Learning to Use Function Words in Signaling Games

Shane Steinert-Threlkeld
Department of Philosophy, Stanford University
shanest@stanford.edu

Abstract

In this paper, we explore a Lewis-Skyrms signaling game with function words (instead of just atomic signals) and reinforcement learning therein. Simulation results are presented and connections to other approaches are drawn.

Keywords: signaling games; function words; reinforcement learning

Signaling Games

Following Lewis 1969, the tools of game theory have been used to examine the conditions under which various forms of meaningful communication can emerge. We consider a simple game involving two players — a sender and a receiver — and n states, n acts and m signals. States are ways the world may be, signals are performances producible by the sender and discriminable by the receiver, and acts are potential responses on the part of the receiver. The game begins when nature selects a state. The sender observes the state and sends a signal. The receiver observes the signal and responds with an act. The sender and receiver are then each rewarded based on the pairing of state and act. A sender strategy is a function that maps each state to a signal, and a receiver strategy is function from signals to acts. We restrict our attention to games of pure coordination in which the payoffs are identical in each act/state pairing for sender and receiver. The reason for considering this class of games is that we are interested in the emergence of communication in circumstances where the participants in the signaling process begin with interests that are aligned.

Lewis considers only games with the same number of states, acts and signals. Supposing that for each state there is a unique act that yields a maximal payoff, we can then identify signaling systems as pairing of sender and receiver strategies whose composition always maps states to their ideal acts. If the number of signals does not match the number of states and acts there will be no signaling systems in this sense, although we can still identify equilibria as sender-receiver strategies such that a change on either the sender or receiver side will result in at least one player doing strictly worse and neither doing better.

Learning In Signaling Games

The static view of signaling games explored by Lewis allows us to identify equilibria for particular signaling games, but provides little insight into how participants might learn to coordinate their actions so as to enter these equilibria. Skyrms 2010 summarizes a body of work which investigates both evolutionary dynamics and learning in signaling games. Here we consider repeated play of signaling games with reinforcement learning. Instead of a sender strategy yielding a single determinate signal for each state, we instead let the sender strategy associate with each state a probability space over possible signals; similarly, a receiver strategy will associate with each signal a probability space over possible acts.

It may help to think of a sender strategy as a set of urns containing differently colored balls, one urn corresponding to each state and one color corresponding to each signal. The receiver strategy will then be another set of such urns, one for each signal, with the colors of the balls corresponding to possible acts. In a single iteration of the signaling game nature chooses a state as before, then the sender draws a ball blindly from the urn corresponding to that state and sends the signal corresponding to the color of that ball. The receiver then draws blindly from the urn corresponding to that signal and performs the act corresponding to the ball's color. If the act is appropriate for the state (if sender and receiver are rewarded) then the sender and receiver each place an additional ball of the same color into the urn from which it was drawn. Otherwise the number of balls in of each color in the urns is left unchanged. This simple model of reinforcement learning is called Roth-Erev reinfocement.[1]

In simple n state, n act, n signal games where a pairing of state i with act i yields a payoff of 1 and all other pairings receive payoff 0, simulations using this sort of reinforcement learning generally converge to signaling systems. When $n = 2$, there is even an analytic proof that this basic reinforcement learning is guaranteed to converge to a signaling system in the long run.[2] With successive iterations of the game the weights of the signals and acts will grow arbitrarily high, so that further reinforcement has an ever-diminishing effect. There will always be a non-zero probability of any signal being sent in a given state and of any act being per-

[1] It is first formalized in Roth and Erev (1995).

[2] See Argiento et al. (2009).

formed for a given signal, but the probability distribution for a given state or signal may approach a pure sender or receiver strategy arbitrarily closely. In general, however, reinforcement learning may yield mixed strategies, with non-negligible probabilities assigned to two or more signals for a single state or two or more acts for a single signal.

We need not limit reinforcement learning to games with a binary payoff structure (where each act/state pair yields a payoff of zero or one). In games where the payoffs span the interval $[0, 1]$ we can reinforce sender and receiver strategies proportional to the payoff. For example, following an iteration of a signaling game with payoff 0.8, we would add '0.8 balls' of the appropriate colors to the urns for the state and signal that occurred in the round of play. In other words, we keep track of *accumulated rewards*. For the sender, these will be values $ar(s, sig)$ for each state s and signal sig. We then have that

$$p(sig|s) \propto ar(s, sig)$$

and we update $ar(s, sig)$ after each iteration where that pair is played. Similarly, the sender has values $ar(sig, a)$ for every signal-act pair. Call this choice rule the *proportionality* rule.

There are, however, other choice rules available. For instance, we can use an exponential choice rule, so that the probability of a signal given a state or an act given a signal is not directly proportional to the accumulated rewards for the signal or act, but rather varies as $x^{\lambda * ar}$ for some x. When $x = e$, we call this the *exponential* rule. In other words:

$$p(sig|s) \propto \exp(\lambda \cdot ar(s, sig))$$

and similarly for signal-act pairs for the receiver. The λ parameter controls sensitivity of choice to noise. If $\lambda = 0$, then noise washes out all other considerations and all choices are equiprobable. When λ is given a very small value learning initially explores a large number of options, although after many iterations choice generally yields the most highly weighted option.

Disjunctive States and Cautious Acts

We have considered players in a signaling game confronted with a world partitioned into n possible states, each of which has its ideal act: the response which represents the best possible outcome for sender and receiver in that state. In the case of actual communication, however, the participants are not guaranteed to have perfect information about the state of the world. It may be that, for all the sender knows, one of some set of states is actual, though it is not known which one. In this case nature has put the sender in a state of imperfect knowledge with respect to the state of the world. We can reflect this situation in the payoff structure of a signaling game. In a game where ideal act/state pairs yield payoff one, we introduce the disjunctive state $i \vee j$ — in which nature has revealed to the sender that the state is either i or j, but not which — and reflect its disjunctive status in the payoff structure by assigning payoff 0.5 for a pairing of act i with state $i \vee j$ and similarly for a pairing of act j with state $i \vee j$.[3] This represents the fact that a sender-receiver strategy which always maps the disjunctive state onto one or the other of its disjuncts will achieve success half of the time.

We can further complicate the picture by introducing acts that are 'cautious' with respect to some number of states. An act cautious over states i and j will achieve a better payoff in state $i \vee j$ than either of act i or act j, but will not active the ideal payoff of 1 in any state. That is, the act cautious over i and j will have a payoff in the interval $(0.5, 1)$ in state $i \vee j$. There is some flexibility about how we assign payoffs for cautious acts in non-disjunctive states. For our purposes we will simply stipulate that an act cautious over i and j in state i, state j, state $i \vee k$ or state $j \vee k$ ($k \neq i$, $k \neq j$) will achieve $1/2$ the payoff for the cautious act in state $i \vee j$.

We can illustrate the setup with a somewhat contrived example: the sender has climbed an apple tree and is shaking a branch so that apples fall towards the ground where the receiver attempts to catch them in a basket. Each time the sender shakes the branch he can see that the apple will fall in one of three spots. The basic states are those in which an apple is falling into spot 1, 2, or 3. The basic acts involve the receiver moving to spot 1, 2, or 3 with the basket. Sometimes, however, the sender can see that the apple will fall in one of two spots, although he cannot tell which. In this case the state is disjunctive over those two spots, and the appropriate cautious act is for the receiver to move to a location in between the two spots. Pairings of acts 1, 2, and 3 with states 1, 2, and 3 respectively yield payoffs of 1, reflecting the fact that the receiver is in the ideal position to catch the apple before it hits the ground and is bruised. Pairings of the cautious acts with cautious states receive a payoff of 0.8, since these acts place the receiver in the best position to catch the apple given the imperfect information, although he is guaranteed to be standing some distance from whatever spot the apple falls in. If the receiver is standing between two spots i and j and it is known by the sender that the apple will definitely fall in one of those spots, or that the apple will fall in one of spots i or k (or one of spots j or k) but not which one, then the act is suboptimal given the sender's knowledge, and the payoff will be 0.4. The payoff matrix for the game will be as follows:[4]

	s_1	s_2	s_3	$s_{1\vee 2}$	$s_{1\vee 3}$	$s_{2\vee 3}$
a_1	1	0	0	0.5	0.5	0
a_2	0	1	0	0.5	0	0.5
a_3	0	0	1	0	0.5	0.5
$a_{1\vee 2}$	0.4	0.4	0	0.8	0.4	0.4
$a_{1\vee 3}$	0.4	0	0.4	0.4	0.8	0.4
$a_{2\vee 3}$	0	0.4	0.4	0.4	0.4	0.8

Figure 1: Payoff matrix for signaling game with three basic states and disjunctive states.

[3] In the fully general case, we can think of states coming from a Boolean algebra without top and bottom elements. But it's important to first get the simplest case working.

[4] See (Skyrms, 2010, p. 116).

Signaling Games with Function Words

The simple sorts of signaling games explored by Lewis involve signals that have no internal linguistic structure. In particular, the signals may be thought of as atomic content words. Following Skyrms, we can identify the informational content of a given signal in a given context with the change it yields in the prior probabilities of the various possible states in a particular sender-receiver strategy. Yet in human linguistic communication utterances contain repeatable subparts which may be combined with one another to systematically yield new meaningful expressions. If game-theoretic models are to provide us with a story about the circumstances under which communicative practices like human language might emerge, we would like to show how signals can come to display an internal structure such that the meanings of complex signals systematically vary with their constituents. In particular, we would like an explanation of the emergence of function words. These are words which have little lexical meaning but which combine with other phrases to alter their meanings. The example to be discussed here are the logical constants.

As in the earlier exposition of signaling games with reinforcement learning, sender and receiver strategies associate with each state and signal a probability distribution over possible signals and acts. Now, however, the sender occasionally sends sequences of two signals. To model this, we add to the 'urn' associated with each state a ball of a new color (i.e. a new signal). When this ball is drawn, the sender does not send it on its own, but rather must then draw a ball from another urn (this is a kind of 'nested urn': each of the state urns contains one of these new-signal urns) containing balls of the original colors.

A round of play begins as before, with nature selecting a state at random. If the sender chooses the new signal from the urn associated with that state, he will then choose a ball from the inner urn, and send both signals in a row: $sig_n sig_1$, for instance. The receiver is then confronted with a pair of signals, the first of which he interprets as a function on the state/act space in the following sense: the receiver's urn for sig_n does not contain acts, but rather *functions* from the set of actions to itself. The receiver will draw a function f from the sig_n urn, then draw an act a from the sig_1 urn, and finally will play act $f(a)$. Payoffs are determined as before by the state/act pairing. If there is a positive payoff, the sender reinforces the choice of sig_n in the state chosen by nature and the choice of the second signal in that state's 'nested urn'. The receiver reinforces his choice of function in the sig_n urn and the act he has chosen for the second signal.

Method

We apply this model to the six-state, six-act signaling game described in the previous section. We let there be three signals plus one new signal sig_4. We initialize the game in a signaling system with respect to the three standard signals, so that with high probability sender chooses signal i in state i, and receiver performs act i given signal i. In particular, the sender's starting strategy is

$$\begin{pmatrix} 100 & 1 & 1 & 1 \\ 1 & 100 & 1 & 1 \\ 1 & 1 & 100 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

where rows are states and columns are signals. The receiver's starting strategy is

$$\begin{pmatrix} 100 & 1 & 1 & 1 & 1 & 1 \\ 1 & 100 & 1 & 1 & 1 & 1 \\ 1 & 1 & 100 & 1 & 1 & 1 \end{pmatrix}$$

where rows are signals and columns are acts. We need to specify the receiver's sig_4 urn of functions. Because the space of all functions has size 6^6, we populate the urn with only three functions: the identity function, a constant function, and a "negation" function, given by:

$$\begin{array}{ccc} a_1 \mapsto a_{2\vee 3} & a_2 \mapsto a_{1\vee 3} & a_3 \mapsto a_{1\vee 2} \\ a_{1\vee 2} \mapsto a_3 & a_{1\vee 3} \mapsto a_2 & a_{2\vee 3} \mapsto a_1 \end{array}$$

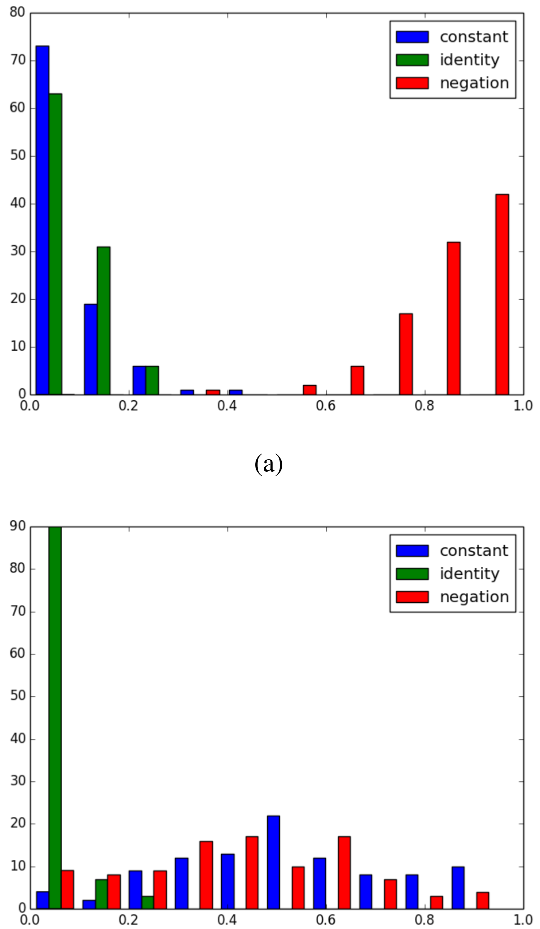
The receiver's sig_4 urn is initialized as $(1, 1, 1)$. We considered two constant functions: one which returned a_1 always and one which returned $a_{1\vee 2}$ – which is the appropriate act in a disjunctive state – always. States were not equiprobable, but rather proportional to $(1, 1, 1, 5, 5, 5)$ where the latter three are the three disjunctive states. These weights ensure that such states will be visited often enough.[5]

This setup is of interest because it provides us with a scenario in which a function word that behaves like a negation operator is of obvious use. If the players can learn to use sig_4 to map states to their complements in the state/act space, then they can achieve the equilibrium given by the diagonal in figure 1 above while only using four signals (instead of six atomic signals as would normally be required). It is clear that the number of discriminable states of the world and appropriate behavioral responses thereto that confront human beings is far greater than the number of distinct signals an individual person can be expected to master over the course of a lifetime. If we can display a simple model where signalers learn to use a negation-like function to achieve equilibria when the number of states and acts exceeds the number of available signals, this lends some plausibility to the idea that communication using this logical function is a natural response to a world whose complexity (in terms of the number of discriminable states and appropriate responses to those states) exceeds the number of available signals, and some of whose states and acts bear the disjunctive/cautious relation to one another.

Results

Figure 2 shows the results for our basic learning set-up with 3 functions and a proportionality choice rule. On the x-axis are 10 bins corresponding to probabilities of the receiver

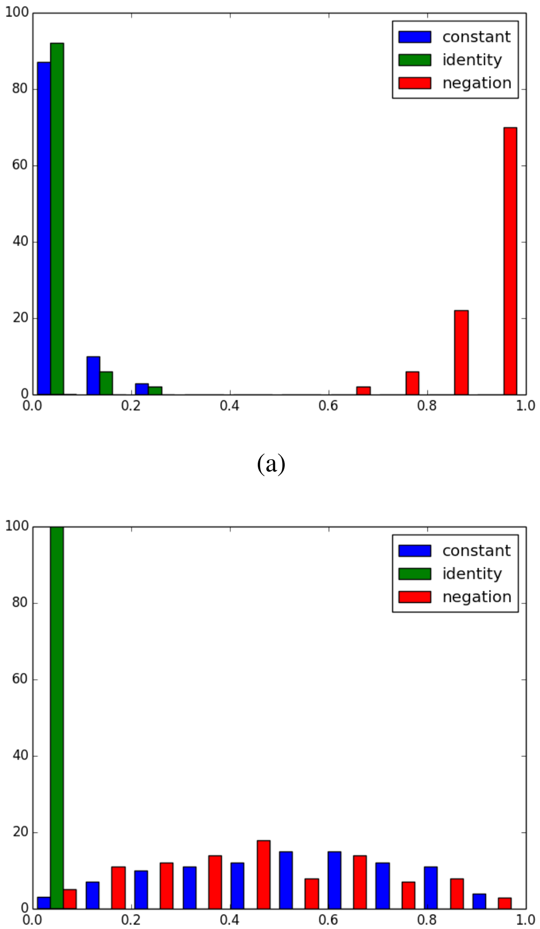
[5] Initial experiments with equal weights on the states suggest that the main results are not effected by this decision.



(a)

(b)

Figure 2: 100 trials of reinforcement learning with 10^4 iterations. Choice rule: proportionality to accumulated rewards. (a) Constant function returns a_1. (b) Constant function returns a_4.



(a)

(b)

Figure 3: 100 trials of reinforcement learning with 10^5 iterations. Choice rule: proportionality to accumulated rewards. (a) Constant function returns a_1. (b) Constant function returns a_4.

choosing a function given that the sender sent sig_4. The y-axis is the number of trials (out of 100) in which the final probability was in the corresponding range. In 2 (a), the constant function returned a_1, which is appropriate in s_1; in 2 (b), the constant function returned $a_{1\vee 2}$ which is appropriate in the disjunctive state $s_1 \vee s_2$.

There are a couple of interesting things to note about these results. In the (a) case, the probability of choosing the negation function is almost always at least 0.5. Moreover, 42 trials ended with negation being chosen over 90% of the time. In the (b) case, things are messier. Only 4 trials ended with $p(neg|sig_4) > 0.9$, compared to 9 with $p(neg|sig_4) < 0.1$. The reason that the constant act being $a_{1\vee 2}$ has such a dramatic effect is that the difference between the payoff of $a_{1\vee 2}$ in $s_1 \vee s_2$ and other acts in that state is much smaller than when the appropriate act is taken in one of s_1, s_2, s_3. Similarly, $a_{1\vee 2}$ does fairly well in all three disjunctive states. Thus, it should take longer for the accumulated rewards to differentiate between doing $a_{1\vee 2}$ in all of $s_{1\vee 2}, s_{1\vee 3}, s_{2\vee 3}$ and using negation to get the appropriate act. To test this, we did two things. First, we re-ran the same experiment with 10^5 iterations. These results are shown in figure 3. Secondly, we took a closer look at how $p(\cdot|sig_4)$ was evolving within individual trials. To this end, figure 4 shows a representative single trial of just 10^4 iterations in the (b) case where the constant function returns $a_{1\vee 2}$.

The differences between figure 2 and figure 3 are quite interesting. In the (a) case, we see a general trend for $p(neg|sig_4)$ to be higher at the end of a trial. For instance, 65 trials ended with $p(neg|sig_4) > 0.9$ in this case (compared to 42). These results suggest that when the constant function returns a_1, the reinforcement learning does have $p(neg|sig_4)$ converging to 1 as the number of iterations increases, but that it is simply a slow learning algorithm. On

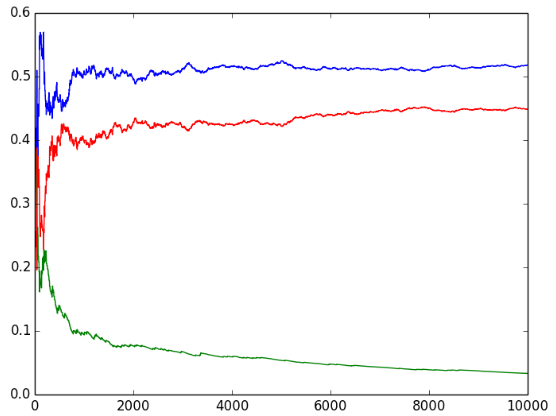


Figure 4: One trial of 10^4 iterations. Blue: $p(const|sig_4)$, red: $p(neg|sig_4)$, green: $p(id|sig_4)$.

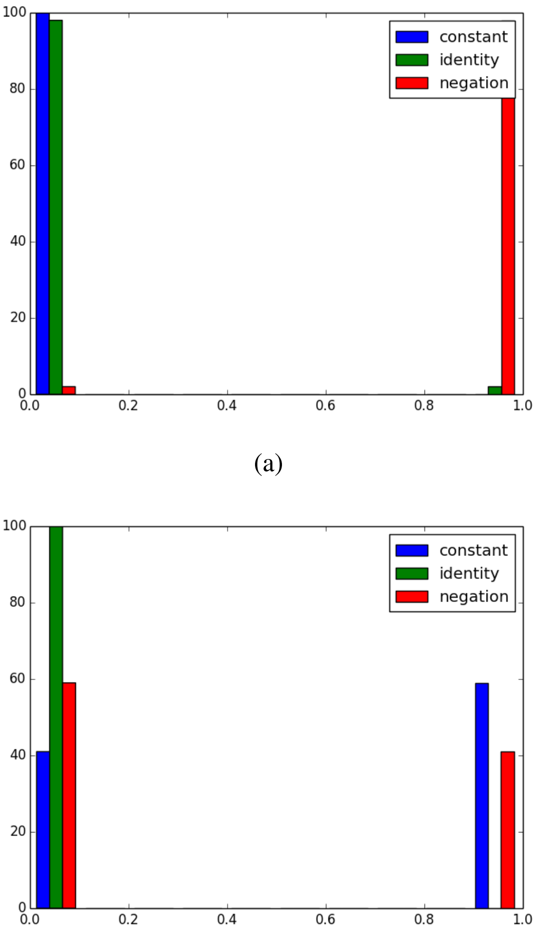
the contrary, figure 3 (b) suggests that something different is happening in the case where the constant function returns $a_{1\vee2}$. In particular, $p(id|sig_4) < 0.1$ in all trials. But the distribution of values that $p(const|sig_4)$ and $p(neg|sig_4)$ converge to appears to be approaching a normal distribution with mean 0.5 as the number of iterations increases. Inspection of individual trials does reveal that in this case, the two probabilities are in fact converging to non-extremal values. Figure 4 shows a sample trial. The two probabilities appear to have nearly converged after just 10^4 iterations.

To attempt to overcome the inability to reliably learn to interpret sig_4 as negation, we tried altering the choice rule. In particular, we re-ran the experiment with 100 trials and 10^4 iterations, but this time with an exponential choice rule:

$$p\left(a|s\right) \propto \exp\left(\lambda \cdot ar(a,s)\right)$$

The thought with this choice rule is that the small differences in payoff between the most appropriate action and other actions in disjunctive states will exert a larger influence on choice since the exponential makes the difference bigger. Skyrms 2010, p. 99 notes that this choice rule ensures convergence in certain basic signaling games where Roth-Erev with a simple proportional choice rule does not.

The results of this experiment are shown in figure 5. This data corresponds to the choice $\lambda = 0.1$. When the constant function returns a_1 (5 (a)), ninety-nine trials reached $p(neg|sig_4) > 0.9$ in just 10^4 iterations. Inspection shows that most of these probabilities were above 0.97. Interestingly, when the constant function returns $a_{1\vee2}$ (5 (b)), we still have that all trials converge to extremal probabilities, but that only 41 of these are with high $p(neg|sig_4)$. This shows that although the exponential choice rule does ensure convergence to extremal values, it does not guarantee convergence to the ideal extremal values. There appears to be strong sensitivity to the states and acts chosen in the early iterations. Tables 1 and 2 show the dependence of these results on the choice of λ for the cases where the constant function returns a_1 and $a_{1\vee2}$ respectively. Note that in these tables, ‘extremal’ means either $p(neg|sig_4) < 0.05$ or $p(neg|sig_4) > 0.95$. In particular, no choice of the learning parameter λ ensures reliable successful convergence when the constant function returns a cautious act $a_{1\vee2}$.



(a)

(b)

Figure 5: 100 trials of reinforcement learning with 10^4 iterations. Choice rule: exponential, with weight $\lambda = 0.1$. (a) Constant function returns a_1. (b) Constant function returns a_4.

Discussion

In certain conditions, we do find that a simple modification of reinforcement learning allows signalers to learn to treat a signal as a function word. In particular, when the constant function returns a_1, both proportional and exponential choice appear to have $p(neg|sig_4) \to 1$, with the latter just converging faster. But with the constant function returning $a_{1\vee2}$, both choice rules fail to deliver this kind of sure convergence. Although one can tinker with the payoff table to try and get the desired result, we think that our simulations can be seen as fairly limitative. Conditions for learning to treat sig_4 as negation were quite optimal: there are only

λ:	0.025	0.05	0.1	0.25	0.5
success:	0	83	93	82	71
extremal:	0	83	97	100	100

Table 1: Effects of λ when constant returns a_1

λ:	0.025	0.05	0.1	0.25	0.5
success:	0	21	39	41	43
extremal:	0	53	98	100	100

Table 2: Effects of λ when constant returns $a_{1\vee 2}$

three functions to learn over (instead of the full space of 6^6), the sender/receiver are already in a signaling system over the basic states and signals, and the disjunctive states are more probable. Given the inability of our learning algorithm to ensure convergence in the presence of the constant $a_{1\vee 2}$ function even under these ideal conditions, one may doubt whether this is the right approach to study the evolution of function words. To that end, we briefly compare this study to others in the literature and mention a few possible future developments.

Connection to Other Work

Nowak & Krakauer 1999 In this paper, the authors study the emergence via a kind of natural selection of signal-object pairings. For our purposes, the most interesting case is the last, where they consider two objects and two properties, for four total combinations. These combinations can be specified by four atomic words $w_1, \cdots w_4$ or with pairs $p_i o_j$. Nowak and Krakauer consider a strategy space where players use the atomic words with probability p and the 'grammatical' constructs with probability $1 - p$. They are able to show that the only two evolutionary stable strategies are when $p = 0$ and $p = 1$ and that their evolutionary dynamics evolves to use the grammatical rule with probability 1. In trying to use syntactic structure to mirror structure in the world, our approach does have something in common with Nowak and Krakauer's. They, however, are still not explicitly interested in function words.

Barrett 2006; 2007; 2009 In a series of papers, Barrett considers 'syntactic' signaling games with reinforcement learning where the number of states and acts exceeds the number of signals, and where the number of senders is greater than 1. These are equivalent to games in which there is one sender who sends a sequence of signals for each state (the length of the sequence is fixed and the strategies differ for each position in the sequence). Perfect signaling is achieved when sender-receiver strategies settle on a coding system with a sequence that elicits the ideal act in each state. Barrett found that in simulations involving 4-state, 4-act, 2-signal, 2- sender signaling games perfect signaling is achieved by reinforcement learning approximately 3/4 of the time, and in 8-state, 8-act, 2-signal, 3-sender games perfect signaling is achieved 1/3 of the time. In the 4-state case, each of the two senders has a partition of size two on the state space; the equilibrium is achieved when the receiver takes the intersection of the sets in which the senders sent their respective signals. The difference, then, between this approach and our own is that there is no explicit signal that serves as a function word in Barrett's set-up; rather, signal concatenation is implicitly treated like conjunction.

Franke 2013 Here, the explicit concern is with using reinforcement learning in signaling games to explain compositional meanings. Although we will not go into the details due to space, Franke's complex signals have the form m_{AB}; the sense in which these are compositional is captured in a distance $s = d(m_{AB}, m_A) = d(m_{BA}, m_A) = d(m_{AB}, m_B) = d(m_{BA}, m_B)$. As these equalities show, the model ignores any kind of word order in the complex signals. Therefore, nothing like the role of a function word can be found.

Future Work

The main dimension that our setup fails to address is the *origin* of function words. Our model effectively builds in that sig_4 is a function word. While we can interpret the constant function as ignoring the second signal sent and the identity function as ignoring sig_4, it would still be nice to have a story about why having function words at all confers some kind of evolutionary advantage. One attempt would modify reinforcement-with-invention as in Alexander, Skyrms, and Zabell (2011) to invent function words; nevertheless, the 'functional' aspect would probably have to be hard-coded into that framework as we currently do. A second attempt would be to do a kind of replicator-mutator dynamic where some of the mutations involved *playing a new game* (our game, with function words) instead of just adopting a new strategy. One would then hope to show that a population using function words can successfully invade one that is not. We leave the general issue of exploring the origin of function words for future work.

References

Alexander, J. M.; Skyrms, B.; and Zabell, S. L. 2011. Inventing New Signals. *Dynamic Games and Applications* 2(1):129–145.

Argiento, R.; Pemantle, R.; Skyrms, B.; and Volkov, S. 2009. Learning to signal: Analysis of a micro-level reinforcement model. *Stochastic Processes and their Applications* 119(2):373–390.

Barrett, J. A. 2006. Numerical Simulations of the Lewis Signaling Game: Learning Strategies, Pooling Equilibria, and the Evolution of Grammar. Technical report, Institute for Mathematical Behavioral Sciences.

Barrett, J. A. 2007. Dynamic Partitioning and the Conventionality of Kinds. *Philosophy of Science* 74:527–546.

Barrett, J. A. 2009. The Evolution of Coding in Signaling Games. *Theory and Decision* 67(2):223–237.

Franke, M. 2013. Compositionality from Reinforcement Learning. In *Proceedings of Games, Interactive Rationality, Learning (G.I.R.L.) 2013*, number i.

Lewis, D. 1969. *Convention.* Blackwell.

Nowak, M. A., and Krakauer, D. C. 1999. The evolution of language. *Proceedings of the National Academy of Sciences of the United States of America* 96:8028–8033.

Roth, A. E., and Erev, I. 1995. Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term. *Games and Economic Behavior* 8:164–212.

Skyrms, B. 2010. *Signals: Evolution, Learning, and Information*. Oxford University Press.