How signaling conventions are established

Calvin T. Cochran*
Jeffrey A. Barrett

November 14, 2020

Abstract

We consider how human subjects establish signaling conventions in the context of Lewis-Skyrms signaling games. These experiments involve games where there are precisely the right number of signal types to represent the states of nature, games where there are more signal types than states, and games where there are fewer signal types than states. The aim is to determine the conditions under which subjects are able to establish signaling conventions in such games and to identify a learning dynamics that approximates how they succeed when they do. Our results suggest that human agents tend to use a win-stay/lose-shift with inertia dynamics to establish conventions in such games. We briefly consider the virtues and vices of this low-rationality dynamics.

1 Introduction

David Lewis (1969) introduced signaling games to study how linguistic conventions might be established without appeal to prior conventions. Lewis signaling games are classical games that presuppose sophisticated players possessing both a high level of rationality and access to natural saliences. Brian Skyrms (2010, 2014) subsequently showed how to represent the basic structure of Lewis signaling games as evolutionary games that may be considered in a population or a learning context and that may lead to the establishment of successful conventions even when played by low-rationality agents without access to natural saliences. Here we focus on how human agents in fact establish linguistic conventions during a Lewis-Skyrms evolutionary game in a learning context. In brief, we find that they often evolve successful conventions by gradually tuning their strategies on the basis of a *win-stay/lose-shift with inertia* dynamics, a low-rationality trial and error learning dynamics. We will discuss how signaling games work, then turn to consider the empirical data.

A Lewis-Skyrms signaling game is a common interest evolutionary game played between a *sender* who can observe nature but not act and a *receiver* who can act but not observe nature. The simplest signaling game is a $2 \times 2 \times 2$ game, where the first number is the number of possible states of nature, the second is the number of possible signals the sender might send, and the third is the number of acts the receiver might perform after getting a signal. Here nature randomly produces one of the two possible states in an

*To contact the authors, please write to: Calvin Cochran, 110 Central St. 32, Wellesley, MA 02482; email:cc4@wellesley.edu

unbiased way, the sender observes the state, then sends one of her two available signals to a receiver who performs one of his two possible actions as illustrated in Figure 1. Each of the two actions corresponds to one of the two states of nature. The players are successful if and only if the receiver's action matches the current state of nature.

In a $2 \times 2 \times 2$ signaling game, neither of the two signal types begins with a meaning. If the agents are to be successful in the long run, they must *evolve* signaling conventions where the sender communicates the current state of nature by the signal she sends and the receiver interprets each signal appropriately and produces the corresponding act. Such conventions may evolve gradually by means of a low-rationality learning dynamics like *simple reinforcement* (SR).

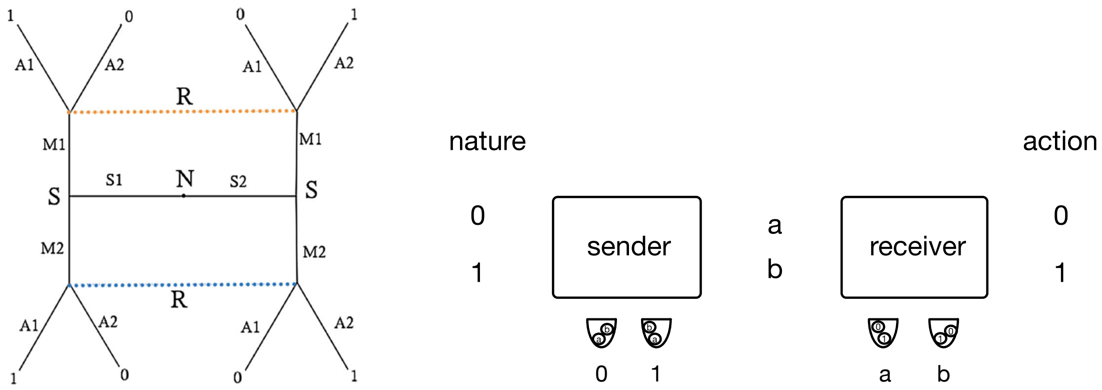


Figure 1: A basic $2 \times 2 \times 2$ signaling game. Following Bruner et al. (2018)), the diagram on the left depicts the extensive form of the game. Nature (N) passes state S1 or S2 to the sender (S), who passes message M1 or M2 to the receiver (R), who chooses act A1 or A2. Choosing the correct act yields payoff of 1 for both players and 0 otherwise. The dotted lines connecting nodes represent information partitions for the receiver: she cannot distinguish which of the nodes she has arrived at if she arrives at either. The diagram on the right gives another view of the two players as they are prepared to update using simple reinforcement. States (and acts) are 0 and 1; signals are a and b. Each agent has two urns corresponding to their two possible inputs and each of those start with two balls, one for each possible output. Additional balls are added when the players succeed.

SR is among the simplest of trial and error learning dynamics, and there is a long tradition of using it to model human learning.[1] On this dynamics one might picture the evolution of the two players' dispositions in a signaling game in terms of balls and urns. In the $2 \times 2 \times 2$ signaling game, one might imagine the sender with two urns, one for each state of nature (0 or 1), each beginning with one a-ball and one b-ball. The receiver has two urns, one for each signal type (a or b), each beginning with one 0-ball and one 1-ball. The sender observes nature, then draws a ball at random from her corresponding urn. This determines her signal. The receiver observes the signal, then draws a ball from his corresponding urn. This determines his act. If the act matches the state, then it is successful and each agent returns the ball drawn to the urn from which it was drawn and adds a duplicate of that ball. If unsuccessful, then, on basic reinforcement learning, each agent simply returns the ball drawn to the urn from which it was drawn. In this way successful dispositions are made more likely conditional on the states that led to those

[1] See Herrnstein (1970) for the basic theory and Erev and Roth (1998) for an example of experimental results for human agents.

actions.

Since the initial urn contents are unbiased, the sender's signals clearly have no pre-established meanings. She begins by signaling randomly and the receiver begins by acting randomly, but as they play, the sender's signals may acquire meanings that allow for increasingly successful actions by the receiver. Indeed, if nature is unbiased, then one can prove that a 2 × 2 × 2 signaling game will almost certainly converge under simple reinforcement learning to a signaling system where each state of nature produces a signal that leads to an action that matches the state.[2]

Importantly, $n \times n \times n$ signaling games do not always converge to optimal signaling systems on SR for $n > 2$. For n = 3 about 9% of runs fail to converge to optimal signaling on simulation . For $n = 4$ the failure rate is about 0.21. And for $n = 8$ it is about 0.59.[3] When the game does not converge to an optimal signaling system, the players get stuck in one of a number of sub-optimal pooling equilibria associated with different success rates.

The behavior of agents in $2 \times m \times 2$ population games are central to the experiments we considered. To give a baseline idea of what the evolution of agent dispositions looks like in such games, Figure 2 shows the relative rates of convergence under SR learning with six players in the sender population and six in the receiver population when players are randomly paired in each round. Note that the simulated SR agents do best with four signals and that it takes a few thousand plays for the agents to evolve firm conventions.

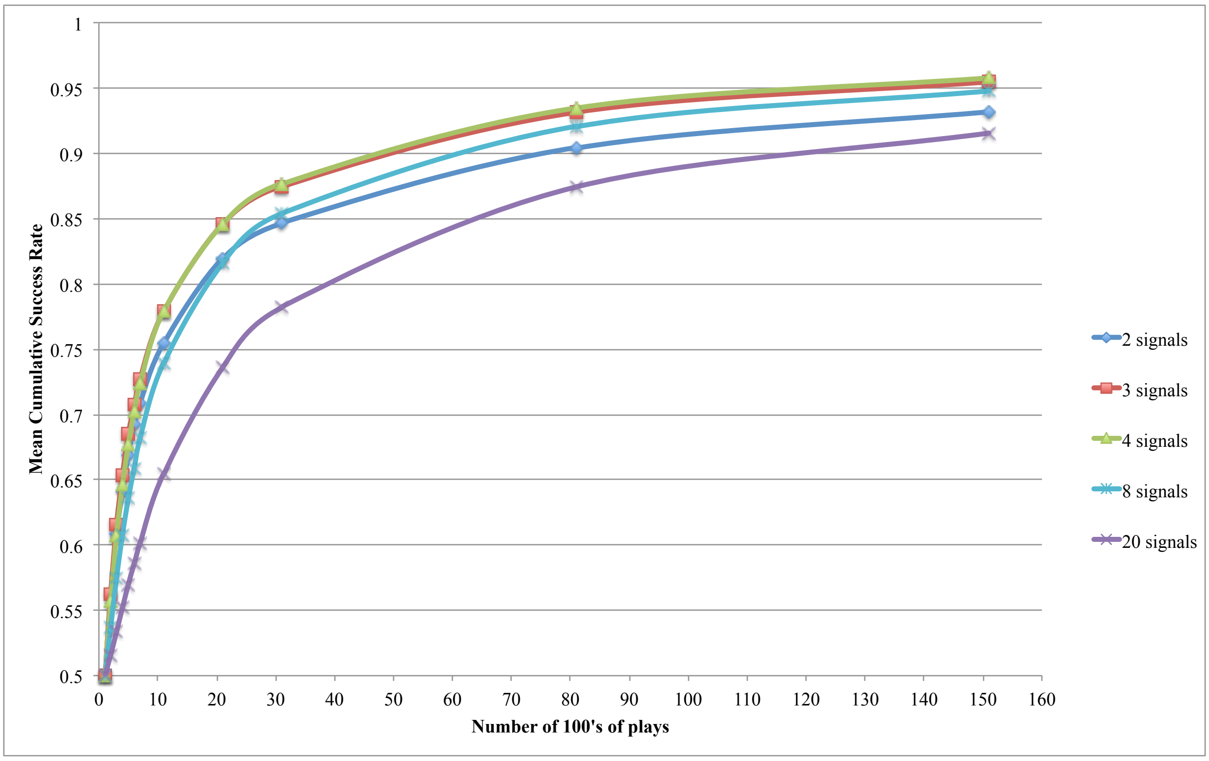


Figure 2: Mean cumulative success rate for $2 \times m \times 2$ population games (six players in each population) across 10,000 runs each under simple reinforcement learning. Agents equipped with four signals learn fastest in the short and medium run.

[2] See Argiento et al. (2009) for the proof.

[3] See Barrett (2006) for a discussion of these results and what they mean.

Figure 2 illustrates that the rate at which agents are observed to approach optimal signaling is a non-monotonic function of n, the number of distinct signals available: four signals allows for the quickest convergence, and having more or less slows things down. This counterintuitive phenomenon can be explained as a trade-off. On the one hand, extra signals afford the players flexibility in the signaling systems they can reach. For instance, if n = 2 then there are only two signaling systems at which players might arrive. If n = 3 then there are many more (some featuring effectively mixed strategies, which we discuss later). This flexibility expedites the speed at which players approach optimal signaling. On the other hand, too many signals may delay the evolution of successful dispositions, especially in the early game. In our simulations, n = 4 appears to strike the optimal balance between these two factors.

SR learning allows for a number of natural variants. One might consider starting with urns populated by more balls or adding more than one ball of the type that led to the successful action on a successful play. On reinforcement learning with punishment, balls of the type used on an unsuccessful play are removed from the urns that were consulted in producing the unsuccessful action. Reinforcement learning with punishment often does significantly better than SR in evolving signaling conventions, but if the level of punishment on failure is too high relative to the level of reward on success, the agents may not be able to learn anything at all.

Lewis-Skyrms signaling games have been studied extensively under a variety of learning dynamics.[4] Experiments have also been done to observe the behavior of human agents when they play such games in the laboratory. Blume et al. (1998) shows that human subjects were able to evolve conventions in 2 × 2 × 2 games relatively quickly when arranged into small sender and receiver populations. Blume et al. (2001) provides further empirical results on up to 3 × 3 × 3 signaling games with only partially aligned interests. And Blume et al. (2002) provides some evidence of learning in signaling games. More recently, Bruner et al. (2018) provide further empirical evidence supporting a number of theoretical predictions for 2 × 2 × 2 and 3 × 3 × 3 games, and Rubin et al. (2019) provide experimental results on the sim-max game, a close cousin of Lewis signaling games in which a similarity metric is defined over the states.

Taken together, these experiments show that human agents are able to form signaling game conventions in a variety of contexts. Our aim here is to investigate *how* human agents in fact learn to form conventions in such games. To this end, we consider experiments involving games where there are precisely the right number of signal types to represent the states of nature, games where there are more signal types than states, and games where there are fewer signal types than states. The present paper represents the first experimental investigation into the effect of message space size on human subjects' speed and ability to reach a convention in complete common interest signaling games[5]. As such it provides special insight into how subjects learn in the context of signaling games.

In brief, we find that human subjects often significantly outperform well-studied low-rationality learning dynamics like SR in the context of Lewis-Skyrms signaling games. Their success, however, is not the result of high-rationality learning. Rather, the evi-

[4] See, for example, Barrett and Zollman (2009), Huttegger et al. (2014), Barrett et al. (2017).

[5] The experiments discussed in Blume et al. (2008) are similar to ours in that they investigate whether the sender having more or less available signal types impacts player behavior in the lab. The game they investigate, though, is a variant of the traditional Lewis signaling game in which players have only partially aligned common interests.

dence suggests that it is primarily the result of their using a low-rationality dynamics *win-stay/lose shift with inertia* (WSLSwI). While WSLSwI is well-suited to establishing conventions quickly and reliably in the context of a broad array of signaling games, there are some where it does not work well at all. That human subjects have difficulty establishing signaling conventions in precisely those games is part of the evidence that they are relying on this learning dynamics.

2 Experimental setup

The basic structure of the present experiments mirrors that of Bruner et al. (2018). Sixteen total sessions were conducted. Twelve subjects participated in each session. Subjects were seated at divider-separated computers and then followed the prompts on their screen until the experiment's conclusion.[6] Subjects engaged in between one and three of the following treatments during each session: $2 \times 2 \times 2$, $2 \times 3 \times 2$, $2 \times 6 \times 2$, and $3 \times 2 \times 3$. The $3 \times 2 \times 3$ treatment was administered 12 times across 12 sessions; each of the others was administered 8 times. Treatment order was varied within each session to minimize ordering effects.

At the beginning of the session, the computer displays instructions for the game and randomly divides subjects evenly into a group of six senders ("Group A"[7]) and a group of six receivers ("Group B"). In part I of the experiment subjects participate in (usually[8]) 60 rounds of the first treatment. During every round each sender is paired randomly with a receiver. Subjects are aware that their opponent is arbitrarily chosen but do not know their identity. Once paired for a round, the computer randomly generates six states of nature (one for each sender) and reveals these to the appropriate senders. Each sender is then prompted to choose a "message" to send to her receiver partner in Group B. She does this by clicking a button with the desired signal (see Figure 4). Each receiver then witnesses her sender partner's signal and chooses an act. All players are then taken to a round results screen in which each pair is informed of the state of nature, the sender's signal, the receiver's act, and whether they succeeded on this round (see Figure 5). Similar to Bruner et al. (2018), and notably different from Blume et al. (1998)[9], subjects are not given access to the decisions of other players or to a history table of their own actions. Once all players press "Next" on the round results screen, the next round commences. This iterated process is illustrated in Figure 3. When all rounds of a treatment have been played, instruction and then play for the next treatment began. All subjects are made aware of the state space and signal space before each treatment.

Once all treatments are complete and a post-play survey has been taken, the session is over and subjects are paid in cash according to their performance. At the beginning of the experiment, the subjects are informed that, in addition to the $7 show-up payment, two rounds from each treatment would be randomly selected and they would be paid $4

[6] The experiment was written using the software oTree (Chen et al., 2016).

[7] Terms such as sender and receiver were avoided in an effort to keep players' potential predisposed biases about communication at bay as they developed their own language.

[8] The exception being the four sessions in which the only administered treatment was $3 \times 2 \times 3$. These ran for 90 rounds to see if extra time might assist players in reaching a stable convention.

[9] In the Blume et al. (1998) experiment, the decisions of all subjects were revealed to all participants at the end of each round. Their subjects could also view all decisions of all participants from past rounds as well. As language users are not generally privy to all conversations and have limited memory, our experiment only grants subjects knowledge of their own current play and that of their present partner.

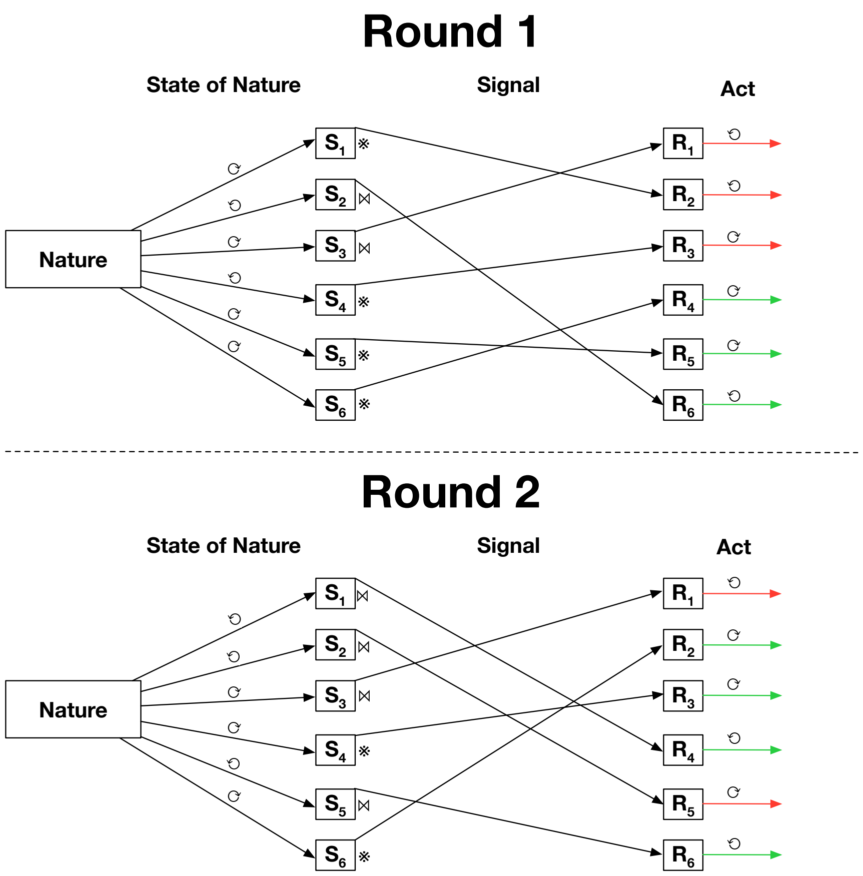


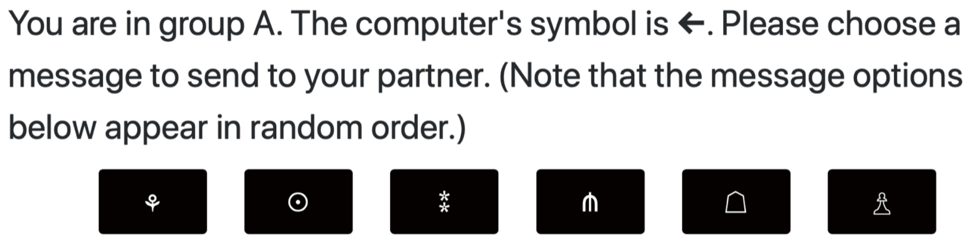
Figure 3: Example rounds of play in a 2 × 2 × 2 treatment. In round 1, each sender is matched with a random receiver. Nature chooses a unique state for each sender (here, ↻ or ↺). Each sender sees their state and sends a signal to their receiver. Each receiver sees their signal and chooses an act. They and their sender partner are either correct (green) if their act matches the state or incorrect (red) otherwise. In round 2, senders are again matched to (perhaps different) random receivers and play starts again.

for each success and $1 for each failure in those rounds at the end of the game.[10]

Much of the theoretical work on signaling games assumes that senders are not predisposed to associate a particular signal with a particular state of nature a priori (and similar for receivers when choosing their act), but this may not be true in practice. When selecting a word to represent the sound made when one hiccups, for instance, the word "hiccup" may have been a preferred candidate because it sounds somewhat like the phenomenon it describes. These potential naming conventions for "Hiccup" and other cases of onomatopoeia are examples of what Schelling (1960) calls *focal points*: game actions which somehow "jump out" at players in coordination games and which therefore may be selected in order to facilitate cooperation. In his original argument, Lewis (1969) appeals to such natural saliences in explaining how agents may arrive at a convention in the context of a single-shot classical game.[11] Signaling game models which assume no natural

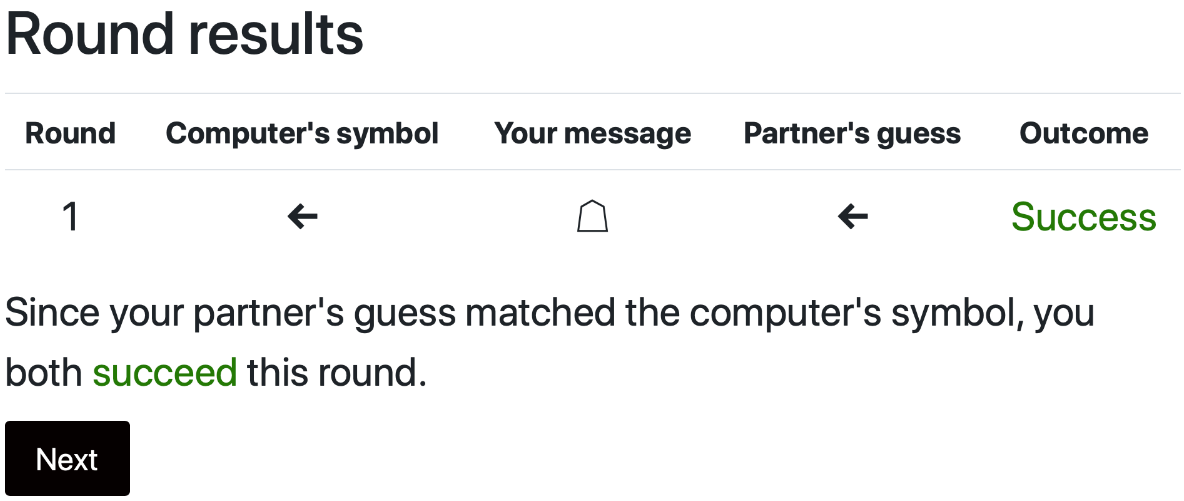
[10] This mimics the payoff scheme in Bruner et al. (2018). In the present case the minimum subject payoff possible across all experiments is $9 (achievable only in the isolated $3 \times 2 \times 3$ session); the maximum is $31. Paying subjects only in certain rounds and not telling them which rounds they are being paid is in part to avoid wealth effects.

[11] Skyrms' evolutionary signaling games have no natural saliences. That they allow one to explain



You are in group A. The computer's symbol is ←. Please choose a message to send to your partner. (Note that the message options below appear in random order.)

Figure 4: Sender is prompted with state of nature and must select signal.



Round results

Round	Computer's symbol	Your message	Partner's guess	Outcome
1	←	⌂	←	Success

Since your partner's guess matched the computer's symbol, you both succeed this round.

Figure 5: Round results screen.

salience might be thought of as worst-case scenarios in which coordination is maximally difficult to achieve due to a lack of focal points.

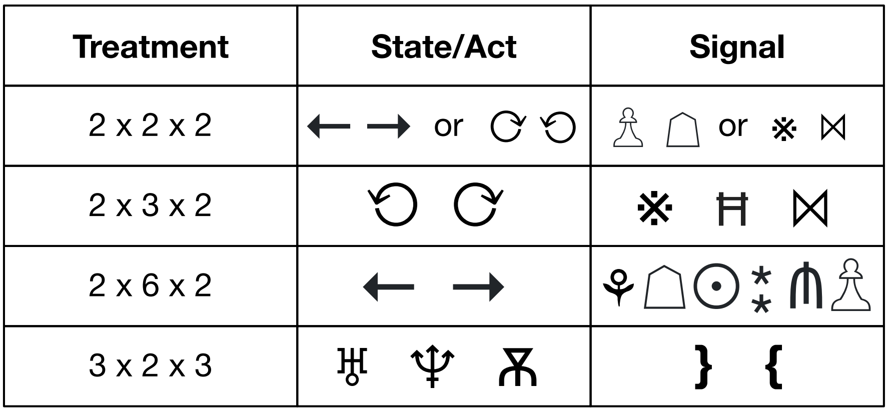
In the present experiment signals were carefully selected to minimize natural salience: each collection of symbols are either reflections of one another or exhibit bilateral symmetry. These[12] symbols are displayed in Figure 6. While an individual participant may immediately associate a particular signal (act) with a state (signal), this poses no problem as it only assists subjects' coordination if the salience is identified systematically by multiple players. The order in which signal (act) buttons were presented to senders (receivers) was also randomized for each subject and in each round to prevent symbol order from becoming a source of salience.

3 Macro-results

Subjects in the 2 × 2 × 2, 2 × 3 × 2, and 2 × 6 × 2 treatments had varying degrees of success in approaching a signaling system, but, in all cases, there were more runs that led to firmly established conventions than not. In the 2 × 2 × 2 treatment, seven of

the evolution of conventions is all the more remarkable. Lacroix (2018) introduces a variant of Skyrms signaling games in which the degree of natural salience can be varied by setting a parameter.

[12] While the 2 × 2 × 2 treatment shares some of its state and signal symbols with the 2 × 3 × 2 and 2 × 6 × 2 treatments, no state or signal was included in multiple treatments of the same session.



Treatment	State/Act	Signal
2 x 2 x 2	← → or ↻ ↺	♙ ☖ or ※ ⋈
2 x 3 x 2	↺ ↻	※ ⛩ ⋈
2 x 6 x 2	← →	♀ ☖ ☉ ⁑ ₥ ♙
3 x 2 x 3	♅ ♆ ⯓	} {

Figure 6: List of state, signal, and act symbols used in different treatments.

the eight sessions approached an optimal convention. In the 2 × 3 × 2, six of the eight did; in the 2 × 6 × 2, five of the eight reached equilibrium. As noted in Bruner et al. (2018), we found that subjects seldom converge to a 100% successful signaling system. Forgetting, experimentation, or stubbornness in a handful of players usually precluded this. In fact, excessive stubbornness and desultory strategies in a few individuals appear to be partially responsible for the few 2 × n × 2 runs which did not succeed[13]. Figure 7 summarizes participants' observed success rate over all 60 rounds for converging runs (Figure 8 shows the same thing but for all runs). Since subjects' success on each round can vary wildly (especially during early plays), data points represent the average success rate over every ten rounds. This provides a more holistic and stable sense of performance over time.

When a convention was reached in the presence of extra signals, synonyms almost always emerged. Out of the six runs that established a signaling convention in the 2 × 3 × 2 treatment, synonyms evolved in five. For the 2 × 6 × 2 treatment, all five successful runs evolved synonyms. Unlike in the 2 × 3 × 2 treatment, there are multiple possible distributions of synonyms in the case of six signals: we could have anywhere from one to five signals representing each state in a signaling system. Interestingly, though, the same synonym distribution emerged in all five successful runs of the 2 × 6 × 2 treatment: three signals came to represent one state, two signals came to represent the other, and one signal remained completely unused. We would have undoubtedly seen every possible distribution of synonyms (including a complete absence of synonyms) if we had run many more sessions. That said, it is possible that certain synonym distributions, like the one we observed five times, are more likely to emerge than others. A natural next step would be to estimate a probability distribution over synonym distributions using a variety of learning dynamics.

It is important to be clear regarding the sense in which synonyms are exhibited. Rarely did an individual sender actually utilize multiple signals to represent the same state once a convention was reached. Most of the time, each sender eventually used exactly one signal to represent each state, leaving the other(s) unused. The receivers,

[13] In one run, multiple senders stubbornly clung to state-signal mappings which were at odds with one another. Across 60 rounds of play, their cumulative success rate was only 0.53. In another run, one sender would sometimes use the same signal to represent different states and would sometimes abandon a state-signal association immediately after it had succeeded! If such behaviors were to persist, the associated agents would never reach a set of fully successful conventions.

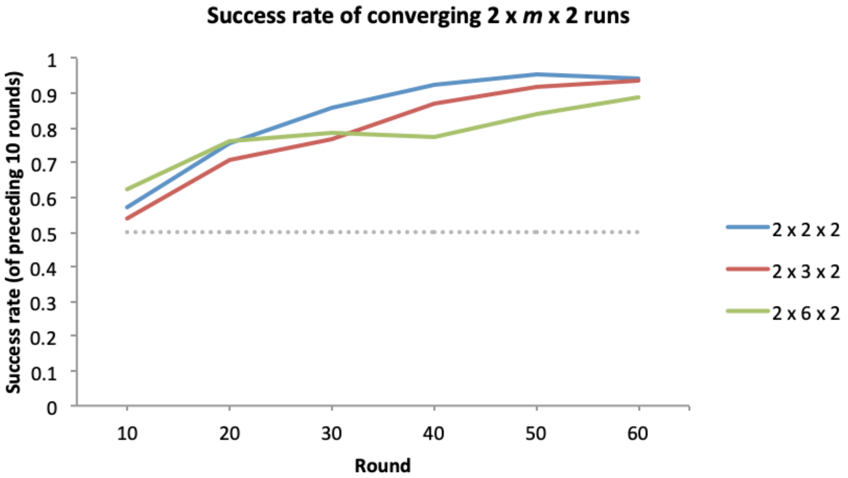


Figure 7: Success rates measured over 10 round periods for the *converging* 2 × 2 × 2, 2 × 3 × 2, 2 × 6 × 2 treatments.

who did not know the identity of their partner each round, might be paired with any one of these senders, each with their own unique vocabulary. From the receiver perspective, their sender opponent each round is then *effectively* mixing over several signals. Indeed, the bulk of the cognitive load in those games with more signals than states, especially the 2 × 6 × 2, is borne by the receivers. To succeed, the receivers must wade through the initially arbitrary flood of signals (one through six), somehow co-establish a convention with senders, and, once a convention is reached, remember what all signals in circulation mean. This can be particularly challenging because certain signals may only be used by one sender and are therefore rarely encountered. We sometimes observed receivers correctly interpret a particular signal when its meaning had been settled but then seem to misremember its meaning several rounds later.

The subjects' behavior regarding synonyms captures an important feature of natural language. An agent may be partial to using certain terms over other equivalent ones, like "pop" over "soda", but still bear the burden of maintaining a mental catalogue of synonyms in order to understand others.

Regarding speed, figures 7 and 8 suggest that having more signals slows progress towards a convention. On average, the 2 × 2 × 2 treatment outperformed 2 × 3 × 2 in terms of speed, which in turn outperformed 2 × 6 × 2.

This is not what one would expect if the subjects were using simple reinforcement learning. As indicated in Figure 2, on simulation, SR learners establish conventions fastest in the experimental set up for two states and two acts when they have four signals to choose from. More to the point here, they do better in the 2 × 3 × 2 than they do in 2 × 2 × 2 game. Note also that when they are successful, human subjects are a full two orders of magnitude faster than simulated SR learners. This is further evidence against modeling the human subjects as gradual reinforcement learners.

Human subjects may have difficulty with games where there are more signals than states because of the increased demands on memory. In the 2 × 6 × 2 game a receiver typically sees a particular signal less often than she would in a 2 × 2 × 2 game. This increases the time between each signal's appearance and the number of signal/act matches each receiver must memorize. This seems to pose a significant challenge for human

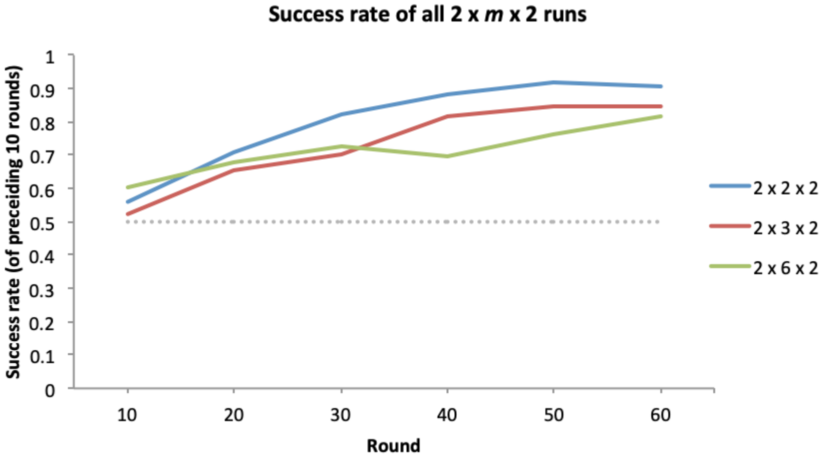


Figure 8: Success rates measured over 10 round periods for *all* $2\times2\times2$, $2\times3\times2$, $2\times6\times2$ treatments.

subjects. In the mid-late game of the $2 \times 6 \times 2$ treatment, receivers were sometimes at a complete loss concerning what to do when they saw lesser-used signals: it was not uncommon to see receivers succeed on a particular signal, only to use a different act when that signal occurred again several rounds later.

In general, the results of recent plays mattered more to the subjects than the results of plays in the more distant past. Whether this is due to memory limitations alone or a combination of factors, the result is that the human subjects are not well-modeled as gradual reinforcement learners. SR learners remember everything, and each of their experiences, regardless of how long ago it occurred, is weighted equally in determining their future actions. We will seek to characterize in more detail how the human subjects learn in the next section, but we will first briefly consider here what happens when there are fewer signals than states.

In contrast with games where there are the same or more signals than states, perfect signaling is impossible in the $3 \times 2 \times 3$ signaling game. Here the best possible expected return is $\frac{2}{3}$. That said, there are numerous signaling conventions that achieve this optimal level of expected return. The sender might use one signal for state 1 with the receiver always doing act 1 when he sees that signal, and the sender might use the second signal for states 2 and 3 with the receiver always doing act 2 when he sees that signal. Such agents would always fail when the state is 3, but with unbiased nature, they would still succeed $\frac{2}{3}$ of the time. Since neither player does better by unilaterally changing their strategies, such a convention is a Nash equilibrium. There are many other similarly successful and stable signaling conventions. See Figure 9 for another example. When paired with each other, SR learners typically evolve a stable signaling convention with an expected success rate of $\frac{2}{3}$.[14]

Importantly, SR learners also do well achieving optimal expected return in the six-sender/six-receiver $3\times2\times3$ game represented in the present experiment. On simulations with six players in each population, each paired randomly with an ideal SR learning partner from the other population in each round, the overall cumulative success rate was nearly optimal at 0.66479 over the course of 10^4 runs with 10^6 plays each. The

[14] See Barrett (2006) for a discussion of this in the one-sender/one-receiver case.

proportion of runs with an average success rate of 0.65 or higher was 0.99. With an average cumulative success rate of 0.6145 after just 10^4 plays, the speed of convergence is roughly comparable with that of simulated SR learners in the six-sender/six-receiver $2 \times m \times 2$ games (as illustrated in Figure 2) remembering that optimal play in the $3 \times 2 \times 3$ game has an expected success rate of $\frac{2}{3}$ rather than 1.

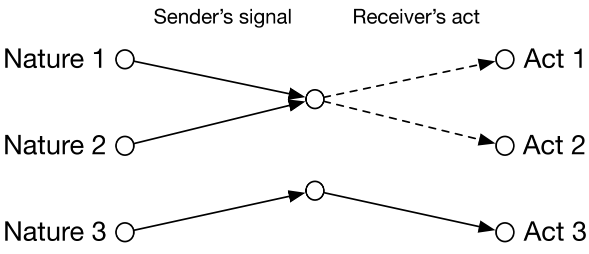


Figure 9: Example of an optimal signaling configuration in a two player $3 \times 2 \times 3$ game. The receiver mixes over acts 1 and 2 upon witnessing signal 1. Success is only guaranteed when the state is 2. The expected payoff is $\frac{2}{3}$, which, while less than 1, is certainly better than the expected payoff from random guessing: $\frac{1}{3}$.

The human subjects, however, struggled in the $3 \times 2 \times 3$ game when compared with their performance in games where they had enough signals. Out of the twelve sessions in which this treatment was featured, only three came anywhere close to reaching optimal conventions. Even the four sessions in which subjects played this treatment for 90 rounds (instead of 60) failed to perform; the extra time yielded no improvements in observed success. In terms of speed this may be significantly better than what one would expect from a gradual reinforcement learner, but it is no where near what the subjects were able to accomplish in the other games.

Of course, having three state and acts increases the difficulty of the game. Bruner et al. (2018) report mixed results from their $3 \times 3 \times 3$ treatments: sometimes agents reach optimal signaling and sometimes they do not. Taking away one of the available signals makes the task yet more difficult. Players in the $3 \times 2 \times 3$ game are more likely to meet with failure for their arbitrary choices in the early game, and it may be unclear how to proceed in the face of these, especially when switching is still sure to result in further failures. In the post-play survey, some subjects admitted to just randomly guessing the whole time ("Group A did not affect my symbols"). That the human subjects had such difficulty with the $3 \times 2 \times 3$ game provides additional evidence against their behaving as SR learners.

In some sessions the $3 \times 2 \times 3$ treatment treatment was first, in some it was second, and in others it was last. The three sessions which *did* converge to maximally efficient signaling were exactly the ones in which subjects (1) encountered the $3 \times 2 \times 3$ last and (2) approached perfect signaling in *both* of their previous treatments. This may be a coincidence. It could also be that these subjects were demonstratively good at these games and so simply performed well in all treatments. Or this may have been an instance of task priming where the subjects' past play influenced their future play. In particular, subjects may have noticed over the course of the first two treatments that a necessary condition for perfect success was an injection from states to signals. This would make them acutely aware that perfect success was not possible in the $3 \times 2 \times 3$ game and that failure on a particular play was not necessarily indicative of a poor strategy. But if success required this sort of sophisticated reasoning by the subjects, then this is again evidence against their being well-modeled as simple reinforcement learners in the context of this game.

The three sessions which did succeed on this difficult treatment also illustrate an important phenomenon in understanding the experiments. While many of the maximally efficient conventions in the $3 \times 2 \times 3$ game require agents to statistically mix over strategies, an *individual* human subject seldom mixed over signals or acts in our experiments (though this was not unseen). Instead, each subject typically played a pure strategy (i.e. mapping each state/signal to a unique signal/act) that was then *effectively* mixed in aggregate over the full population. For example, session 15's convention, though noisy, looked something like Figure 10. All senders mapped nature 2 to signal 1. All senders mapped nature 3 to signal 2. In response to signal 1, though, some senders used signal 1 and some sent signal 2. Although these are pure strategies, they appear mixed to a receiver, whose opponent sometimes uses signal 1 to represent nature 1 and sometimes uses signal 2. The other two sessions which succeeded in the $3 \times 2 \times 3$ treatment evolved a convention most similar to the one depicted in Figure 9 earlier.

In contrast to the behavior of human subjects, *individual* simulated SR agents in the $3 \times 2 \times 3$ game typical play genuine mixed strategies—that is, sometimes a particular sender chooses signal 1 and sometimes signal 2 in response to state 1. This is further evidence against human agents being well-modeled as simple reinforcement learners.

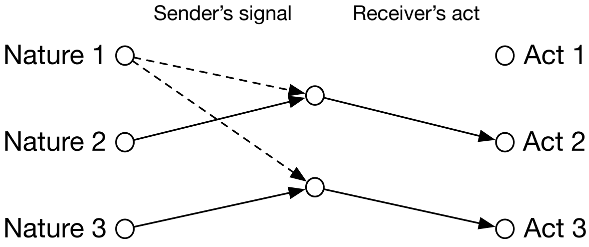


Figure 10: Loosely the signaling convention reached by session 15 in their $3 \times 2 \times 3$ treatment.

4 How human agents learn

While the present experiment does not allow us to directly determine how each participant learns, we can look for persistent patterns in their behavior, then consider what learning rules would explain the patterns we find. One such pattern exhibited by participants was a stronger than expected aversion to changing signals after a period of initial or short-term local success. The sort of aversion to change we saw provides strong evidence against modeling human agents as simple reinforcement learners.

One might naturally expect a degree of aversion to change from a simple reinforcement learner. If an SR receiver sees signal 2, selects act 3, and succeeds on one play, then she will tend to repeat this action the next time she witnesses signal 2. For an SR learner, past successes of actions in a particular context translate to a correspondingly higher likelihood of repeating those actions when that context is encountered in the future. But such gradual reinforcement learning does not mesh well with our observations.

Role (treatment)	Sender (2 x 2 x 2)	Sender (2 x 3 x 2)	Sender (2 x 6 x 2)	Sender (3 x 2 x 3)	Receiver (2 x 2 x 2)	Receiver (2 x 3 x 2)	Receiver (2 x 6 x 2)	Receiver (3 x 2 x 3)
Observed proportion	0.896	0.979	0.833	0.792	0.92	0.792	0.875	0.667
Expected proportion	0.667	0.5	0.2857	0.667	0.667	0.667	0.667	.5
p-value	0.0008***	~0***	~0***	0.0248*	0.0002***	0.0669	0.0022**	0.0047**

Table 1: Observed proportion vs expected proportion (under simple reinforcement) of subjects who did not deviate from the first signal (act) which landed them a success.

To get a sense of how success influences the subjects' future actions, we measured the proportion who repeated the signal/action corresponding to their first success at the next opportunity. For example, if a sender's first success within a treatment occurred when she witnessed state 1 and sent signal 2, what is the probability that she sends signal 2 the next time she sees state 1? If $2 \times 2 \times 2$ subjects reinforced using simple reinforcement, we would expect players to stick with their previous choice $\frac{2}{3}$ of the time. Instead, both senders and receivers stayed loyal to the choice that gave them their first success roughly 90% of the time. In other treatments too, subjects repeat the choice corresponding to their first success far more than one would expect from a simple reinforcement learner with gradually evolving dispositions.

Table 1 depicts the proportion of subjects who do not deviate from the choice that yielded their first success, the expectation for what this proportion would be under simple reinforcement, and p-value the probability that our subjects would have proportions this high if they were using this kind of reinforcement learning. Put simply, the null hypothesis is that players' first reinforcement is done using gradual, one-ball-at-a-time reinforcement. More specifically, in an $n \times m \times n$ signaling game, let X be the state (signal) witnessed by a sender (receiver) on the first round in which that player succeeds and Y be the signal (act) chosen by that player in that succeeding round. The null hypothesis then is as follows: Immediately following this first success, that player updates her dispositions in accordance with simple reinforcement (as we described it). The next time she witnesses X, she will choose Y with probability $\frac{2}{m+1}$ if she is a sender and $\frac{2}{n+1}$ if she is a receiver. For almost all treatments, participants resisted changing signals after their first success much more than simple reinforcement players would have (in expectation).

Of course, one might model this sort of behavior in the context of a modified version of reinforcement learning by making the magnitude of the reinforcement on success very large in comparison to the strength of the subjects' initial dispositions (represented in the SR learning model as the initial urn contents). This would explain the subjects' strong and quickly-forged commitments to their early-game successful decisions. However, the subjects typically abandon their previously successful signal or act strategy if it fails even just a few times in a row regardless of the earlier history of success. One might model this by opting for a form of reinforcement learning with high rewards for success but some significant punishment or forgetting to help erase the triumphs of the past in the face of recent failures. In short, one can go a significant way in simulating the observed behavior of the subjects by supplementing basic SR learning with a variety of mechanisms designed to explain why recent plays have a much greater effect on the current behaviour of subjects than the results of earlier plays.[15]

[15] Blume et al.'s (2002) observed their subjects to exhibit this sort of recency bias as well, and the authors attributed it to forgetting. See Barrett and Zollman (2009) for an account of forms of reinforcement

That said, a close examination of the data from the present experiment suggests a more straightforward learning model: *win-stay/lose shift with inertia* (WSLSwI). WSLSwI is a variant of win-stay/lose shift (WSLS), a very simple learning rule that works just as it sounds.[16] Each sender (receiver) has a mapping from states (signals) to signals (acts). On each play, every agent obeys her mapping for the given stimulus. If she succeeds, her mapping does not change (win-stay). On failure, she maps the stimulus she just observed to a randomly selected new signal (act) (lose-shift).

While the dispositions of agents using SR are determined by their full history of success and failure, WSLS is very forgetful. On WSLS the agents' current strategy for each state and signal type are determined by what happened the last time they saw the state and signal type. This makes WSLS learners flexible and quick. While SR learners slowly converge to optimal signaling behavior on the $2 \times 2 \times 2$ game and may not converge to optimal signaling at all on an $n \times n \times n$ game, under WSLS, agents often achieve perfect signaling on a finite number of plays. In the case of the $2 \times 2 \times 2$ game, it may only take a handful of plays.

There already exists some evidence that various sorts of inertia can improve WSLS's performance and successfully model human decision making. Worthy and Maddox (2014) build on Estes' (1950) learning equations to develop $WSLS_{Learning}$, a variant of WSLS in which players do not immediately abandon their current behavior upon failure nor remain committed to it upon success. Instead, players update their probability of staying on a success and switching on a failure over the course of many trials. Longer strings of success on one action make agents more committed to those actions and less likely to leave them on failure. In the same paper, Worthy and Maddox present a hybrid learning dynamics in which the probability of taking some action is a weighted average of its likelihood under SR and $WSLS_{Learning}$. These novel dynamics yielded a closer fit to human behavior in the decision making experiments they considered than the baseline WSLS model. While these dynamics do not utilize exactly the same sort of inertia used in WSLSwI, they are both modifications of WSLS which prevent agents from immediately abandoning strategies upon one failure. Similarly, in a simulation setting, Barrett et al. (2017) investigate "WSLS with reinforcement" on signaling games, a dynamics in which agents shift (upon failure) to acts which have reaped more benefits in the past. WSLSwI then is yet another dynamics with a mechanism that dampens the often erratic behavior of WSLS. In this case, the suggestion is that human agents may stick to a strategy even after a loss simply because they tend to abandon a previously successful action only if it repeatedly fails.

As on WSLS, but not basic SR, our subjects were seldom observed to change signals (acts) for a given stimulus after a success. But contrary to WSLS, they do not always change after just one failure. Sometimes it takes several, sequential, unsuccessful applications of the same signal (act) before they switch. That is, players display different levels of *inertia* in their willingness to switch to a different signal (act): hence, WSLSwI. This sort of resistance to changing one's ways seems behaviourally plausible, especially if the current mapping has been in place a long time or has wrought many successes.

learning that features varieties of forgetting.

[16] WSLS was initially proposed and applied to bandit-problems (a class of decision theory problems) by Robbins (1952). WSLS was later introduced in a game theoretic context by Nowak and Sigmund (1993). Barrett and Zollman (2009) show that win-stay/lose randomize (WSLR), a similar but somewhat stronger rule, will guide agents to a signaling system in a Lewis-Skyrms signaling game with probability 1.

WSLSwI is slightly less forgetful than WSLS. On WSLSwI players are equipped with an an inertia level i at any given time. Each player's i may be distinct and may change throughout the game, perhaps as a function of play history. When an agent witnesses state (signal) v and selects the signal (act) w associated with v for a failure, their *failure count* for that stimulus increases by 1. On a success, the failure count is reset to 0. When a player's failure count reaches i, she changes her mapping so that v maps to a signal (act) other than w.

One of the virtues of WSLSwI is that it provides a natural role for higher-order reasoning in the case of more sophisticated agents. In particular, an agent might help direct exploration of the strategy space by implementing deterministic or probabilistic constraints on the choice of a new signal (act) when switching on failure. On the basic WSLSwI dynamics, the choice of a new strategy on failure is entirely unconstrained reflecting no higher-order considerations. The human subjects in the present experiment, however, typically choose the new signal (act) in a way that is not uniformly random. For instance, in the $2 \times 3 \times 2$ and $2 \times 6 \times 2$ treatments, a sender tends to choose a new signal that she *is not already using.* That is, senders tend to keep their mappings injective when they switch. Similarly for receivers in the $3 \times 2 \times 3$ treatment. They tend to choose in such a way that different signals map to different acts. Along similar lines, in the $2 \times 2 \times 2$ and $3 \times 2 \times 3$ treatments, senders would sometimes perform a signal "swap": state v now maps to a new signal, and another state now maps to w if no others do. That is, senders switched in such a way that their resulting mappings were surjective; otherwise, the sender would be mapping all states to the same signal. Similar for receivers in the $2 \times 2 \times 2$, $2 \times 3 \times 2$, and $2 \times 6 \times 2$ treatments. In other words, subjects tended to shift in such a way as to not preclude optimal signaling.[17]

WSLSwI has at least two highly desirable properties that WSLS lacks. Although often very fast in establishing signaling conventions, WSLS is unstable. When a convention is reached, if either player makes a mistake, this leads to failure, which leads to a shift, which likely leads to further failures, unraveling the players' hard-won convention (Barrett and Zollman, 2009).[18] WSLSwI is much more stable than WSLS at equilibrium. Once a convention is reached, if a mistake and resulting failure occurs on a single play (or more with a high enough inertia), a WSLSwI player will resist changing her mapping, preserving the equilibrium. In short, inertia provides a mechanism whereby players might tolerate random mistakes and maintain coordination.

Another weakness of WSLS is that on the simple $2 \times 2 \times 2$ signaling game it may not converge to a signaling system due to a "revolving door" problem where WSLS players get caught in an infinite loop where they shift simultaneously and continually miscoordinate. WSLSwI players, especially when they have different inertias, however, will typically not encounter this problem.[19]

WSLSwI also has several advantages over simple reinforcement learning. While not as flexible and quick as WSLS, WSLSwI is much faster than SR. In particular, it is

[17] Of course, one would expect lower cognition species not to implement such sophisticated constraints. Indeed, in many cases, even the human subjects we observed seemed to shift strategies randomly with no manifest constraints.

[18] Evolved dispositions under WSLS are fragile, unlike those of simple reinforcement learners. Barrett et al. (2017) discuss the competing virtues of speed and stability in the WSLS class of dynamics and give results on a hybrid learning dynamics *win-stay/lose shift with reinforcement.* This dynamics addresses the stability issue but is substantially more complicated and memory-demanding than WSLSwI. See also Cochran and Barrett (2020).

[19] See Cochran and Barrett (2020) for an investigation of the formal properties of WSLSwI.

fast enough to explain the rapid convergence of human subjects to signaling conventions observed in the present experiment.[20] Also, WSLSwI typically converges to a pure convention in a finite time while the strategies of SR agents are always mixed at finite times. This also helps to explain the behavior of human subjects who exhibit near optimal behavior after a relatively short time, particularly in the simpler games. Finally, unlike SR, WSLSwI is not prone to get stuck in suboptimal pooling equilibria in signaling games with more than two state-act pairs.[21] The present experiments would have to be extended to more complicated games to determine the degree to which this accords with human behavior.

Returning to the results of the present experiments, Figure 11 displays the success rates of players in a 2 × 2 × 2 population game with six senders and six receivers. The blue line depicts the average success rates of our converging human subject runs every 10 plays. The red and yellow lines represent the progress of simulated WSLSwI learners (with inertia 2) and SR learners, respectively. Measuring the mean absolute deviation between our subjects' performance and those of the simulated agents reveals that WSLSwI is a better than eight times closer fit than SR on this measure. Variants of SR with higher reinforcement rates and/or punishment may better match human subject performance than basic SR as such tunings inevitably inherent some properties of WSLSwI—namely, quick convergence to success and flexible abandonment of failing strategies. Given this, the claim is just that WSLSwI better represents human leaning in this context than gradual SR learning.

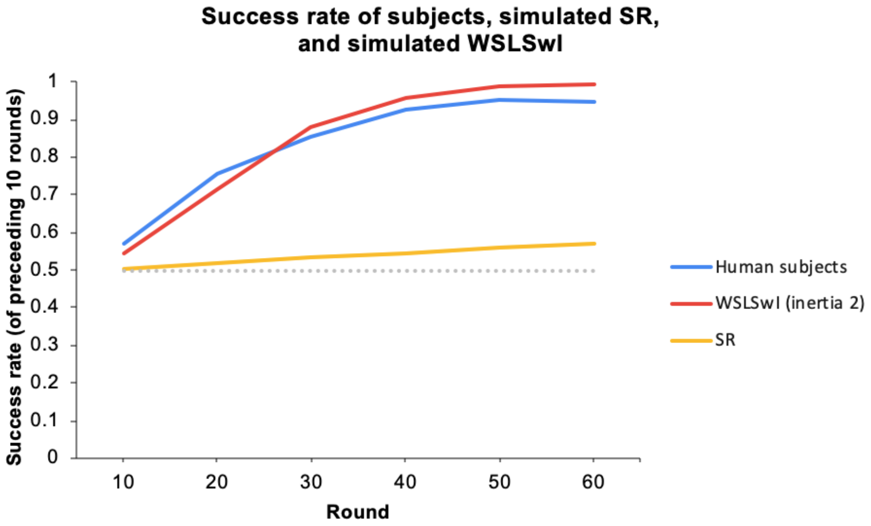


Figure 11: Success rates measured over 10 round periods for the converging 2 × 2 × 2 human subject treatments, simulated WSLSwI learners with inertia 2, and simulated simple reinforcement learners.

The most direct way to estimate a subject's inertia is simply to count how many consecutive failures we observe before she switches to another signal (act). We call this the subject's *observed inertia* at that moment. For example, suppose that Figure 12 (left) represents six consecutive rounds for a WSLSwI sender in a 2 × 2 × 2 game with states

[20] See Cochran and Barrett (2020).

[21] See Barrett (2006). Hofbauer and Huttegger (2008) and Cochran and Barrett (2020).

$\{Blue, Green\}$ and signals $\{1, 2\}$. In the first round, the sender acts on her mapping which sends state Blue to signal 1. This results in success. The next two times she sees Blue, though, she sends signal 1 but does not succeed. Finally, on the fourth time she sees Blue she chooses 2, revealing that she has changed her mapping so that Blue maps to 2 now. Since it took two failures for this change to happen, this is an observed inertia of 2. Note that the round in which the state was Green does not count towards the inertia for Blue. We might (and should) similarly count the inertia for Green. Figure 13 (left) gives a frequency histogram of observed inertia across all subjects and for all states (signals) they might witness. Note that most subjects will switch signals as a result of failures more than once; these subjects have more than one recorded instance of observed inertia and their observed inertia may be different each time (even within the same treatment). The graph on the left was recorded by plotting each of these instances for all players.

State	Signal	Act
Blue	1	Blue
Blue	1	Green
Green	2	Green
Blue	1	Green
Blue	2	Green

State	Signal	Act
Blue	1	Blue
Blue	1	Green
Green	2	Green
Blue	1	Green
Blue	1	Blue

Figure 12: Observed (left) and minimum (right) inertia example.

Figure 13 (left) gives the impression that most subjects have an inertia of 1; that is, they are playing regular WSLS. But this is not the whole story. Suppose that a sender has inertia 5 for a certain state at some point. In order for us to actually observe this inertia of 5, we would need this sender to witness the appropriate state five times and to fail each time sequentially before switching. If five failures do not happen, we cannot witness their observed inertia, meaning Figure 13 (left) is biased towards lower inertias.

To provide a more complete picture, we also record the *minimum inertia* for each player. Suppose a sender experiences the sequence of plays depicted in Figure 12 (right). She starts by sending signal 1 when the state is Blue. The next two times she sees Blue, she sends 1 and fails. On the 4th occasion that she sees Blue, she still sends 1 and succeeds. While we cannot directly ascertain what this sender's inertia was at this point, we know that it was at least 3. If her inertia had been, say 2, she would have switched to signal 2 in the last round that she saw Blue. We therefore say that this sender has a *minimum* inertia of 3 here. Figure 13 (right) gives the distribution of minimum inertia.

While minimum inertia helps fill in some of the gaps left by observed inertia, the information it does provide is not very precise because it is simply a lower bound for actual inertia. If a receiver exhibits a minimum inertia of 3 at some point, this means that her actual inertia could be 3, 4, 5, 6, or higher. This information is not worthless, though. Inspecting both graphs in Figure 13 together reveals that, if subjects were using WSLSwI, the majority of them had inertia higher than 1. As a reminder, there are not actually two different kinds of inertia, observed and minimum. Rather, there is only standard inertia; observed and minimum inertia are just tools we use to estimate it.

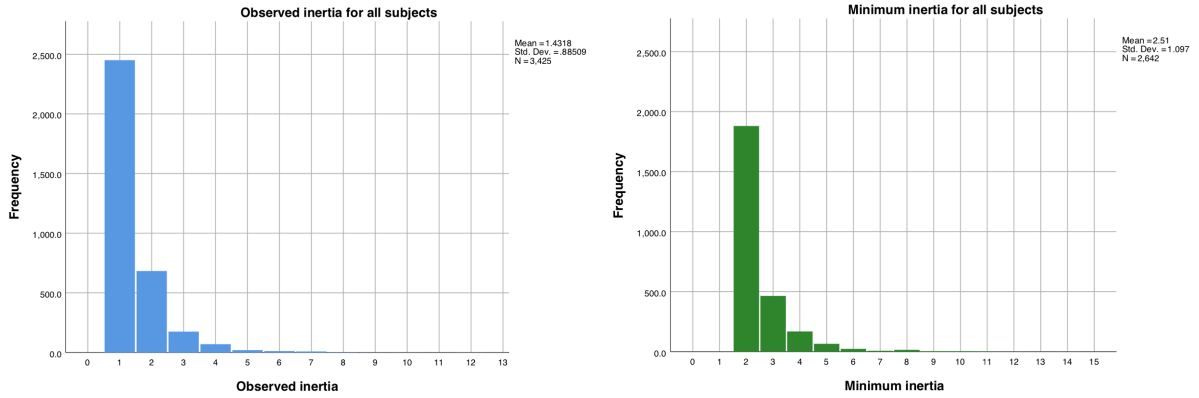


Figure 13: Frequency for observed inertia (left) and for minimum inertia (right) for all subjects.

We also considered how subjects' inertias might depend on their roles, the game they are playing, and how long they have been playing it. We found a statistically significant ($p < 0.01$) difference between senders' and receivers' average observed inertia levels: senders have a slightly higher observed inertia than receivers. Figure 14 displays relative frequencies[22] for observed inertia levels of senders (dark) and receivers (light) on the left; on the right are the minimum inertia levels for the two roles. Though the difference is subtle, we see that senders on average have higher observed and minimum inertia (though remember, this last one is noisy). This may have to do with broader human behavior. Since players are rewarded for the receiver's choice, not the sender's, receivers may feel guiltier about round failures than senders and feel more pressured to switch if incorrect. Additionally, since they move first, senders may feel more empowered to set and enforce conventions i.e. if a mistake is made then the receivers should change, not them. Indeed, in the post experiment survey one sender confessed that this (choose a mapping and let the receivers learn it) was her strategy, and three senders were observed to never deviate from their signal choices at the beginning of the game. No receivers did this with their initially chosen acts.

Table 2 provides descriptive statistics for observed inertia across all treatments. While the means for all four groups are close, there is a statistically significant difference between treatment $2 \times 6 \times 2$ and the other two treatments featuring two states of nature ($p < 0.01$). 2 gives the relative frequency distributions for $2 \times 6 \times 2$ against $2 \times 2 \times 2$; the former treatment has lower inertia. Again, we can only speculate as to why this might be. It is possible that players in the $2 \times 6 \times 2$ treatment somehow feel more pressure to explore their wealth of signal options when they start failing. The grass is always greener on the other signals.

We also found that subjects typically exhibit different levels of inertia within the same treatment. A significant parameter here is the length of time they have been playing the game. Figure 16 plots average observed inertia for all players in the $2 \times 3 \times 2$ treatment against round number. In brief, subjects tend to have higher observed inertia in later rounds. This may be because they tend to be succeeding more in later rounds and are

[22] We use relative frequencies here instead of raw frequencies because the number of sender and receiver data points within each graph is not the same, making the two groups hard to compare at a glance in a raw frequency histogram. By relative frequency of, say, observed inertia, we mean that, given a sender/receiver's inertia is observed, what is the probability that it is a 1, 2, 3, etc.

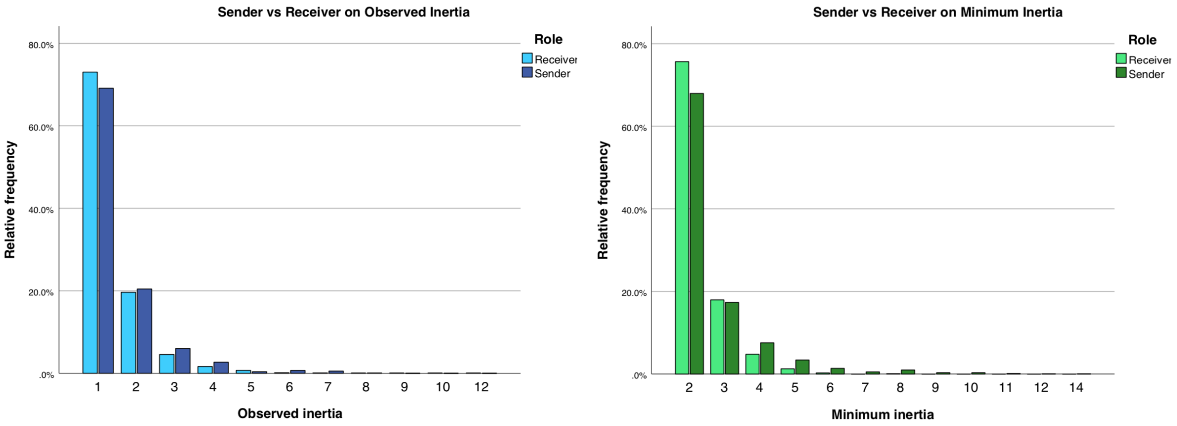
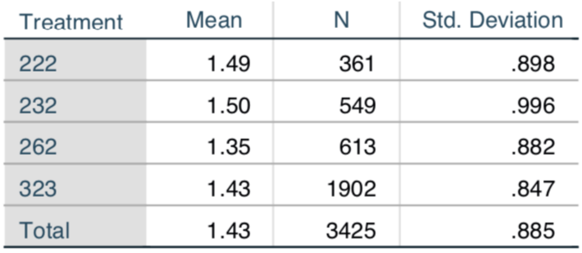


Figure 14: A relative frequency comparison of senders' and receivers' observed inertia (left) and minimum inertia (right).



Treatment	Mean	N	Std. Deviation
222	1.49	361	.898
232	1.50	549	.996
262	1.35	613	.882
323	1.43	1902	.847
Total	1.43	3425	.885

Table 2: Descriptive statistics for observed inertia across all treatments.

harder to dislodge from their mappings. Another possibility is that over time subjects tire of the cognitive work involved in shifting their mappings.

5 Discussion

Win-stay/lose-shift with inertia (WSLSwI) explains a number of the behaviors of the human subjects in the present experiment. In general, it captures subjects' tendency to count their most recent experience more than earlier experience. More specifically, it explains why subjects tend to stick with a strategy that led to success until that strategy fails. And it explains why subjects will readily give up a strategy after a sequence of failures even in cases where it has delivered significant past success.

Just as important, WSLSwI explains how human subject are able to reach stable conventions orders of magnitude faster than one would expect from gradual reinforcement. While gradual reinforcement has virtues for investigating stable features of the world, WSLSwI is better-suited to investigating more transient features of the world since it can glom onto patterns quickly, then quickly shift when those patterns no longer hold. This makes it in many ways ideal for forging conventions in the context of the shifting behaviors of other agents.

The speed of WSLSwI, however, comes at a cost. Inasmuch as there are no strategies that deliver uniform success in the context of the $3 \times 2 \times 3$ game, there are no stable strategies for it to glom onto. So WSLSwI also explains why subjects find this game

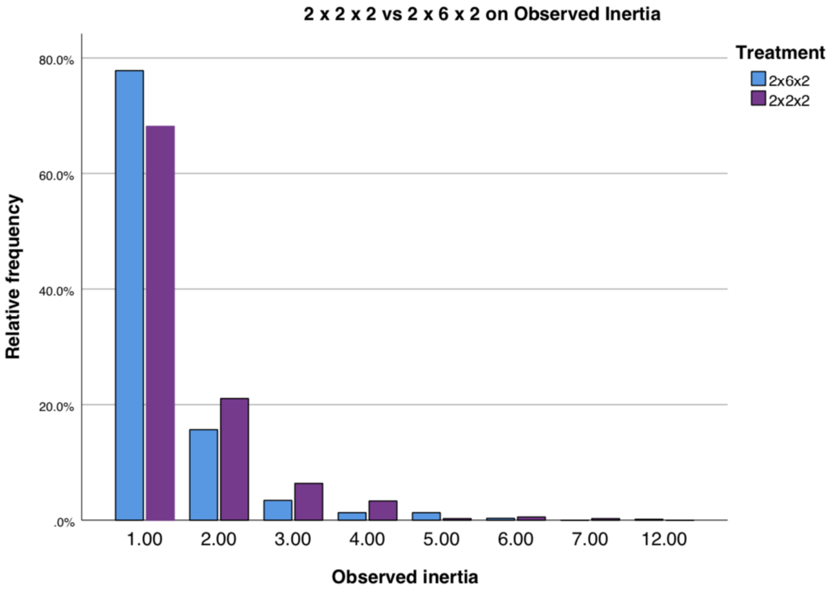


Figure 15: Comparison of observed inertia relative frequencies for treatments featuring 2 states with 2 signals versus with 6 signals.

so difficult—as the dynamics predicts, they often end up shifting between strategies randomly with no sustained success.[23]

We also found that when subjects had more signals than they needed to represent states, they were still usually able to approach a signaling convention, albeit slower than they would have with exactly enough signals. WSLSwI explains this too. The extra signals provide more options for agents to try and still miscoordinate while the delays produced by inertia slow the process of finding a successful convention.

In addition, WSLSwI provides a degree of freedom for explaining some of the relatively sophisticated strategies that subject sometimes exhibited. In particular, when they shifted on failure, subjects sometimes constrained their choice to always allow for the possibility of the result representing an optimal signaling system given their current experience. Senders employ this sort of higher-order rationality when they avoid signals that they are using for other states in choosing a new signal for the current state, and receivers employ it when they avoid mapping different signals to the same act. That it allows agents to impose higher-order constraints on the exploration of potential strategies makes WSLSwI a flexible tool for modeling trial-and-error learning. It also makes it a very powerful learning dynamics as implemented by subjects—one that is supremely well-suited to establishing conventions in the context of population-based Lewis-Skyrms signaling games where there are enough terms to represent the state-act pairs.

WSLSwI provides yet another degree of freedom in that it allows for subjects to apply higher-order considerations: namely, they may tune their level of inertia over time to reflect their pragmatic aims. Once one finds an optimal convention or sense that one is close, one may wish to maintain the progress one has made and hence might increase the level of inertia to provide additional stability while still allowing for the possibility that one will want to evolve new dispositions if the community changes how it behaves.

[23] On the other hand, outside of the laboratory, where agents are often able to invent new signal types when needed, one might expect this feature of WSLSwI to be unproblematic.

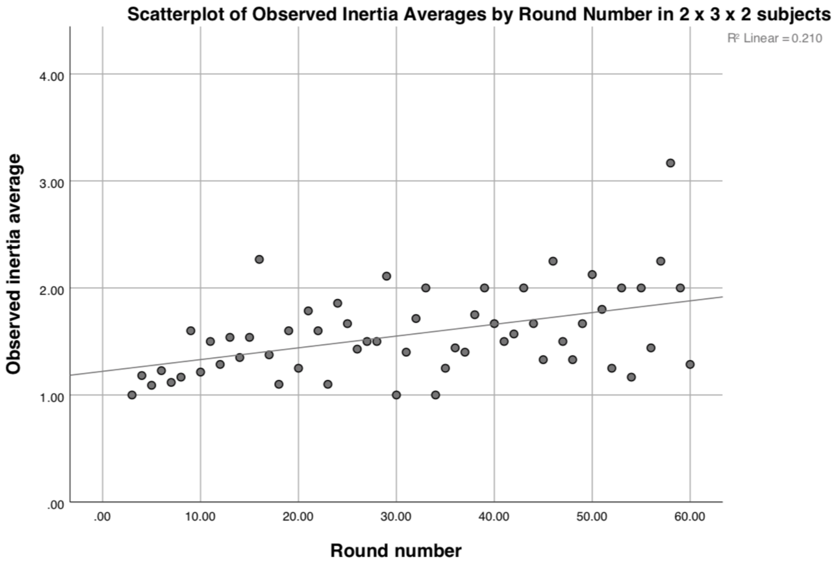


Figure 16: Scatter plot of mean observed inertia as predicted by round number for $2\times3\times2$ subjects.

And, again, subjects tended to exhibit this sort of behavior.

In short, WSLSwI provides a strong low-rationality learning model that explains how human subjects are able to establish conventions in a broad assortment of Lewis-Skyrms signaling games, and it explains why they sometimes don't. And, along the way, it meshes well with the details of the subjects' observed behavior.

The present experiment suggests a number of directions for future research. To begin, we do not know how convergence speed for human subjects scales with a signaling game's complexity in, say, $n\times n\times n$, games where $n > 2$ nor do we know whether and the extent to which subjects encounter the sort of sub-optimal partial pooling predicted by simple reinforcement learning. If they do not encounter significant partial pooling or if they are still able to establish conventions more quickly than predicted by SR in complex games, this might provide additional evidence against SR and in favor of something with the speed and flexibility of WSLSwI. One should also want to investigate whether agents utilize WSLSwI when repeatedly paired with the same partner (as opposed to a random member of a larger community) as it may well be that agents exhibit less (or no) inertia when the uncertainty about their partner is removed. .

Another direction for research would involve focusing on the variety of ways that human subjects learn in signaling games. WSLSwI describes the behavior of some subjects well and captures the coarse-grained behavior of the subjects in aggregate, but for all of its virtues, it fails to describe the behavior of all of the human subjects in the present experiment. Some subjects never deviated from their initial strategies. At least one subject seemed to be utilizing a form of probe and adjust learning. And others used strategies that we have been unable to clearly identify. Further, within subjects whose behavior was compatible with WSLSwI there was notable heterogeneity in the details. Some had virtually no inertia while others demonstrated a great deal. And for some the level of inertia evolved significantly over the course of a game treatment. In light of this, further experimental work might classify subjects based on the sort of learning dynamics they exhibit.

One of the morals here is that experiments on subjects playing various Lewis-Skyrms signaling games can be especially relevant to our understanding of learning. Learning a convention in the context of shifting behaviors of others in the community poses a compelling problem. And there is an epistemic purity to the problem—the lack of focal points and the complete symmetry among equilibria in such games mean that learning from past experience is the only path players have to reach a successful convention.

Bibliography

Argiento, Raffaele, Robin Pemantle, Brian Skyrms, and Stanislav Volkov (2009). "Learning to Signal: Analysis of a Micro-level Reinforcement Model." *Social Dynamics*, 225–249.

Barrett, Jeffrey (2006). "Numerical Simulations of the Lewis Signaling Game: Learning Strategies, Pooling Equilibria, and the Evolution of Grammar." *UC Irvine: Institute for Mathematical Behavioral Sciences Technical Report.*

Barrett, Jeffrey and Kevin Zollman (2009). "The role of forgetting in the evolution and learning of language." *Journal of Experimental and Theoretical Artificial Intelligence*, 293–309.

Barrett, Jeffrey A., Calvin T. Cochran, Simon Huttegger, and Naoki Fujiwara (2017). "Hybrid learning in signalling games." *Journal of Experimental and Theoretical Artificial Intelligence*, *29*(5), 1119–1127.

Blume, Andreas, Douglas DeJong, Yong-Gwan Kim, and Geoffrey Sprinkle (1998). "Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games." *The American Economic Review*, *88*(5), 1323–1340.

Blume, Andreas, Douglas V. Dejong, Yong-Gwan Kim, and Geoffrey B. Sprinkle (2001). "Evolution of Communication with Partial Common Interest." *Games and Economic Behavior*, *37*(1), 79–120.

Blume, Andreas, Douglas V. Dejong, George R. Neumann, and N. E. Savin (2002). "Learning and communication in sender-receiver games: an econometric investigation." *Journal of Applied Econometrics*, *17*(3), 225–247.

Bruner, Justin, Cailin O'Connor, Hannah Rubin, and Simon M. Huttegger (2018). "David Lewis in the lab: experimental results on the emergence of meaning." *Synthese*, *195*(2), 603–621.

Chen, Daniel L., Martin Schonger, and Chris Wickens (2016). "oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments." *SSRN Electronic Journal.*

Cochran, Calvin T. and Jeffrey A. Barrett (2020). "The efficacy of human learning in Lewis-Skyrms signaling games." *manuscript.*

Erev, I. and A.Ẽ. Roth (1998). "Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria." *American Economic Review*, *88*, 848–881.

Estes, William K. (1950). “Toward a statistical theory of learning..” *Psychological Review*, *57*(2), 94–107.

Herrnstein, R.J̃. (1970). “On the law of effect.” *Journal of the Experimental Analysis of Behavior*, *13*, 243–266.

Hofbauer, Josef and Simon M. Huttegger (2008). “Feasibility of communication in binary signaling games.” *Journal of Theoretical Biology*, *254*(4), 843–849.

Huttegger, S., B. Skyrms, P. Tarres, and E. Wagner (2014). “Some dynamics of signaling games.” *Proceedings of the National Academy of Sciences*, *111*(Supplement 3), 10873–10880.

Lacroix, Travis (2018). “On salience and signaling in sender–receiver games: partial pooling, learning, and focal points.” *Synthese.*

Lewis, David Kellog (1969). *Convention: a philosophical study.* Harvard University Press.

Nowak, Martin and Karl Sigmund (1993). “A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoners Dilemma game.” *Nature*, *364*(6432), 56–58.

Robbins, H. (1952). “Some aspects of the sequential design of experiments.” *Bulletin of the American Mathematical Society*, *58*, 527–535.

Rubin, Hannah, Cailin O’Connor, and Justin Bruner (2019). “Experimental Economics for Philosophers.” *Methodological Advances in Experimental Philosophy.*

Schelling, Thomas C. (1960). *Strategy of Conflict.* Harvard University Press.

Skyrms, Brian (2010). *Signals: Evolution, Learning, and Information.* Oxford University Press.

Skyrms, Brian (2014). *Evolution of the social contract.* Cambridge University Press.

Worthy, Darrell A. and W. Todd Maddox (2014). “A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W.K. Estes.” *Journal of Mathematical Psychology*, *59*, 41–49.