not know* what a best response is. They know what they did, they know whether or not they got a payoff, and they know the structure of the game. So if they did get a payoff signaling worked, and the best response to the other player's last act is to do the same thing in the same situation. In the special case where each player only has two choices, if they did not get a payoff the best response is clearly to try the other thing.

But in the general case, where there are more than two states, acts, and consequences, plays which lead to no payoff leave the players somewhat in the dark. The receiver knows what she did, and what signal she got, but not which of the states the sender saw. The sender knows what the state was and which signal was sent, but not which of the inappropriate acts was done. Neither knows enough to determine the best response. In such a
p. 104 case, we might consider a weak version of the rule in which she chooses at random ↳ between alternatives which might be a best response consistent which what she knows. Call this *best response for all we know*.

The special case of two states, signals and acts is different. If signaling does not succeed, each player can figure out what the other did since the other had only two choices. Best response dynamics here is well defined. How does it do? The following analysis is due to Kevin Zollman.

Pure *best response* dynamics can get trapped in cycles and never learn to signal.[18] Our primitive belief learners are outsmarting each other! Let us try making them a little less eager to be rational. Every once and a while a player best-responds to the other's previous action, but most of the time he just keeps doing the same thing mindlessly. This is *best response with inertia*. You can think of each as flipping her own coin to decide whether to best respond on this round or not. (The coins can be biased.) Acting in accord with best response with inertia, our agents now always learn to signal. With probability one, they sooner or later hit on a signaling system, and then stick with it forever.

What about signaling games with N signals, states and acts? Now the closest our players can come to best response is *best response for all we know*. On getting a payoff, they know that they did a best response to the other's act, so they stick with it. On a failure all they know is that they didn't do a best response, but they don't know which of the other possible actions was the best response—so they choose at random between those alternatives. Already in the case where N=2, that we have already considered, this kind of learning gets trapped in cycles. So, for the general case, we are led to consider *best response for all we know with inertia*. Individuals either just keep doing what they were doing, or—at random times—best respond for all they know. This is an exceedingly modest form of belief learning, but Zollman shows that here (numbers of states, signals and acts
p. 105 are equal), it always learns to signal. It locks on to ↳ successful pieces of a signaling system when it finds them, and it explores enough to surely find them all.

Now let us think about where all this thinking about belief learning has led. *Best response for all you know with inertia* just comes to this: *Keep doing what you have been doing except once in a while pay attention and if you fail try something different* (at random). So redescribed, this learning rule does not require beliefs or strategic thinking at all! The cognitive resources required are even more modest than those required for reinforcement learning, since one need not keep track of accumulated payoffs—and it always works.

Learning to signal

How hard is it to learn to signal? This depends on our criterion of success for the learning rule. If success means spontaneous generation of signaling in many situations, then all the kinds of learning that we have surveyed pass the test. In particular, all forms of reinforcement learning work, although some work better than others. If it means learning to signal with probability one in all Lewis signaling games, a simple payoff-based learning rule will do the trick. It is easy to learn to signal.

Notes

1 As Dante has him in the Divine Comedy, Canto IV:

> Then when a little more I rais'd my brow,
> I spied the master of the sapient throng,
> Seated amid the philosophic train.
> Him (*Aristotle*) all admire, all pay him rev'rence due.
> There Socrates and Plato both I mark'd,
> Nearest to him in rank; Democritus,
> Who sets the world at chance, Diogenes,
> With Heraclitus, and Empedocles,
> And Anaxagoras, and Thales sage,
> Zeno, and Dioscorides well read
> In nature's secret lore.

I would put Democritus higher.

2 Argiento et al. 2009.

3 Barrett 2006.

4 These simulations are for 1000 trials, with 100,000 iterations of reinforcement learning on each trial.

5 At state one probabilities of .8, .7, and .6 we always get signaling for initial propensities .0001, and .001.

6 Hopkins and Posch 2005; Borgers and Sarin 1997; Izquierdo et al. 2007.

7 1,000 trials, 100,000 iterations per trial.

8 For example, suppose the choice is between two acts, A and B and that A is reinforced three times for every two times B is reinforced. Let lambda in the exponential response rule be .001, the propensity for A be 3n, and the propensity for B, 2n. Then for n=10 the probability to choose A would be .5025 — just a little more than one half. But for n=100 this probability would be .5250; for n=1000, .7311; for n=10,000, .999955. Since the ratio of the responses has been kept constant at three to two in this example, the linear response rule would have kept probability of A at 2/3.

9 This is not true for larger values of λ.

10 The mechanism is somewhat different from that in win-stay, lose-randomize.

11 Bereby-Meyer and Erev 1998.

12 Barrett 2006, 2007a, 2007b. Barrett and Zollman 2007.

13 There are other competing complex models with their own parameters to estimate from the data. It would be nice to have a definitive realistic model to apply to signaling. At this time there seems to be no clear winner. The models all fit the data reasonably well. Salmon 2001; Feltovich 2000; Erev and Haruvy 2005.

14 Grim et al. 2002.

15 Zollman 2005.

16 Grim et al. 2000; See also the review in Grim et al. 2004.

17 Wagner 2009.

18 Try working out the possibilities yourself.