Dyn Games Appl (2012) 2:129–145
DOI 10.1007/s13235-011-0027-2

# Inventing New Signals

**Jason McKenzie Alexander · Brian Skyrms · Sandy L. Zabell**

Published online: 20 August 2011


**Abstract** A model for inventing new signals is introduced in the context of sender–receiver games with reinforcement learning. If the invention parameter is set to zero, it reduces to basic Roth–Erev learning applied to acts rather than strategies, as in Argiento et al. (Stoch. Process. Appl. 119:373–390, 2009). If every act is uniformly reinforced in every state it reduces to the Chinese Restaurant Process—also known as the Hoppe–Pólya urn—applied to each act. The dynamics can move players from one signaling game to another during the learning process. Invention helps agents avoid pooling and partial pooling equilibria.

**Keywords** Signals · Invention · Reinforcement · Chinese Restaurant Process · Hoppe–Pólya urn

## 1 Introduction

Sender-receiver signaling games were introduced by Lewis [11] and in more general form by Crawford and Sobel [5]. Nature picks a state of the world, with some fixed probability, from a set of states. One player, the sender, observes the state and picks a signal from some arbitrary set of signals. (Signals are arbitrary in the sense that they are not assumed to have preexisting meaning or salience.) A receiver observes the signal and chooses one of a set of acts. Payoffs are jointly determined by the state of the world and the act taken. It is

J. McKenzie Alexander
London School of Economics and Political Science, London, UK

B. Skyrms (✉)
University of California Irvine, Irvine, CA, USA
e-mail: bskyrms@uci.edu

B. Skyrms
Stanford University, Stanford, CA, USA

S.L. Zabell
Northwestern University, Evanston, IL, USA

interesting to investigate whether some form of adaptive dynamics of evolution or learning can spontaneously generate meaningful signaling.

Recent investigations have demonstrated unexpected complexity in the dynamics of very simple signaling games with strong common interest. Suppose that there are the same number of states, signals and acts. Suppose that for each state there is a unique act, such that the payoff for both sender and receiver is 1 if the act is chosen, and 0 otherwise. (In this case one says there is pure or strong common interest.) A *sender strategy* is a map from states to signals, a *receiver strategy* a map from signals to acts. If we give the act that pays off in a state the same index as the state, we can define a *signaling system equilibrium* as a pair of sender and receiver strategies whose composition maps the $i$th state to the $i$th act for each index $i$. This is obviously the most desirable situation to be in.

At the opposite extreme, there are also *complete pooling equilibria,* in which the sender sends the signals with probability independent of the state, the receiver does acts with probability independent of the signal received, and consequently the probability of an act being taken does not depend on the state Nature has chosen. Provided there are more than two states, there are also *partial pooling equilibria* in which some but not all states are pooled. (This means that the states and their corresponding acts can be divided into subsets, such that within a subset one has complete pooling: the probability of doing acts in the subset does not depend on states in the subset, and for each state in the subset the sum of the probabilities of the acts in the subset sum to one.)

In these special games, signaling system equilibria are distinguished from an evolutionary point of view by being the unique evolutionarily stable states [18]. It might then seem plausible that replicator dynamics [17] would always lead to a signaling system equilibrium, but this turns out not to be true [9, 12]. In the special case of 2 states, 2 signals and 2 acts with pure common interest and where nature chooses states with equal probability, it is true. But if states are not equiprobable, the connected component of *pooling* equilibria has a basin of attraction of positive measure. And with 3 states, 3 signals, and 3 acts, *partial pooling* equilibria have a basin of attraction of positive measure even if nature chooses states with equal probability.

With reinforcement learning, the situation is more complicated. Argiento et al. [2] consider the basic reinforcement learning scheme of Roth and Erev [15], Erev and Roth [7], applied to 2 state, 2 signal, 2 act signaling games with pure common interest and equiprobable states. It generates a stochastic process as follows. The sender has an urn for each state, each containing two colors (say yellow and blue). When the sender observes a state, she draws a ball from the urn for that state. If she draws a yellow ball she sends signal one; if she draws a blue ball she sends signal two. The receiver has an urn for each signal, each containing two colors (say red and green). Upon receiving a signal, the receiver draws a ball from the urn corresponding to the signal. If the ball is red he does act one; if green, act two. In the case of both sender and receiver, the original balls drawn are then returned. In addition, if the right act for the state is done both sender and receiver also add an extra ball of the same color drawn to the selected urns. This is then repeated. (The initial number of yellow and blue balls in the sender's urns, and the initial number of red and green balls in the receiver's urns can be arbitrary.) Using stochastic approximation theory, Argiento et al. [2, Theorem 1.1] prove that with probability one the players converge to a signaling system; if the initial distribution of colors is uniform, each signaling system has equal probability of being selected. Simulation studies suggest this result does not carry over to the case where states are not equiprobable, because players may sometimes converge to pooling.

Pooling is inefficient and undesirable. Why can't the agents simply *invent new signals* to remedy the situation? We would like to have a simple, easily-studied model of such a process. That is to say, we want to move beyond closed models where the theorist fixes the signals, to an open model in which the space of signals itself can evolve. We would like to suggest such an open model here. This involves a kind of hybrid of the Roth–Erev urn process and the Chinese restaurant process—the latter being also known in another guise as the Hoppe–Pólya urn.

## 2 The Chinese Restaurant Process and the Hoppe–Pólya Urn

Imagine a Chinese restaurant, with an infinite number of tables, each of which can hold an infinite number of guests. People enter one at a time and sit at either an occupied table or an unoccupied one. Imagine there is also a ghost—a phantom guest—who is always sitting at one, otherwise unoccupied, table. The probability that a guest sits down at a table is proportional to the number already at that table, including the phantom guest. (So that if $n$ guests have been seated, the probability of the next guest joining the phantom is $1/(n+1)$.)

The first person to enter sits at the first unoccupied table, since no one but the phantom guest is there. The phantom guest then moves to an unoccupied table. The second person to enter now has equal probability of either sitting with the first, or sitting at the table with the phantom, resulting in a new occupied table. Should the second person join the first, the third person entering the room has a $2/3$ chance of sitting at their table and a $1/3$ chance of starting a new occupied one. Should the second person start a new table, the phantom guest moves on, and the third person will join one or the other, or start a third occupied table, all with equal probability. This is the *Chinese Restaurant Process* which has been studied as a problem in abstract probability theory [1, 13]. It is equivalent to a simple urn model, and this urn scheme can be modified to represent reinforcement learning with invention.

In 1984 Hoppe introduced what he called "Pólya-like Urns" in connection with "neutral" evolution—evolution in the absence of selection pressure. In the classical Pólya urn process, we start with an urn containing various colored balls. Then we proceed as follows: A ball is drawn at random and then returned to the urn with another ball of the same color. All colors are treated in exactly the same way. We can recognize the Pólya urn process as a special case of reinforcement learning in which there is no distinction worth learning—there are no states, no acts, and reinforcement always occurs. It is a standard result that the probabilities in a Pólya urn process (the fraction of each color present) almost surely converge to some random limit. (That is, they are guaranteed to converge to something, but that something can be anything.)

To the Pólya urn, Hoppe [10] adds a black ball—the *mutator*. The mutator brings new colors into the game. If the black ball is drawn, it is returned to the urn and a ball of an entirely new color is added to the urn. (Hoppe allowed that there might be more than one black ball, corresponding to multiple phantom guests in the Chinese restaurant process. Here, however, we will stick to the simplest case.) The Hoppe–Pólya urn model was meant as a model for neutral evolution, where there are a vast number of potential mutations which convey no selective advantage. (The same urn model has an alternative life in the Bayesian theory of induction, having essentially been invented in 1838 by the logician Augustus de Morgan to deal with the prediction of the emergence of novel categories. It generalizes Bayes–Laplace rules of succession, known to philosophers as Carnap's continuum of inductive methods [20, 21].)

It is evident that the Hoppe–Pólya urn process and the Chinese Restaurant process are essentially the same process described in two different ways. Hoppe's colors correspond to the tables in the Chinese Restaurant; the mutator ball to the phantom guest. After a finite number of iterations $N$, the $N$ guests in the restaurant or the $N$ balls in Hoppe's urn (other than the phantom guest or the black ball) are partitioned into some number of categories. The categories are colors for the urn, tables for the restaurant. But the partitions we end up with can differ each time; they depend on the luck of the draw. We have *random partitions of the guests or balls*, each time having a different number of categories, different numbers of individuals in each category, and different individuals filling out the numbers.

The possible patterns that can arise after $n$ guests are seated are described by the *partition vector* $\langle a_1, \ldots, a_n \rangle$ where $a_j$ denotes the number of tables with $j$ guests. For example, if there are four guests sitting at two tables, the pattern of one table with one guest and one table with three guests means that $a_1 = 1$ (the one table with one guest), $a_2 = 0$ (there are no tables with two guests), $a_3 = 1$ (the one table with three guests), and $a_4 = 0$ (there are no tables with four guests); this corresponds to the partition vector $\langle 1, 0, 1, 0 \rangle$.

Note that there are four ways of realizing this pattern and that all four are equally likely. If the two tables are labeled A and B, and the four guests arriving in order are labeled 1, 2, 3, 4, then the four patterns and their probabilities are (the first guest always sits at the first table):

| Table A | Table B | Probability |
|---|---|---|
| 1, 2, 3 | 4 | $1 \times 1/2 \times 2/3 \times 1/4 = 1/12$ |
| 1, 2, 4 | 3 | $1 \times 1/2 \times 1/3 \times 2/4 = 1/12$ |
| 1, 3, 4 | 2 | $1 \times 1/2 \times 1/3 \times 2/4 = 1/12$ |
| 1 | 2, 3, 4 | $1 \times 1/2 \times 1/3 \times 2/4 = 1/12$ |

It is in fact generally true of the process that all realizations of a given partition vector are equally likely to occur. All that affects the probability is a specification of the number of tables that have a given number of guests. The fact that any arrangement of guests with the same partition vector has the same probability is called *partition exchangeability*, and it is the key to mathematical analysis of the process.

There are explicit formulas to calculate probabilities and expectations of classes of outcomes after a finite number of iterations. The expected number of categories—of colors of ball in Hoppe's urn or the expected number of tables in the Chinese restaurant—will be of particular interest to us, because the number of colors in a sender's urn will correspond to number of signals in use. This is given by a very simple formula: after $N$ iterations, the expected number is $\sum_{(i=1 \text{ to } N)} 1/i$, the $N$th partial sum of the harmonic series, which grows logarithmically in $N$. Results are plotted in Fig. 1.

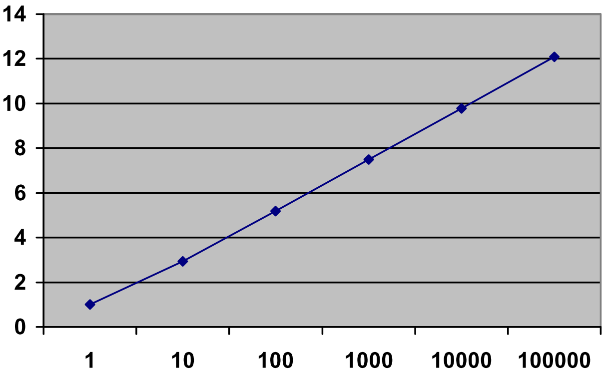


**Fig. 1** Expected number of categories

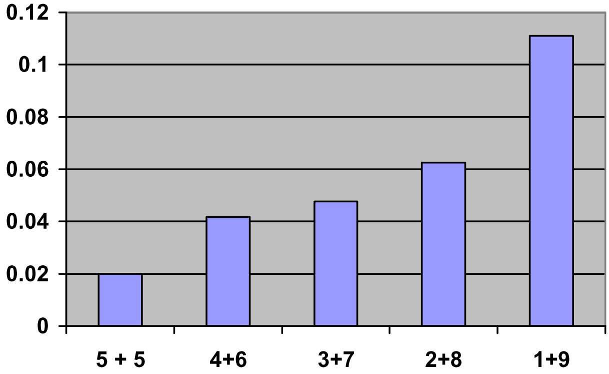


**Fig. 2** Probability of partitions of 10 into two categories

Although it is known that with probability one the limiting number of categories is infinite, for even quite large numbers of iterations the expected number of categories is relatively modest.

There is something else that we would like to emphasize. For a given number of categories, the distribution among those categories is not uniform. We can illustrate this with a simple example. Suppose we have ten iterations and the number of categories turns out to be two (two colors of ball, two tables in the restaurant), something that happens about 28% of the time. This can be realized in five different ways as a partition of 10: $5+5$, $4+6$, $3+7$, $2+8$, $1+9$. (These correspond to the partition vectors: $a_5=2$ and $a_j=0$ otherwise, $a_4=a_6=1$ and $a_j=0$ otherwise, $a_3=a_7=1$ and $a_j=0$ otherwise, $a_2=a_8=1$ and $a_j=0$ otherwise, and $a_1=a_9=1$ and $a_j=0$ otherwise, respectively.) There is a simple way of calculating the probability of each—the *Ewens sampling formula*—which gives the probability of a partition vector for $n$ draws:

$$\Pr\langle a_1,\dots,a_n\rangle = \prod_{j=1 \text{ to } n} 1/\big[\big(j^{a_j}\big)(a_j!)\big]$$

The results in the case of our example are graphed in Fig. 2.

The more unequal a division is between the categories, the more likely it is to occur. Some colors are numerous, some are rare. Some tables are much fuller than others. This can be seen as the result of a kind of *preferential attachment* process. In the Chinese restaurant, fuller tables are more likely to attract new guests. This generates a power-law distribution, similar to those that are ubiquitous in word frequencies in natural language and elsewhere [22].

## 3 Reinforcement with Invention

We remarked that the Pólya urn process can be thought of as reinforcement learning when there is no distinction worth learning—all choices (colors) are reinforced equally. The Hoppe–Pólya urn, then, is a model which adds useless invention to useless learning. That was its original motivation, where different alleles confer no selective advantage.

If we modify the Pólya urn by adding differential reinforcement—where choices are reinforced according to different payoffs—we get the basic Roth–Erev model of reinforcement learning, used by Argiento et al. If we modify the Hoppe–Pólya model by adding differential reinforcement, then—as discussed below—we get reinforcement learning capable of invention.

## 4 Inventing New Signals

We use the Hoppe–Pólya urn as a basis for a model of inventing new signals in signaling games. For each state of the world, the sender has an additional choice: *send a new signal.* A new signal is always available. The sender can either send one of the existing signals or send a new one. Receivers always pay attention to a new signal. (A new signal means new signal that is noticed, failures being taken account of by making the probability of a successful new signal smaller than one.) Receivers, when confronted with a new signal, just act at random. We equip them with equal initial propensities for the acts.

Now we specify how learning proceeds. Nature chooses a state and the sender either chooses a new signal, or one of the old signals. If there is no new signal the model works just as with basic Roth–Erev reinforcement. If a new signal is introduced, it either leads to a successful action or not. When there is no success, the system returns to the state it was in before the experiment with a new signal was tried.

But if the new signal leads to a successful action, both sender and receiver are reinforced. The reinforcement now consists of the sender's increased propensity to send the signal in the state in which it was just sent, and the receiver keeping track of the successful acts when receiving the new signal. In terms of the urn model, in the case of a successful act, the receiver activates an urn for the signal, with one ball for each act, and a second ball for the successful act. The sender now considers the new signal not only in the state in which it was tried out, but also considers it a possibility in other states. So, in terms of the urn model, a ball for the new signal is added to each sender's urn, as well as a reinforcement ball added to the urn for the state that just occurred. The new signal has now established itself. We have moved from a Lewis signaling game with $N$ signals to one with $N + 1$ signals. See Fig. 3.

In summary, one of three things can happen:

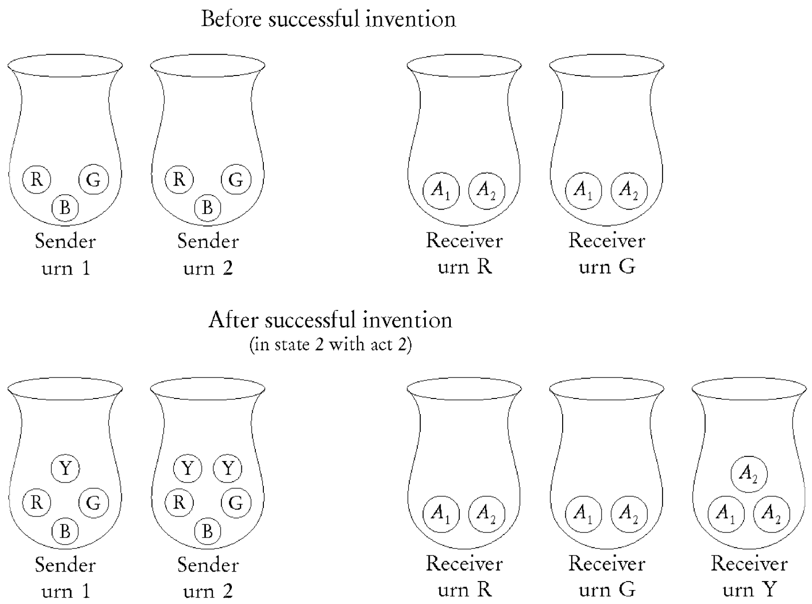


**Fig. 3** R denotes a red ball, G—green, B—black, Y—yellow. In State 2 a black ball is drawn, act 2 is tried and is successful. A yellow ball is added to both sender's urns and a reinforcement yellow ball is added to the urn for state 2. The receiver adds an urn for the signal yellow, and adds an extra ball to that urn for act 2

1. No new signal tried, and the game is unchanged. Reinforcement proceeds as in a game with a fixed number of signals.
2. A new signal is tried but without success, and the game is unchanged.
3. A new signal is tried with success, and the game changes from one with $n$ states, $m$ signals and $p$ acts to one with $n$ states, $m + 1$ signals, $p$ acts.

## 5 Starting with Nothing

If we can invent new signals, we can start with no signals at all, and see how the process evolves. We can expect that—like the simple Hoppe–Pólya urn—the limiting number of different signals will be infinite. The appendix gives a proof for the case of $m$ states having unequal probability and $n$ acts. Starting with no signals, the limiting number of different types of signals is almost surely infinite, and each signal is almost surely sent an infinite number of times.

But if we run a large finite number of iterations, we do not expect a comparably large number of signals. For learning with invention, we do not have anything like the stochastic approximation theory analysis of Argiento et al. to prove this rigorously, and such an analysis looks very hard to come by. We therefore proceed with a preliminary investigation by simulation.

Consider the 3 state, 3 act Lewis signaling game with states equally probable. As before, we have strong common interest—exactly one act is right for each state. In simulations of our model of invention, starting with no signals at all, the number of signals after 100,000 iterations ranged from 5 to 25. A histogram of the final number of signals in 1,000 trials is shown in Fig. 4. This behavior is close to that which would be expected from a pure Chinese Restaurant process.

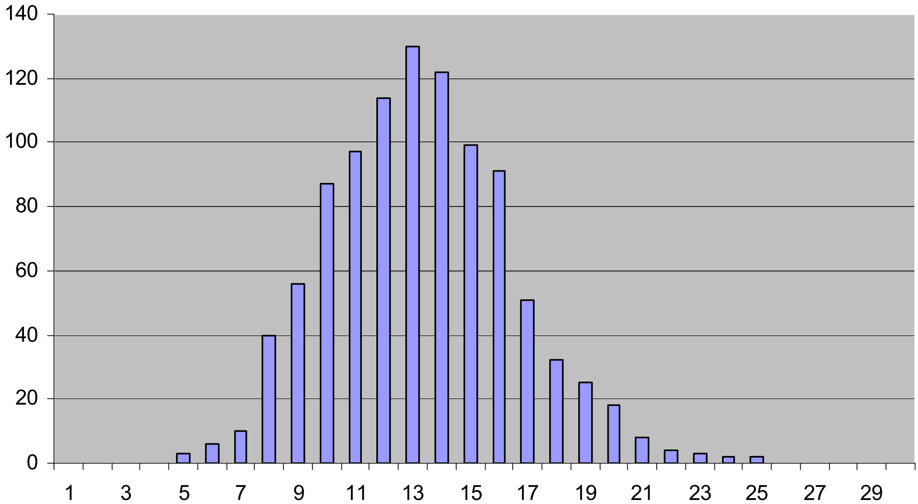


**Fig. 4** Number of signals after 100,000 iterations of reinforcement with invention. Frequency in 1000 trials

## 6 Avoiding Pooling Traps

In a version of this game with the number of signals fixed at 3, reinforcement learning sometimes falls into a partial pooling equilibrium. In simulations of basic Roth–Erev reinforcement learning with initial propensities of 1, 9.6% of the trials led to imperfect information transmission [[3], 1000 trials, 1,000,000 iterations of learning per trial]. In these cases the average payoff approaches $2/3$ and the players appear to approach a partial pooling equilibrium. (In the remaining trials the average payoff was close to 1.) Using reinforcement with invention, starting with no signals, 1,000 trials *all* ended up with efficient signaling (using a success rate of at least 99% after one million trials as a proxy for efficient signaling). Signalers went beyond inventing the three requisite signals. Lots of synonyms were created. By inventing more signals, they avoided the traps of partial pooling equilibria.

In the game with 2 states, 2 acts, and the number of signals fixed at 2, if the states had unequal probabilities agents sometimes fell into a complete pooling equilibrium—in which no information at all is transmitted and the average payoff is $1/2$. In such an equilibrium the receiver would simply do the act suited for the most probable state and ignore the signal, and the sender would send signals with probabilities that were insensitive to the state.

The probability of falling into complete pooling increased as the disparity in probabilities became greater. Our simulations of basic Roth–Erev reinforcement learning, 1,000 trials, 100,000 iteration of learning per trial, give us the following picture. When one state has probability 0.6, failure of information transfer hardly ever happens. At probability 0.7 it happens 5% of the time. This number rises to 22% for probability 0.8, and 38% for probability 0.9. Highly unequal state probabilities appear to be a major obstacle to the evolution of efficient signaling.

If we take an extreme case in which one state has probability 0.9, start with no signals at all, and let the players invent signals as above, then they reliably learn to signal. In 1000 trials they never fell into a pooling trap; they always learned a signaling system (again using a success rate of at least 99% after one million trials as a proxy for efficient signaling). The invention of new signals makes efficient signaling a much more robust phenomenon.

## 7 Causes of Efficiency

The ability to avoid partial pooling traps and evolve efficient signaling might be due to two different mechanisms, alone or in combination. It might be the case that an excess of signals might, by itself, make partial pooling much less likely. If this is the case, then starting with a fixed number of signals larger than the number of states (invention disabled) would make it less likely that individuals would get near partial pooling. Or individuals falling into a partial pooling equilibrium might be able to invent themselves out of it. If this were the case, then we should find that if we start the process with invention near partial pooling, new signals enable the evolution to efficient signaling. In fact, both mechanisms can operate.

Consider the case of two states, signals and acts, in which state 1 has probability 0.9, and state 2 probability 0.1. We will initialize the system near a pooling equilibrium, and run learning with invention. The degree of entrenchment of the pooling equilibrium can be varied by changing the initial weights: with more balls in the sender's and receiver's urn, it becomes more difficult to wander away from pooling.

We start the sender with $(1/2n) + 1$ balls for signal 1 and $(1/2n) + 1$ balls for signal 2 in the urn for each state; the receiver with $n + 1$ balls for act 1 and 1 ball for act 2 in her urn for each signal. The higher the entrenchment parameter, $n$, the more difficult it is for a new

**Table 1** The effect of extra signals on efficient signaling

| Number of initial signals | % of pooling | % of signaling |
|---|---|---|
| 2 | 38.1 | 61.9 |
| 3 | 12.0 | 88.0 |
| 4 | 4.5 | 95.5 |
| 5 | 1.7 | 98.3 |
| 6 | 0.5 | 99.5 |
| 7 | 0.5 | 99.5 |
| 8 | 0.2 | 99.8 |
| 9 | 0 | 100.0 |

signal to become established. For $n = 10$, 100, 1,000, 10,000 learning with invention always converged to a signaling system equilibrium (1,000 trials, 1,000,000 iterations per trial). For comparison, with no invention learning always converged to the pooling equilibrium with $n > 100$. (The results were similar if the system was initialized near the pooling equilibrium where signal 1 was always sent.)

Even in simulations with 1,000,000 learning steps, invention produces only a modest number of signals. So we might just start the process with extra signals already there. We now keep the initial weights as in the original setup and vary the initial number of signals. Extra signals promote learning to signal efficiently, as shown in Table 1.

It appears that both (i) excess initial signals make it less likely to fall into a pooling equilibrium and (ii) invention of new signals allows escape from the vicinity of a pooling equilibrium.

## 8 Synonyms

Let us return to signaling with invention. Typically we get efficient signaling with lots of synonyms. How much work are the synonyms doing? Consider the following simulation of $3 \times 3 \times 3$ signaling, starting with no signals and proceeding with 100,000 iterations of learning with invention (see Table 2).

Notice that a few of the signals (shown in boldface) are doing most of the work. In state 1, signal 5 is sent 87% of the time. Signals 1 and 2 function as significant synonyms for state 2,

**Table 2** In invention, only a few synonyms are used to achieve efficiency

| | | |
|---|---|---|
| Signal 1 | Probabilities in states 0, 1, 2 | 0.000, **0.716**, 0.000 |
| Signal 2 | Probabilities in states 0, 1, 2 | 0.000, **0.281**, 0.000 |
| Signal 3 | Probabilities in states 0, 1, 2 | 0.096, 0.000, 0.001 |
| Signal 4 | Probabilities in states 0, 1, 2 | 0.009, 0.000, 0.000 |
| Signal 5 | Probabilities in states 0, 1, 2 | **0.868**, 0.000, 0.000 |
| Signal 6 | Probabilities in states 0, 1, 2 | 0.000, 0.000, **0.810** |
| Signal 7 | Probabilities in states 0, 1, 2 | 0.024, 0.000, 0.000 |
| Signal 8 | Probabilities in states 0, 1, 2 | 0.000, 0.000, **0.143** |
| Signal 9 | Probabilities in states 0, 1, 2 | 0.000, 0.000, 0.044 |
| Signal 10 | Probabilities in states 0, 1, 2 | 0.000, 0.000, 0.000 |
| Signal 11 | Probabilities in states 0, 1, 2 | 0.000, 0.000, 0.000 |
| Signal 12 | Probabilities in states 0, 1, 2 | 0.001, 0.000, 0.000 |
| Signal 13 | Probabilities in states 0, 1, 2 | 0.000, 0.000, 0.000 |

being sent more that 99.5% of the time. Signals 6 and 8 are the major synonyms for state 3. (All of these signals are highly reinforced on the receiver side.) The pattern is fairly typical (in 1000 trials). Very often, many of the signals that have been invented end up little used.

This is just what we should expect from what we know about the Hoppe–Pólya urn. Even without any selective advantage, the distribution of reinforcements among categories tends to highly unequal, as was illustrated in Fig. 2. Might not infrequently used signals simply fall out of use entirely? Simulations that suggest they do not fade away, but at present we do not have an analytic proof.

## 9 Noisy Forgetting

Nature forgets things by having individuals die. Some strategies (phenotypes) simply go extinct. This cannot really happen in the replicator dynamics—an idealization where unsuccessful types get increasingly rare but never actually vanish. And it cannot happen in Roth–Erev reinforcement where unsuccessful acts are dealt with in much the same way.

Evolution in a finite population is different. In the models of Sebastian Shreiber [16], a finite population of different phenotypes is modeled as an urn of balls of different colors. Successful reproduction of a phenotype corresponds to addition of balls of the same color. So far this is identical to the basic model of reinforcement learning. But in Schreiber's models individuals also die. We transpose this idea to learning dynamics to get a model of reinforcement learning with noisy forgetting.

For individual learning, this model may be more realistic than the usual model of geometrical discounting. (Geometrical discounting was by Roth and Erev as a modification of the basic model to incorporate forgetting or "recency" effects.) That model may be best suited for aggregate learning, where individual fluctuations are averaged out. But individual learning is noisy, and it is worth looking at an urn model of individual reinforcement with noisy forgetting.

## 10 Inventing and Forgetting Signals

We can put together these ideas to get learning with invention and with noisy forgetting, and apply it to signaling. It is just like the model of inventing new signals except for the random dying-out of old reinforcement, implemented by random removal of balls from the sender's urns.

The idea may be implemented in various ways. Here is one. With some probability, nature picks an urn at random and removes a colored ball at random. (The probability is the forgetting rate, and we can vary it to see what happens.) Call this *Forgetting A*. Here is another. Nature picks an urn at random, picks a color represented in that urn at random, and removes a ball of that color. Call this *Forgetting B*.

Now it is possible that the number of balls of one color, or even balls of all colors could hit zero in a sender's urn. Should we allow this to happen, as long as the color (the signal) is represented in other urns for other states? Here is another choice to be made. But in the simulations we are about to report, we never hit a zero.

We simulated *invention* with *No Forgetting, Forgetting A, Forgetting B,* starting with no signals, for the number of states (= acts) being 3, 4, 5. States are equiprobable and the forgetting rate is 0.3 for all simulations. A run consists of 1,000,000 iterations of the learning process, and each entry is the average of 1,000 runs. The results are shown in Table 3.

**Table 3** Using forgetting to prune unneeded signals

| | Average number of signals remaining | | |
|---|---|---|---|
| | No Forgetting | Forgetting A | Forgetting B |
| 3 states | 16.276 | 19.879 | 3.016 |
| 4 states | 17.491 | 21.079 | 4.005 |
| 5 states | 18.752 | 22.686 | 4.982 |
| 6 states | 20.097 | 24.069 | 5.975 |
| 7 states | 21.336 | 25.82 | 6.96 |
| 8 states | 22.661 | 27.14 | 7.941 |
| 9 states | 23.815 | 28.684 | 8.929 |
| 10 states | 24.925 | 30.663 | 9.928 |

Relative to *No Forgetting*, *Forgetting B* is highly effective in pruning unused signals. (However, as noted before, even with no forgetting the number of excess signals after a million trials is not that large.) In contrast, *Forgetting A* does not help at all and even appears to be detrimental. This difference is real: even after a run of only 1,000 iterations of the learning process and 100 runs for both *No Forgetting* and *Forgetting A*, the difference in the average of the number of signals ($10.90 - 9.18 = 1.72$) is highly statistically significant. (Using the Welch—that is, unpooled—test, $t = 4.4579$ on 197 degrees of freedom, with a $p$-value of $1.388 \times 10^{-5}$; the corresponding 95% confidence interval for the difference is 0.96 to 2.48. The difference remains highly significant even using a nonparametric procedure: the Wilcoxon rank sum test gives an even smaller $p$-value of $5.061 \times 10^{-6}$.)

Why is there a difference between *No Forgetting* and *Forgetting A*? A plot of the average number of signals for the two cases as one goes from 1 to 1,000 iterations, as shown in Fig. 5, is suggestive.

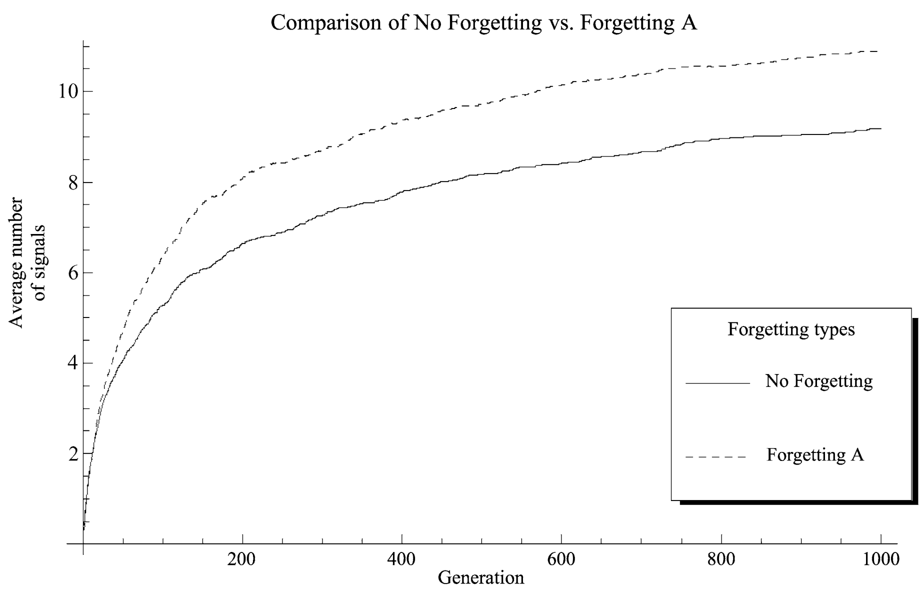


**Fig. 5** *No Forgetting* vs. *Forgetting A*, 100 runs of 1,000 generations, states equiprobable

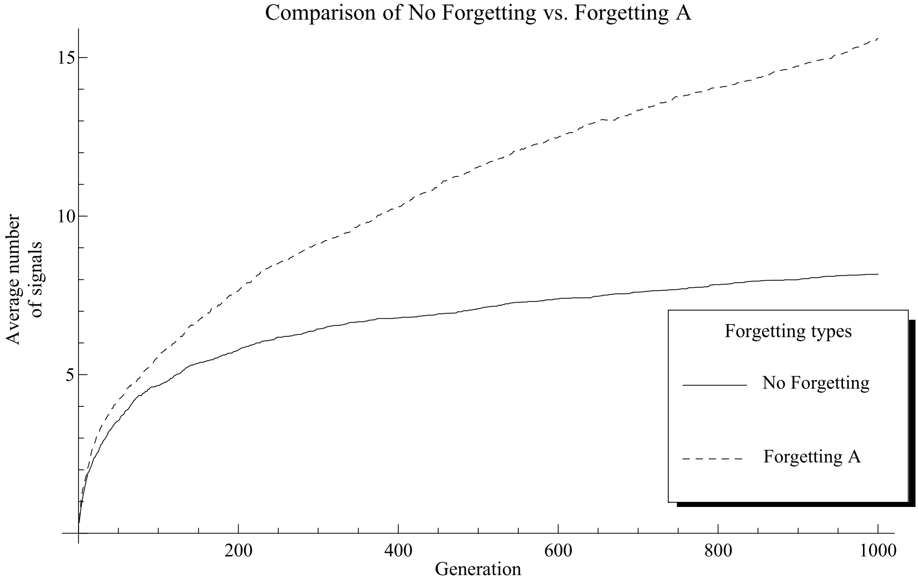


**Fig. 6** *No Forgetting* vs. *Forgetting A*, 100 runs of 1,000 generations. Initial state probabilities selected at random from a uniform distribution over the simplex. The same initial conditions were used for the $n$th run of each type of forgetting

This behavior does not appear to depend on any special properties of the equiprobable state case just discussed: if one does 1,000 iterations of *No Forgetting* and *Forgetting A* for 100 randomly chosen state probabilities instead, a similar graph arises, as shown in Fig. 6.

We hypothesize that one important reason for why *No Forgetting* does as well as it does is that it is relatively easy for a successful signal to lock in, and the number of the colored balls that represent it to grow fast. With forgetting A, the more balls there are of a given color, the more likely it is that one of that color will be thrown away. This tends to prevent lock-in, make reinforcement of a successful color less likely, and hence increase the variety of colored balls in the urn.

## 11 Related Work: Infinite Numbers of States or Acts

Bhaskar [4] analyzes an evolutionary model with an infinite number of signals. There is noisy, cost-free pre-play communication added to a base game. Because of the infinite number of signals, there must be signals with arbitrarily low probability of use. Then, because of special properties of the model, there are always signals that are as good as new in functioning as a "secret handshake" [14]. The "secret handshake" destabilizes inefficient equilibria. This suggests investigation of learning with inventing new signals in the setting of cost-free pre-play communication.

Worden and Levin [19] analyze a model in which there are an infinite number of potential acts. Trying an unused act changes the game. It is assumed that players can only try an unused act that is "close" to a used act in that its payoff consequences are epsilon-close to those of one of the old acts. The game can then change slowly so that a Prisoner's Dilemma can evolve into a non-dilemma. Formally, we could extend our model to include inventing new acts by giving the receiver a Hoppe–Pólya urn. The question is how to extend the payoff matrix. Worden and Levin supply one sort of answer. Others might be of interest, depending on the context of application.

## 12 Conclusion

We move from signaling games with a closed set of signals to more realistic models in which the number of signals can change. New signals can be invented, so the number of signals can grow. Little used signals may be forgotten, so the number of signals can shrink. A full dynamical analysis of these models is not available, but simulations suggest that these open models are more conducive to the evolution of efficient signaling than previous closed models.

## Appendix: Infinite Number of Signals

This appendix recapitulates some definitions in more mathematical form, and provides proofs of some statements.

### A.1 Signaling Equilibria

Suppose there are $s$ states, $s$ acts, and $t$ signals. A *sender strategy* is an $s \times t$ stochastic matrix $A = (a_{ij})$; $a_{ij}$ is the probability that the sender sends signal $j$ given state $i$. A *receiver strategy* is a $t \times s$ stochastic matrix $B = (b_{jk})$; $b_{jk}$ is the probability that the receiver chooses act $k$ given he receives signal $j$. In this case the pair of strategies $(A, B)$ is said to be a *signaling system*. Note that the matrix product $C = AB$ is an $s \times s$ stochastic matrix; if $C = (c_{ik})$, then $c_{ik}$ is the probability that the receiver chooses act $k$ given state $i$.

If $\pi_i$ is the probability that nature chooses state $i$, then $\sum_i \pi_i c_{ii}$ is the probability that a correct state-act pairing occurs. This is the same as the expected payoff for the signaling system (given the assumed payoff structure). By an *equilibrium* we mean a Nash equilibrium: i.e., neither the sender nor the receiver can increase the common expected payoff by one changing their strategy while the other does not. A *signaling system equilibrium* is a pair of strategies $(A, B)$ such that $C = I$, the identity matrix. In this case $t \geq s$ (there must be at least as many signals as states), and if $a_{ij} > 0$, then $b_{ji} = 1$.

A group of states is said to be *pooled* in a sender strategy if for every signal the probability of that signal being sent is the same for every state in the group. A *complete pooling equilibrium*, an equilibrium in which all the states are pooled; that is, it is a sender strategy $A$ all of whose rows are the same (the probability of sending of a particular signal is independent of the state), and therefore the rows of $C$ are also equal (the probability of choosing of a particular act is also independent of the state). A *partial pooling equilibrium* is an equilibrium in which some but not all of the states are pooled.

### A.2 The Ewens Sampling Formula

The Chinese restaurant process/Hoppe–Pólya urn provides a stochastic mechanism for generating the class of *random partitions* described by the *Ewens sampling formula*. Given $n$ objects divided into $t$ classes (each class containing at least one object), with $n_i$ in the $i$th class, let $a_j$, $1 \leq j \leq n$, record the number of classes containing $j$ objects (that is, $a_j$ is the number of $n_i = j$). Then $\langle n_1, \ldots, n_t\rangle$ is the vector of (class) *frequencies* and $\langle a_1, \ldots, a_n\rangle$ the corresponding *partition vector*. Clearly one has

$$\sum_{i=1}^{t} n_i = \sum_{j=1}^{n} j a_j = n, \quad \sum_{j=1}^{n} a_j = t$$

In the case of the Chinese restaurant process the number of classes present after $n$ objects are generated is itself random. In the most general case there is a free parameter $\theta$, and if $S_n(\theta) = \theta(\theta+1)(\theta+2)\cdots(\theta+n-1)$, then the probability of seeing a particular partition vector $\langle a_1, \ldots, a_n\rangle$ is given by the Ewens formula

$$P\big(\langle a_1, \ldots, a_n\rangle\big) = \frac{n!\theta^{\sum_{j=1}^{n} a_j}}{\prod_{j=1}^{n} j^{a_j} a_j! S_n(\theta)}.$$

If $\theta = 1$, then this reduces to the formula given earlier in this paper. If instead $\theta = 2$, say, then this would correspond to there being two black mutator balls (or phantom guests in the CRP version), rather than just one; and similarly for $\theta$ any positive integer. (But the formula makes sense for any positive value of $\theta$, integral or not; this corresponds to giving the black ball a relative weight of $\theta$ versus the unit weight given every colored ball.)

There is a rich body of theory that has arisen over the last four decades concerning the properties of random partitions whose stochastic structure is described by the Ewens formula.

If $T_n$ denotes the (random) number of different types or colors (other than black) in the urn after $n$ draws, then one can show that its expected value is

$$E[T_n] = \theta\left(\frac{1}{\theta} + \frac{1}{\theta+1} + \cdots + \frac{1}{\theta+n-1}\right).$$

If $\theta = 1$, this reduces to the $n$th partial sum of the harmonic series,

$$\sum_{k=1}^{n} \frac{1}{k} = \ln n + \gamma + \epsilon_n;$$

here $\gamma = 0.57721\ldots$ is Euler's constant, $\epsilon_n \sim 1/2n$ (that is, $\lim_{n\to\infty} 2n\epsilon_n = 1$).

## A.3 Two Limit Theorems

In the following we derive some of the properties of the signaling with invention process described in the section on inventing new signals.

### *A.3.1 The Number of Different Colors Diverges*

First some notation. Let $\mathcal{F}_n$ be the history of the process up to time $n$ (the $n$th trial). Let $A_n$ denote the event that on the $n$th trial a new signal is tried with success. Let $\omega$ denote the entire infinite history of a specific realization of our reinforcement process. (So if, instead, one were considering the process of tossing a coin an infinite number of times, $\omega$ would represent a specific infinite sequence of heads and tails.) Let

$$P(A_n \mid \mathcal{F}_{n-1})$$

denote the conditional probability that $A_n$ occurs, given the past history of the process up to time $n-1$. This is not a number, but a random quantity, since it depends on the realization $\omega$, which is random. Finally, let

$$P(A_n \mid \mathcal{F}_{n-1})(\omega)$$

denote this conditional probability for a specific history or realization $\omega$; this is a number.

By the martingale generalization of the second Borel–Cantelli lemma (see, e.g., [6, p. 249]), one has

$$\{A_n \text{ i.o.}\} = \left\{ \sum_{n=1}^{\infty} P(A_n \mid \mathcal{F}_{n-1}) = \infty \right\} \quad \text{almost surely.}$$

That is, consider the following two events. The first is

$$\{\omega : \omega \in A_n \text{ infinitely often}\};$$

the second event is

$$\left\{ \omega : \sum_{n=1}^{\infty} P(A_n \mid \mathcal{F}_{n-1})(\omega) = \infty \right\}.$$

The assertion is that the two events are the same, up to a set of probability zero.

We claim that $A_n$ occurs infinitely often with probability one; that is, $P(\{A_n \text{ i.o.}\}) = 1$. By the version of the Borel–Cantelli lemma just cited, it suffices to show that

$$P\left( \left\{ \omega : \sum_{n=1}^{\infty} P(A_n \mid \mathcal{F}_{n-1})(\omega) = \infty \right\} \right) = 1.$$

In fact we show more: that

$$\sum_{n=1}^{\infty} P(A_n \mid \mathcal{F}_{n-1})(\omega) = \infty$$

for *every* history $\omega$.

To see this, suppose that there are $k$ states, and that the sender selects these with probabilities $p_j$, $1 \le j \le k$. Suppose that initially there is just one ball, the black ball, in each of the $k$ urns. Then at each stage there are $a_j \ge 1$ balls in each urn, one black and the remaining $a_j - 1$ some variety of colors. The probability that at stage $n$ a new signal is generated and successfully used depends on the values of $a_1, \ldots, a_k$ at the start of stage $n$ (that is, before selection takes place), and is

$$\sum_{j=1}^{k} p_j \left( \frac{1}{a_j} \right) \left( \frac{1}{k} \right).$$

(That is, you pick the $j$th urn with probability $p_j$, you pick the one ball out of the $a_j$ that is black, and there is a one chance in $k$ that the receiver chooses the correct act.)

Now use the generalized *harmonic mean–arithmetic mean inequality* (see, e.g., [8]); this tells us that for $a_j > 0$, one has

$$\frac{1}{\sum_{j=1}^{k} (\frac{p_j}{a_j})} \le \sum_{j=1}^{k} p_j a_j.$$

Further, the total number of balls in the state urns, $a_1 + \cdots + a_k$, is greatest when a black ball has been selected every time and the receiver chooses the right act (since then one adds $k + 1$ balls of a new color at each stage rather than just one). Thus at the start of stage $n$ one has

$$a_1 + \cdots + a_k \leq (k + 1)(n - 1) + k$$

Thus if $p^\star = \max\{p_j, 1 \leq j \leq k\}$, it is apparent that

$$\sum_{j=1}^{k} p_j a_j \leq p^\star(a_1 + \cdots + a_k) \leq p^\star\big[(k + 1)n - 1\big].$$

Putting this all together gives us that the probability that at stage $n$ a new signal is generated and successfully used is

$$\frac{1}{k}\sum_{j=1}^{k}\left(\frac{p_j}{a_j}\right) \geq \frac{1}{kp^\star}\frac{1}{(k + 1)n - 1}.$$

It follows that

$$\sum_{n=1}^{\infty} P(A_n \mid \mathcal{F}_{n-1})(\omega) \geq \frac{1}{kp^\star}\sum_{n=1}^{\infty}\frac{1}{(k + 1)n - 1} > \frac{1}{k(k + 1)p^\star}\sum_{n=1}^{\infty}\frac{1}{n} = \infty,$$

using the well-known fact that the harmonic series diverges.

*A.3.2 The Number of Balls of a Given Color Diverges*

Suppose a color has been established. The above proof can be easily modified to show that the number of balls of a given color in all the state urns, say green, tends to infinity almost surely as $n$ tends to infinity. Let $A_n$ now denote the event that on the $n$th trial green is selected and reinforced; and $p_\star = \min\{p_j, 1 \leq j \leq k\} > 0$. Then in the receiver's green urn there must be at least one state that is represented at least $1/k$ of the time (since there are $k$ states and the sum of the fractions must sum to one). Let $a_{j_\star}$ denote the index of this state. (Note this state can vary with $n$.) Then one has

$$P(A_n \mid \mathcal{F}_{n-1}) \geq p_\star \frac{1}{a_{j_\star}}\left(\frac{1}{k}\right) \geq \frac{p_\star}{k[(k + 1)n - 1]},$$

and this again diverges. (This bound is obviously quite crude.)

It follows as a simple corollary that the signal corresponding to a given color is necessarily sent (almost surely) an infinite number of times (even though each time the signal is sent its color may or may not be reinforced).

## References

1. Aldous D (1985) Exchangeability and related topics. In: l'École d'été de probabilités de Saint-Flour, XIII–1983. Springer, Berlin, pp 1–198
2. Argiento R, Pemantle R, Skyrms B, Volkov S (2009) Learning to signal: analysis of a micro-level reinforcement model. Stoch Process Appl 119:373–390
3. Barrett JA (2007) Dynamic partitioning and the conventionality of kinds. Philos Sci 74:526–546

4. Bhaskar V (1998) Noisy communication and the evolution of cooperation. J Econ Theory 82:110–131
5. Crawford V, Sobel J (1982) Strategic information transmission. Econometrica 50:1431–1451
6. Durrett R (1996) Probability: theory and examples, 2nd edn. Duxbury Press, Belmont
7. Erev I, Roth AE (1998) Predicting how people play games: reinforcement learning in games with unique mixed quilibria. Am Econ Rev 88:848–881
8. Hardy GH, Littlewood JE, Polya G (1988) Inequalities, 2nd edn. Cambridge University Press, Cambridge
9. Hofbauer J, Huttegger S (2008) Feasibility of communication in binary signaling games. J Theor Biol 254:843–849
10. Hoppe FM (1984) Pólya-like urns and the Ewens sampling formula. J Math Biol 20:91–94
11. Lewis D (1969) Convention. Harvard University Press, Cambridge
12. Pawlowitsch C (2008) Why evolution does not always lead to an optimal signaling system. Games Econ Behav 63:203–226
13. Pitman J (1995) Exchangeable and partially exchangeable random partitions. Probab Theory Relat Fields 102:145–158
14. Robson A (1989) Efficiency in evolutionary games: Darwin, Nash and the secret handshake. J Theor Biol 144:379–396
15. Roth A, Erev I (1995) Learning in extensive form games: experimental data and simple dynamical models in the intermediate term. Games Econ Behav 8:164–212
16. Shreiber S (2001) Urn models, replicator processes and random genetic drift. SIAM J Appl Math 61:2148–2167
17. Taylor PD, Jonker LB (1978) Evolutionarily stable strategies and game dynamics. Math Biosci 40:145–156
18. Wärneryd K (1993) Cheap talk, coordination, and evolutionary stability. Games Econ Behav 5:532–546
19. Worden L, Levin S (2007) Evolutionary escape from the prisoner's dilemma. J Theor Biol 245:411–422
20. Zabell S (1992) Predicting the unpredictable. Synthese 90:205–232
21. Zabell S (2005) Symmetry and its discontents: essays in the history of inductive probability. Cambridge University Press, Cambridge
22. Zipf G (1932) Selective studies and the principle of relative frequency in language. Harvard University Press, Cambridge