THE CORRECTION GAME OR, HOW PRE-EVOLVED COMMUNICATIVE DISPOSITIONS MIGHT AFFECT COMMUNICATIVE DISPOSITIONS

TRAVIS LACROIX

Abstract. How might pre-evolved communicative dispositions affect how individuals learn to communicate in a novel context? I present a model of learning that varies the reward for coordination in the signalling game framework under simple reinforcement learning as a function of the agents' actions. The model takes advantage of a type of modular compositional communicative bootstrapping by which the sender and receiver use pre-evolved communicative dispositions—a "yes/no" command—to evolve new dispositions.

Keywords — signalling games, reinforcement learning, modular composition, invention, cue-reading, evolution of language negation

Introduction

The signalling game was introduced by Lewis (1969) to explain how communication conventions might arise without having an antecedent language in place to agree upon the convention.[1] We suppose that there are two players, called the sender and receiver. Nature picks a state of the world at random. The sender observes the state of the world and chooses a signal from a set of arbitrary (in the sense of lacking pre-existing salience) signals. The receiver observes the signal, though not the state, and chooses from a set of possible actions. Payoffs are jointly determined by state and the action.

The signalling game has been modified using evolutionary dynamics (Skyrms, 1996, 2010). A simple model, which we will concern ourselves with in this paper, is Roth-Erev reinforcement learning (Roth and Erev, 1995; Erev and Roth, 1998), which has a long psychological pedigree and is a standard learning model in behavioural psychology.[2] Under this dynamic, the probability of an actor's choosing a particular action is proportional to the accumulated reward for that action.[3] Consider an atomic 2×2 signalling game—where there are 2 equiprobable states (s_0

Department of Logic and Philosophy of Science, University of California, Irvine
Mila, (Québec AI Institute / Institut Québécois d'Intelligence Artificielle)
E-mail address: tlacroix@uci.edu.
Date: Draft of April 2019. *PLEASE CITE PUBLISHED VERSION, IF AVAILABLE.*

[1]This is given a more general form in Crawford and Sobel (1982).

[2]See, for example, Bush and Mosteller (1955); Suppes and Atkinson (1960); Arthur (1993); Börgers and Sarin (1997, 2000); Bruner et al. (2018). For an overview of reinforcement learning from a computational perspective, see Kaelbling et al. (1996); Sutton and Barto (1998)

[3]This model is based on the *matching law* (Herrnstein, 1970), which in turn was a formalisation of *law of effect* (Thorndike, 1905, 1911, 1927).

and s_1), 2 signals (m_0 and m_1), 2 actions (a_0 and a_1), and the reward for coordination is given by $u(s_i, a_j) = \delta_{ij}$.[4] I will sometimes refer to this as an *atomic* 2*-game*, and in general, for an arbitrary number of signals and state-act pairs, an *atomic* n*-game*.[5]

This can be illustrated simply by an urn-learning procedure. We assume that the sender has an urn for each of the two states—labelled s_0 and s_1. Similarly, the receiver has an urn for each of the two messages (signals)—labelled m_0 and m_1. At the outset of the game, each of the sender's urns is equipped with one ball for each possible message at her disposal—m_0 and m_1. Similarly, each of the receiver's urns contains a ball for each of her possible actions—a_0 and a_1.

On each play of the game, the state of the world is chosen at random. The sender then chooses a ball at random from the urn corresponding to the state of the world and sends that message to the receiver. The receiver then chooses a ball at random from the urn corresponding to the message received. If the action matches the state of the world, then the sender and the receiver both reinforce their behaviour by returning the ball to the urn from which it was chosen, and additionally adding another ball of the same type to the urn from which the original ball was chosen. If the action does not match the state, then each player simply returns the drawn ball to the urn from which it was drawn.[6] The game is then repeated.

A sender strategy is a function mapping states to a probability distribution over signals; a receiver strategy is a map from signals to acts. The dynamic shifts strategies to the extent that adding balls to an urn for a successful action shifts the relative probability of picking a ball of that type on a future play of the game. Adding balls to a particular urn changes the conditional probabilities of the sender's signals (conditional on the state) and the receiver's acts (conditional on the signal). Thus, the players are more likely to perform actions that were previously successful. In this special case, the sender and receiver will coordinate upon a *signalling system*—a pair of sender and receiver strategies whose composition maximises the possible expected payoff—with probability 1 (Argiento et al., 2009).

However, sender and receiver strategies can also be *pooling*. In complete pooling equilibria, the sender sends signals with probabilities that are independent of the state. Similarly, the receiver chooses actions independent of the signal received. Thus, the probability of an action is independent of the state of nature. When there are more than two possible states, signals, and actions, there are also *partial* pooling equilibria where some, but not all, of the states are pooled.[7] When $n > 2$, these partial-pooling equilibria have a basin of attraction of positive measure (Pawlowitsch, 2008; Hofbauer and Huttegger, 2008).[8]

---

[4] δ_{ij} is the Kronecker delta, defined as

$$u(s_i, a_j) = \delta_{ij}, \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

.

[5] This terminology is due to Steinert-Threlkeld (2016).

[6] In the most basic case, there is no penalty for miscoordination, though it is possible to model punishment by discarding a ball when it led to a failure.

[7] Note that in general, there are $n!$ possible signalling systems in the $n \times n$ signalling game (Lewis, 1969), but there are n^{2n} possible combinations of strategies (Huttegger, 2007a). Thus, the number of possible strategies quickly outpaces the number of possible signalling systems as n increases.

[8] Note, this result is under a slightly different dynamic, called the *replicator dynamic* (Taylor and Jonker, 1978). However, there is a close relation between the replicator dynamic and the reinforcement learning dynamic presented here: Beggs (2005) and Hopkins and Posch (2005)

Partial pooling is undesirable to the extent that it is inefficient. It is, perhaps, further undesirable insofar as the signalling game framework is supposed to give a plausible evolutionary explanation of how communication might actually have arisen in nature, rather than a *just so* story. Partial pooling can be avoided by, e.g., altering the underlying dynamic; however, there is a trade-off here between the explanatory power of a model and the results obtained by the model.[9] Thus, we have a twofold desideratum: to avoid partial-pooling while maintaining empirical sensitivity to the potential precursors of linguistic communication.

Communication is a unique evolutionary process in the following sense: once a group of individuals has learned some set of simple communication conventions, those learned behaviours may be used to influence future communication, giving rise to a feedback loop. This results from the *reflexivity* of natural language: Languages can be used to communicate *about* languages (Hockett, 1963). When faced with a novel context, an evolutionary agent can always evolve a brand new disposition from scratch. However, might the sender and receiver in a signalling context not take advantage of pre-evolved dispositions in order to help them learn to signal more efficiently? In this paper, I present a model of learning that varies the reward for coordination in the signalling game as a function of the agents' actions. The model takes advantage of a type of compositional communicative bootstrapping by which they use pre-evolved communicative dispositions to learn new dispositions.

We might note that coordination for the purpose of communication is generally *goal-directed*. This need not be understood in terms of something as high-level as, e.g., *intentions*. For example, in the *cue-reading game*—a modification of the signalling game where the sender has a fixed set of dispositions which the receiver must learn to interpret (in the sense of "react to") in the appropriate way (Barrett and Skyrms, 2017)—the receiver's *goal* is to interpret the sender's cues correctly; the sender, on the other hand, is static—she does not have a goal in the way the receiver does, but reacts in a fixed way to the states of nature, regardless of whether the receiver interprets her actions correctly.

However, might the sender in the cue-reading game not also have a goal—namely, for the receiver to *understand* her fixed signalling disposition. How might she achieve this goal? If she can communicate *that* her signal *means* such-and-such, then this would help the receiver toward the goal of interpreting the signal appropriately. However, this would be putting the cart before the horse, so to speak: the entire premise of the signalling [cue-reading] game is that we do not presuppose the sender is able to communicate the meaning of her signal [cue]; rather, it is precisely the "intended" meaning, based on the sender's signalling disposition, that the receiver must learn. In the signalling game, the meaning of the signal co-evolves as a function of *both* the sender's and receiver's respective dispositions. If the sender could communicate her disposition, then the sender and receiver would have already arrived at a signalling convention.

Suppose that the sender and receiver have already evolved a signalling system in some *other* context. Might the sender not then use *those* communicative capacities to try to express to the receiver what her meaning is in the new context? This does not presuppose that the sender and receiver have already solidified the meanings

show that the mean-field dynamics of Roth-Erev learning is a version of the replicator dynamic. See also, Barrett (2006).

[9] See, for example, D'Arms et al. (1998), and the further discussion in LaCroix (2018a).

of the signals in the *main* context; rather, when the receiver fails to perform the action that the sender *wants* her to perform (i.e., the one that is appropriate for the particular state), the sender may be able to communicate *that* the receiver did something wrong.[10]

The following story makes more clear the sort of phenomena that I have in mind here. Consider two actors in a signalling context. Suppose they have already evolved up some rudimentary communicative capacity. For example, they may have learned how to communicate some simple command for an action. This might be interpreted, at least for the purposes of this story, as a command that represents some holophrastic binary distinction—e.g., "stop/go", "yes/no", "correct/incorrect", "true/false", "right/wrong", etc.[11]

In such a context, there are two relevant states of the world, with corresponding appropriate actions, and there are two possible signals to represent these state-act pairs. One or the other signalling system will evolve with certainty, given that this is a 2×2 signalling game. Now, if we imagine this sort of communicative context has already evolved, it stands to reason that individuals in a new signalling context (where no dispositions have yet evolved) might learn to take advantage of their previously evolved communicative convention in the following sense. Suppose the sender and receiver are in a novel signalling context, where they must evolve dispositions from scratch. In the normal signalling game model, they may learn to coordinate upon a signalling convention simply by trying things and reinforcing those actions that led to a success.

However, since (*ex hypothesi*) they have already evolved a communicative disposition to communicate *that* an action is appropriate or not, they already have at their disposal a signalling game which takes a correct or incorrect action as input and outputs a signal that states that the action was correct or incorrect: when an agent performs an incorrect action, the "state of the world" is such that it would be appropriate for the sender to send the "no/stop/wrong" signal, which the receiver will appropriately interpret—since she already understands *this* signal. Thus, miscoordination in our new context is an appropriate input for the pre-evolved context. This is exactly the notion of modular composition discussed in Barrett and Skyrms (2017). Therefore, on the presupposition that the actors have already evolved such a capacity, they need only to be able to compose the two separate games into a single game in order to communicate that, e.g., *corrective* action should take place.

Below, I present several variations of a base model, which I will call the *correction game*, that are built on this intuitive story, and analyse the results of this ability to take advantage of a previously evolved disposition. In particular, I compare learning rates and occurrence of suboptimal partial-pooling equilibria with the atomic signalling game of the same dimension, where the individuals do not take

[10] There is a lot of *intentional* talk in this paragraph; however, I take this to be harmless for the reasons given by Dennett (1971, 1987); according to Dennett (1971), presupposing beliefs and desires on the part of such an agent—one who is not rational, *per se*—is a form of "conceptually innocent anthropomorphizing" (93). It should be fairly clear that I am not presupposing, as a matter of fact, that a sender or receiver in the signalling game have any human-level cognitive capacities.

[11] Note that equally a command, or imperative, in a signalling system can be interpreted as an indicative statement. We will not worry about this distinction too much here, but see the discussion in Harms (2004a,b); Millikan (2005); Huttegger (2007b); Zollman (2011).

advantage of a previously evolved disposition, to show whether and in what ways this is advantageous to the players.

1. The Correction Game Model

Before getting into the details of how the model works, we might consider the following "high-level" interpretation of what is going on here. Suppose two agents want to communicate. We model this with a signalling game of some arbitrary dimension, depending upon the case under consideration. The sender sends a signal to the receiver in an attempt to transfer information about the state of the world. Suppose the receiver performs an action that is inappropriate for the state under consideration. In the normal signalling game model, they move on and a new state of the world is chosen for a new attempt. However, suppose with some probability the sender tries to correct the action. Note that this, in a sense, presupposes that the sender "knows" what the correct action is; however, this is not problematic because if the sender were the one attempting the action, she has perfect information about the state of the world and so, even if she does not know *a priori* what the correct action is, she could hit upon the correct action quickly via some simple trial-and-error experimentation—the purpose of the signalling game model is to show how such state-act pairs might become associated with signals, thus giving rise to information transfer. So, rather than moving on to a new round, with some probability the sender will attempt to take advantage of a previously evolved communicative capacity for the purpose of "correcting" the inappropriate action of the receiver.

We will start by supposing that the agents in the signalling game have already evolved up some command capacity, which we will take to be analogous, in some respect (i.e., the intended outcome action of the command), to "stop/go" or "right/wrong" or "yes/no", etc.

This model is built upon the base of a normal atomic n-game, discussed in the introduction. The correction game proceeds as the atomic n-game normally does: nature picks a state of the world without bias, the sender chooses a signal at random, and the receiver chooses an action at random. If they coordinate, they receive payoff 1, and shift their dispositions proportional to their accumulated rewards. However, the correction game diverges from the atomic n-game when the sender and receiver miscoordinate. When the actors fail to coordinate, the sender attempts to "correct" the action in question, with some probability, μ. Namely, with probability μ, the agents take advantage of the previously evolved capacity to direct actions via some command—i.e., the sender takes the failure as input for the sub-game, and sends the pre-evolved signal corresponding to "wrong". With probability $(1 - \mu)$, they simply move on to the next play of the game, as they normally would, with payoff 0.

Thus, the reward is 1 for one-shot coordination, and if the actors fail and abandon their failure (probability $(1 - \mu)$), then the reward is 0. This "segment" of the correction game is just the normal signalling game procedure with payoff 1 for coordination and no punishment for miscoordination. The main difference between the correction game and the signalling game is that there is a chance (μ) that the receiver attempts a new action, with the state and the signal remaining fixed. This is under the assumption that, in light of the failure, the sender sends the additional signal that the receiver has done something "incorrect", as it were. (Note that if the underlying command is already evolved to a signalling system, then we can assume the sender and receiver always coordinate on *this* signal—i.e., the receiver knows how to react to the additional command, by, e.g., trying something new, since this is a pre-evolved disposition.) This extra command from the sender gives the receiver complete knowledge that the particular action she chose was incorrect for the state; however, she still lacks complete knowledge about *which* of the remaining actions *is* appropriate for the state. Thus, this set-up does not presuppose anything about the meanings of the signals being evolved in the main game, nor the sender's ability to communicate these meanings.

There are several possible ways of modelling this process. I will suggest the following. For each run, we will take the reward on the first "cycle" to be the usual reward for coordination: $u(s, a)$. If the sender tries to correct the receiver's action, due to miscoordination, then the sender and receiver will get some discounted reward conditional upon coordination. This will be given by a discount factor, $\gamma \in [0, 1]$. This discount might be understood as decreasing marginal utility for the additional cost of having to play an extra game—i.e., taking the time to try to correct the receiver's action. For $\gamma = 1$, we have cost-less correction. For $\gamma = 0$, this extended game reduces to the normal atomic signalling game. Thus, the full specification of the reward is given by $R_{t_n} = \gamma^n \cdot u(s, a)$—the reward, R, on the nth cycle, t_n.

To make clear what I have in mind here, consider the following possible play. Suppose we have a discount factor, $\gamma = \frac{1}{2}$, and a base-payoff, $u(s, a) = 1$, for coordination. The signalling game begins as normal. Nature picks a state of the world, the sender picks a signal, the receiver picks an action. If they coordinate, then they both receive a payoff of 1, and they move on to the next play. If they miscoordinate, then with probability $(1 - \mu)$, they receive a payoff of 0, and they move on to the next play. However, with probability μ, if they miscoordinate, then they play a "correction game", which can be understood as the sender utilising a previously evolved capacity to inform the receiver that she did something wrong. Here, we assume they always coordinate on the correction game, since it is a pre-evolved 2×2 game, so the receiver tries a new action—namely, if A is the set of actions available to the receiver, and if action $a_i \in A$ led to a miscoordination, the receiver samples stochastically from the set $A - \{a_i\}$, with an associated re-normalised probability distribution $\Delta(A - \{a_i\})$. Suppose the action chosen on the first repetition is a_j.

If the sender and receiver coordinate on the first repetition, then they both receive a discounted payoff $R_{t_1} = \gamma^1 \cdot u(s, a) = \frac{1}{2}$. If they miscoordinate, then, again, with probability $(1 - \mu)$, they abandon the attempt to coordinate, receive payoff 0, and move to the next play. Again, with probability μ, the sender tries to correct the receiver a second time. The receiver tries a new action, sampled from the set $A - \{a_i, a_j\}$, with an associated probability distribution $\Delta(A - \{a_i, a_j\})$. If coordination occurs on the second retry, the sender and receiver get a discounted payoff of $R_{t_2} = \gamma^2 \cdot u(s, a) = \frac{1}{4}$. This continues with the general discounted reward being given by $R_{t_n} = \gamma^n \cdot u(s, a)$ for n attempts to correct the action. See Figure 1. Note that the sender strategy (and the state of nature) are fixed during the correction component of the game; only the receiver tries to correct her action.

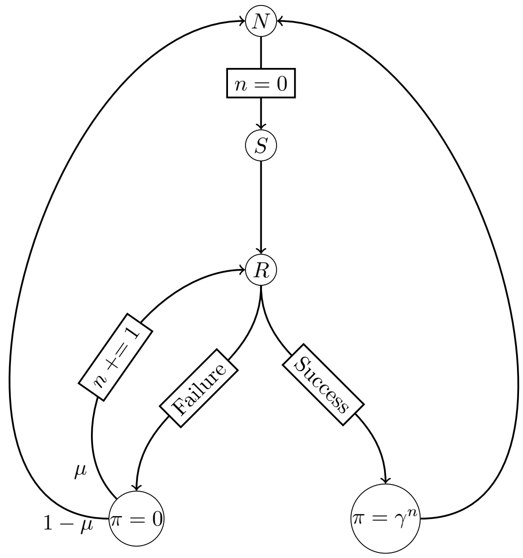

FIGURE 1. Basic *correction game*, with $\pi = R_{t_n} = \gamma^n u(s,a)$, and $u(s,a) = 1$. N denotes nature, S denotes the sender, and R denotes the receiver. $\mu \in [0,1]$ is a probability. n is the "counter" that is used to discount the rewards.

2. The Simple Correction Game: Cue-Reading

We begin by examining a correction game where the sender's dispositions are already fixed. This is, in effect, a cue-reading game, with the possibility for the sender to attempt to correct the receiver's action when her chosen action fails to achieve coordination on that particular state.

We examine an 8×8 cue-reading game. States are equiprobable, and the payoff for success is $u(s,a) = 1$. The sender begins with dispositions such that

$$P(m_i|s_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

Each run consists in 500 individual plays of the game.[12] We examine the results of 1000 runs.[13]

We have two new parameters that can be varied. First, the probability, μ, with which the sender and receiver repeat a failed play; second, the cost for repetition, γ. To get a reasonable picture of how these parameters affect learning in the underlying

[12]Note that this is an extremely low number of plays, but individuals learn quickly under reinforcement learning when the sender's dispositions are already fixed. In an 8×8 cue-reading game, after 10,000 plays, the sender and receiver have a cumulative success rate greater than 0.95 on almost all (0.975) of the runs, and every run results in a cumulative success rate greater than 0.90. As such, signalling systems are guaranteed in a fairly short amount of time in this particular case—thus, we examine shorter-run results to see whether we cannot arrive at signalling even faster with correction. The question of partial pooling is less of a concern here.

[13]The simulations were run in Python 2.7, and the resultant data was compiled using MatLab. Source code for the correction game can be found here: [INSERT GITHUB LINK].

cue-reading game, we examine the 16 combinations of $\mu = [0.25, 0.50, 0.75, 1.00]$ and $\gamma = [0.25, 0.50, 0.75, 1.00]$.

2.1. **Results.** We must be careful in interpreting the results of our simulations. What is common is to calculate the *cumulative success rate* of a particular run by simply counting the number of plays where the sender and receiver successfully coordinated and dividing this by the total number of plays. Early failures get washed out as the number of plays per run increases. We can then examine the *proportion* of runs that have a cumulative success rate surpassing some threshold.

However, the threshold for success is not arbitrary. The 8×8 cue-reading game has a large number of partial pooling equilibria. These are polymorphic traps where the sender and receiver might get caught. The most efficient sub-optimal strategy for the reciever (given the sender's dispositions are fixed in the cue-reading game) occurs when she performs the appropriate action for 7/8 of the signals, and pools her strategy on the 8th signal. These pooling equilibria allow for a maximum communicative success rate (and a maximum expected payoff) of 0.875. Thus, we ought to set our threshold for success at 0.875 to see whether the sender and receiver have escaped these polymorphic traps.

However, we note that when $mu = 1$, the receiver will necessarily retry actions until she hits upon a successful action. Thus, we should expect that for $\mu = 1$, the sender and receiver will *always* surpass the threshold for success. Indeed, this is exactly what happens (after 1000 plays per run), as shown in Table 1. More complete data are shown, for comparison, in Figure 2.

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic	**0.000**			
Repetition	0.25	0.000	0.000	0.000	0.000
Probability	0.50	0.003	0.021	0.082	0.214
μ	0.75	0.954	0.999	0.999	1.000
	1.00	1.000	1.000	1.000	1.000

TABLE 1. Proportion of successes for short-run simulation results for correction game with pre-evolved sender dispositions (cue-reading) under a variety of discount factors and repetition probabilities (10^3 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

However, one might worry that the cumulative success rate is not accurately capturing successes in the cue-reading game with correction since 10^3 plays is really $10^3 + C$ plays, where C is the number of repeat attempts at success which take place on a given run. The average number of repetitions in each case is shown in Table 2. This further highlights the effects of cost-less correction—because the accumulated rewards are shifted more for lower-cost correction, the likelihood of choosing the correct action in a future play is increased more than when correction is expensive. (For example, a reduction in the number of repetitions when the sender is guaranteed to try to correct the receiver's behaviour ($\mu = 1.00$) implies that the sender and receiver are failing to coordinate less often; here, we see a

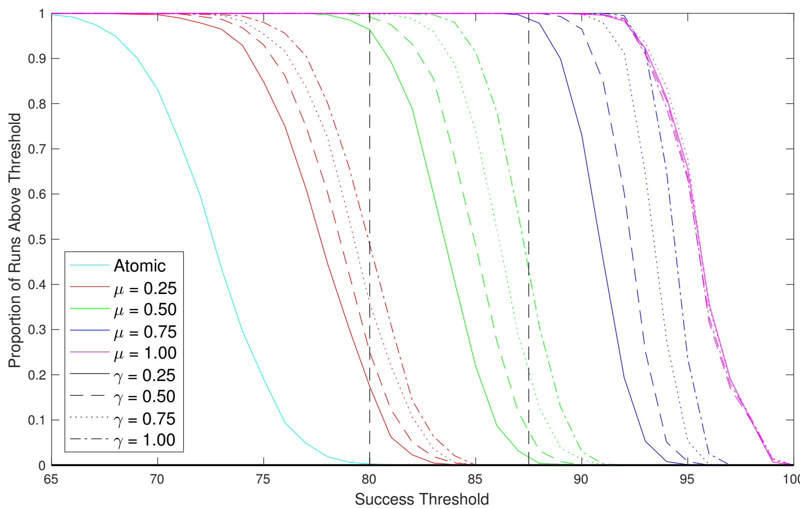

FIGURE 2. The proportion of successful runs above thresholds $[0.65, 1.00]$ shown for each combination of parameters, $[\mu, \gamma]$. The vertical dashed lines indicate the thresholds 0.80, and 0.875, respectively

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0**		
Repetition	0.25	78	76	73	70
Probability	0.50	176	161	149	138
μ	0.75	303	264	228	199
	1.00	484	386	307	247

TABLE 2. Average number of repetitions made in the cue-reading correction game

monotonic decrease in repetitions as the cost for repetition decreases) In the worst case, we see almost a 50% increase in "plays".

There is perhaps good reason not to interpret our data this conservatively: for one, a repeat does not actually constitute a full play of the game to the extent that no new state nor signal is chosen during a repeat. Even so, we can correct for this in the following way.

The simulations were re-run, and new data was gathered thus. The sender and receiver are allowed 500 plays to try to learn a signalling convention. We then let them communicate according to whatever convention they have settled upon (or begun to settle upon) for 1000 plays. We count successes and failures during the communication period only, not during the learning period—thus, we ignore the failures that occur during learning. This approximates the expectation of success in the same way as looking at the urn contents after 500 rounds and calculating the exact expectation. Since the strategies that evolve are going to vary stochastically, we take an average of 1000 runs. The adjusted success rates for 1000 runs under this success measure are shown in Table 3. The data vary significantly from those

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.000**		
Repetition	0.25	0.006	0.017	0.040	0.091
Probability	0.50	0.017	0.089	0.371	0.701
μ	0.75	0.045	0.359	0.874	0.991
	1.00	0.113	0.752	0.995	1.000

TABLE 3. Adjusted proportion of successes for short-run simulation results for correction game with pre-evolved sender dispositions (cue-reading) under a variety of discount factors and repetition probabilities (5×10^2 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

of Table 1. In particular, there is less pooling across all discount factors when the correction probability $\mu = 0.25, 0.50$, whereas more pooling (than the data in Table 1) appears to be exhibited for more probable repetitions, $\mu = 0.75, 1.00$. However, we should note three things here: first, there are half as many plays where learning occurs (500 as opposed to 1000); second, initial miscoordination during learning is not counted in the latter case—thus, we should expect slightly more successes than if initial failures are counted; finally, since the sender and receiver do not take advantage of the correction capacity during the communication period, successes here really do constitute successes.

The general qualitative results still hold: fixing the discount factor, γ, an increase in μ corresponds to an increase in success; fixing the repetition probability, μ, a decrease in cost for repetition also corresponds to an increase in success. Thus, these results are robust regardless of what one counts as a success. Again, more complete data are shown, for comparison, in Figure 3. Note that the successes for the atomic case are shifted up, since we are not counting the initial failures during the learning period. The data is less clearly differentiated; however, again, holding fixed one parameter, we see a monotonic increase in successes as we vary the other parameter (either holding fixed the cost and increasing the probability of correction, or holding fixed the probability and decreasing the cost).

We might note that this cue-reading game with correction is really only half of the model suggested by the story at the outset. Correction (modelled as a pre-evolved "yes/no" meta-game) only occurs when the receiver fails to coordinate with the sender's intended meaning. This is the "no" component: the receiver is corrected with some probability only when she does something wrong. We might add the "yes" half of the correction as follows: suppose the receiver coordinates on the first try; with some probability, μ, the sender further reinforces this behaviour by telling the receiver that she did something *right*, by using their pre-evolved disposition. Thus, the receiver receives an additional payoff, given by $\gamma \cdot u(s, a)$. Note that there

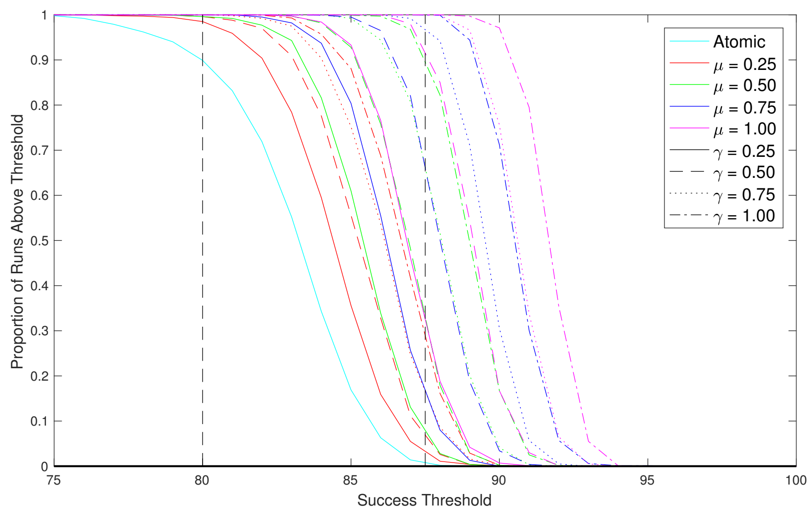

FIGURE 3. The adjusted proportion of successful runs above thresholds $[0.75, 1.00]$ shown for each combination of parameters, $[\mu, \gamma]$. The vertical dashed lines indicate the thresholds 0.80, and 0.875, respectively

is no repetition when the action is successful, so there is a one-shot reinforcement, which occurs with probability μ.[14]

In this case, even under the worst parameter combinations, every combination of parameters resulted in 100% of the runs exceeding the pooling-threshold cumulative success rate of 0.875 after only 500 plays. Indeed, most combinations do significantly better than this. The proportion of runs resulting in a cumulative success rate greater than 0.95 are shown in Table 4.[15]

This model was built upon a cue-reading game, rather than a signalling game; thus, the sender's dispositions were fixed at the outset. Can the sender and receiver co-evolve their strategies, while taking advantage of their pre-evolved corrective dispositions?

3. The Simple Correction Game: Signalling

In this section, we examine the effects of combining the possibility for correction with the full signalling game, as opposed to the asymmetric cue-reading game. However, some care is required here. If the sender and receiver begin correcting too early, then correction will obviously not help them to evolve a signalling convention, because the sender would effectively be trying to correct the receiver's behaviour

[14]We might imagine that the normal payoff for coordination is given by nature, as is the case in the atomic signalling game, whereas this additional payoff is given by the sender; however, it need not be the sender who tries to correct the receiver's behaviour—see the discussion in Section 5 below.

[15]Note the increase in the success threshold; every run results in 100% of the plays having a cumulative success rate greater than 0.9 (and so greater than 0.875). No pooling whatsoever occurs after 500 plays.

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic	**0.000**			
Repetition	0.25	0.670	0.792	0.877	0.933
Probability	0.50	0.820	0.950	0.981	0.998
μ	0.75	0.883	0.987	0.999	1.000
	1.00	0.947	0.994	1.000	1.000

Table 4. Adjusted proportion of successes for short-run simulation results for correction game ("yes" *and* "no") with pre-evolved sender dispositions (cue-reading) under a variety of discount factors and repetition probabilities (5×10^2 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.95

while she is still not yet fixed upon what her signal actually means. Early on, she might use m_0 to mean s_0, and correct the receiver when she chooses an act other than a_0. However, later in the game she might use m_0 to mean s_1. The propensities are highly variable at the outset. Thus, we must allow the sender and receiver to start to learn a signalling convention before they can utilise the correction game. Unfortunately, the answer to the previous question is decidedly: *no.* The efficiency seen in the cue-reading game with correction does not generalise to the signalling game, so the results here are limitative. Even so, what happens is somewhat subtle, so it is worth going through with some care.

3.1. **Results.** Here we examine the short-term results for a simple correction game built on top of a full atomic 8-game, under a variety of parameters. States are equiprobable, and the payoff for success is $u(s, a) = 1$. Each run consists in 10^5 individual plays of the game, and we examine the results of 1000 runs. The sender and receiver are allowed a learning period of $25,000$ plays prior to trying to correct behaviour using their pre-evolved dispositions. Again, we examine the correction game with 16 combinations of $\mu = [0.25, 0.50, 0.75, 1.00]$ and $\gamma = [0.25, 0.50, 0.75, 1.00]$. Again, when *either* $\mu = 0$ or $\gamma = 0$, the correction game is equivalent to the atomic signalling game. In the former case, the probability of repetition is 0, so the sender and receiver never retry. In the latter case, any number of repetitions results in a payoff of 0, so even if the sender and receiver repeat until a success, they do not reinforce on that success.

The cumulative success rates, with a threshold of 0.875 for success, of these several parameters are shown in Table 5. In general, it appears that correction helps the sender and receiver to learn a signalling convention; however, this is again under the assumption that a "success" is just coordination on a given play, ignoring the repetitions. Thus, when the repetition probability is 1 the proportion of successes is going to trivially be 1—the sender and receiver repeat a failure until it turns into a success. We can obtain more accurate results of whether the sender and receiver are avoiding pooling by examining their success during a communication period, after an initial learning period.

Successes are re-calculated as follows: The sender and receiver have an initial learning period of $25,000$ plays where they learn atomically. They learn for the rest of the 10^5 plays by using correction. Finally, we count successes during a

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.331**		
Repetition	0.25	0.413	0.432	0.406	0.428
Probability	0.50	0.765	0.812	0.805	0.765
μ	0.75	0.917	0.914	0.901	0.892
	1.00	1.000	1.000	1.000	1.000

TABLE 5. Proportion of successes for short-run simulation results for correction game under a variety of discount factors and repetition probabilities (10^5 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

1000-play "communication period", which approximates the actual *expectation* of success; results of 1000 runs are examined. These adjusted data are displayed in Table 6 Note, first and foremost, that the success in the atomic case are increased.

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.548**		
Repetition	0.25	0.468	0.409	0.341	0.260
Probability	0.50	0.385	0.249	0.164	0.078
μ	0.75	0.349	0.183	0.037	0.000
	1.00	0.279	0.076	0.000	0.000

TABLE 6. Adjusted proportion of successes for short-run simulation results for correction game with co-evolved sender dispositions (signalling) under a variety of discount factors and repetition probabilities (10^5 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

This is because we are not counting the failures during the initial learning period. In comparison to the atomic game under this success measure, the correction game does categorically worse. In a way, however, this makes sense. The sender is correcting behaviour without herself knowing what a signal is supposed to mean. Thus, correction is too aggressive. Note also that the correction game here performs worse when the cost for payoff is decreased. This is the opposite of what happens in the cue-reading game. Again, this is because cost-less correction has a larger effect on propensities, which, we have now seen, is detrimental when the sender's disposition is not yet fixed. This is further highlighted by the fact that the number of repetitions *increases* as the cost of repeating goes down in the full signalling game. The average number of repetitions in each case are shown in Table 7.

We are obtaining a clearer picture of how and when correction, in the form of a pre-evolved disposition, might positively affect learning a new disposition. In the cue-reading game, the sender is determined that the signal means such-and-such, so correction is appropriate. In the signalling game, she is also learning a conventional

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0**		
Repetition	0.25	2136	2158	2199	2154
Probability	0.50	4608	4459	4572	4731
μ	0.75	6889	6924	7223	7534
	1.00	9409	10044	10317	12239

Table 7. Average number of repetitions made in the signalling correction game.

meaning for her signals, so it makes little sense for her to insist very early on that the receiver has done something wrong. This is further highlighted by the fact that, when we include the "yes" component, so that the sender reinforces correct behaviour on the receiver's part, the results are even than those in Table 6. This is because the correcting behaviour on the full correction game is even more aggressive than the behaviour on the correction game with only the "no" component.

However, these results are more subtle than just that they fail to help avoid pooling. In particular, in spite of the fact that the sender and receiver end up pooling their strategies more often when the sender is too aggressive, the *expected payoff* remains largely unchanged in every case. These data are shown in Table 8 The variance between the expected payoff between these 1000 runs is effectively

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.8982**		
Repetition	0.25	0.8975	0.8961	0.8926	0.8944
Probability	0.50	0.8910	0.8927	0.8890	0.8844
μ	0.75	0.8936	0.8909	0.8850	0.8780
	1.00	0.8947	0.8849	0.8797	0.8587

Table 8. Average expected payoff for short-run simulation results for correction game under a variety of discount factors and repetition probabilities (10^5 plays per run, 1000 runs).

equivalent—approximately 0.005—in every case. Indeed, when the sender and receiver escape pooling equilibria the correction game does no worse than the atomic signalling game. However, they tend to get caught in pooling more often the more aggressive the sender is in trying to correct the receiver's behaviour.

If we decrease the initial period in which the sender and receiver learn atomically, they do even worse still. This is because their dispositions are even less fixed than when they start out with a period of atomic learning. If we increase the initial period where they learn atomically, then, as the period of atomic learning approaches the total number of plays, the results limit toward the atomic results. Thus, it is not possible that the sender and receiver do better than the atomic case when they have a pre-evolved corrective disposition at their disposal. The best they can do is as good as the atomic case.

4. The Simple Correction Game: Signalling With Invention

The reason why the results of the general signalling game were limitative, it was suggested, is because the sender is unwarranted in attempting to correct the receiver's behaviour: she is also learning what the signals mean, and so it makes little sense for her to insist upon a particular meaning at the outset when the meanings of the signals are still fluctuating. In this section we examine the general signalling game *with invention*, which is a modified version of the Hoppe-Pólya urn model (Hoppe, 1984) of *neutral* evolution—where many mutations do not convey a selective advantage.[16]

In the atomic case, the signalling game with invention works in this way. Suppose we have 8 states of nature and 8 appropriate actions. The sender begins with no signals; she has 8 urns for each of the states, and each urn contains one black ball—the *mutator*. The receiver begins with no urns. On a particular round, nature picks a state of the world with some probability—again, we assume nature is unbiased so each state is equiprobable. When the receiver sees the state, and selects a ball at random from the corresponding urn. If she selects the black ball, she invents a new signal, by placing a ball for that signal in the urn. This is the signal that she sends to the receiver. The receiver is attentive to new signals: when the signal sent is novel, she creates a new urn for that signal, containing 8 balls for each of the possible actions and then selects an act from that urn. When the sender and receiver coordinate, they reinforce by adding another ball of the same type to the urn from which it was chosen. The game is then repeated with a new state of nature. This is a Hoppe-Pólya urn model with *differential* reinforcement.

Note that the sender never reinforces her propensity to invent, so the rate at which the sender invents new signals decreases over time. Thus, when a state of nature is seen for the first time, the sender invents a signal to communicate with the receiver. If they coordinate, then there is a 2/3 probability in the future that the same signal will be sent in that state, and 1/3 probability that a brand new signal will be invented in that state. If they miscoordinate, then there is a 1/2 probability that the sender will retry the same signal in that state, and 1/2 probability that the sender will send a new signal in that state.[17]

4.1. Results.

Here we examine the short-term results for the full correction game built on top of a full atomic 8-game, with invention, under a variety of parameters. The sender begins with no signals. States are equiprobable, and the payoff for success is $u(s, a) = 1$. Each run consists in 1.5×10^4 individual plays of the game, and we examine the results of 1000 runs. The sender and receiver we begin with no atomic learning period, since this was the worst-case in the general signalling game with correction. Again, we examine the correction game with 16 combinations of $\mu = [0.25, 0.50, 0.75, 1.00]$ and $\gamma = [0.25, 0.50, 0.75, 1.00]$, as compared with the atomic signalling game with invention.

The sender and receiver learn signalling dispositions with invention over the course of 1.5×10^4 individual plays, and then, to gain a more accurate representation of what counts as a success, they communicate for 1000 plays, and we calculate the average number of successes over the course of the communication period, where the

[16] See also the discussion in Skyrms (2010); Alexander et al. (2012).

[17] This is related (though due to the differential reinforcement not equivalent) to the *Chinese Restaurant Process*; see, e.g., (Aldous, 1985; Pitman, 1995).

threshold for success is 0.875. These results are shown in Table 9. As opposed to the

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic	**0.071**			
Repetition	0.25	0.166	0.378	0.528	0.664
Probability	0.50	0.316	0.551	0.547	0.409
μ	0.75	0.446	0.518	0.025	0.004
	1.00	0.495	0.313	0.007	0.000

Table 9. Adjusted proportion of successes for short-run simulation results for correction game ("yes" *and* "no") with co-evolved sender dispositions (signalling) plus invention under a variety of discount factors and repetition probabilities (1.5×10^4 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

atomic signalling game, where the sender's unjustified aggressiveness in correcting the receiver is detrimental to them both, we see that correction again helps the sender and receiver to coordinate, when the correction is not too often and not too inexpensive. In the case where the sender always tries to correct the receiver, and correction is cost-free, they do worse. Again, this should be unsurprising, given that the sender and receiver are still learning to coordinate. Thus, the sender's being too aggressive is still detrimental to them both; however, in almost every other case, correction has a significant impact on learning to signal. More complete data are displayed in Figure 4. We look more closely at the results that are beneficial in Figure 5, centred about the success threshold. Correction almost always helps when the sender invents new signals.

However, it is also known that inventing new signals can help to avoid pooling equilibria in general (Alexander et al., 2012). There are further subtleties to the signalling game with invention; we can look at the average number of signals invented in each case to see how efficiently the sender is inventing while the sender and receiver are learning. The average number of signals invented in each case are shown in Table 10. As is evident, correction not only helps the sender and receiver

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic	**89**, (100.00%)			
Repetition	0.25	83, (93.25%)	80, (89.89%)	74, (83.15%)	69, (77.53%)
Probability	0.50	78, (87.64%)	69, (77.53%)	62, (69.66%)	57, (64.04%)
μ	0.75	70, (78.65%)	59, (66.29%)	47, (52.81%)	38, (42.70%)
	1.00	72, (80.90%)	58, (65.17%)	48, (53.92%)	38, (42.70%)

Table 10. Average number of signals at the end of 1.5×10^4 plays of the signalling game with invention across a variety of parameters, and comparison with atomic case

to avoid pooling equilibria, it helps them to do so more efficiently—i.e., by creating fewer signals at the outset. In the best case in terms of proportion of successes

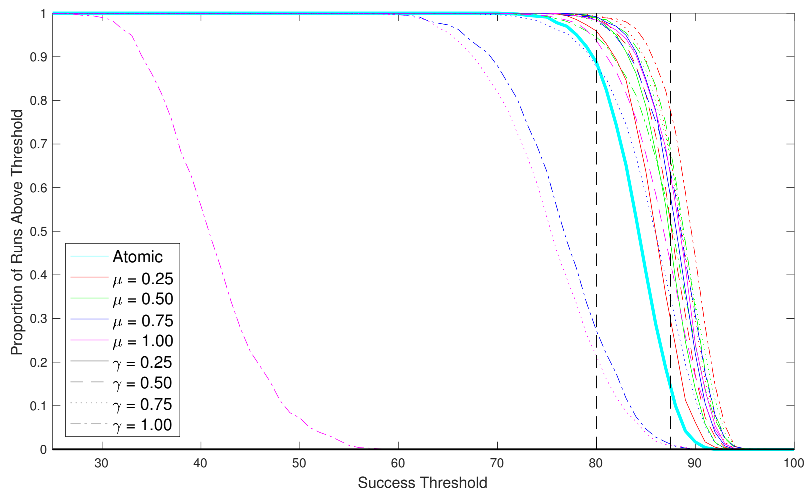

FIGURE 4. The proportion of successful runs above thresholds $[0.25, 1.0]$ shown for each combination of parameters, $[\mu, \gamma]$, compared with the atomic game. The vertical dashed lines indicate the thresholds 0.80, and 0.875, respectively. The atomic case is bold for clarity.

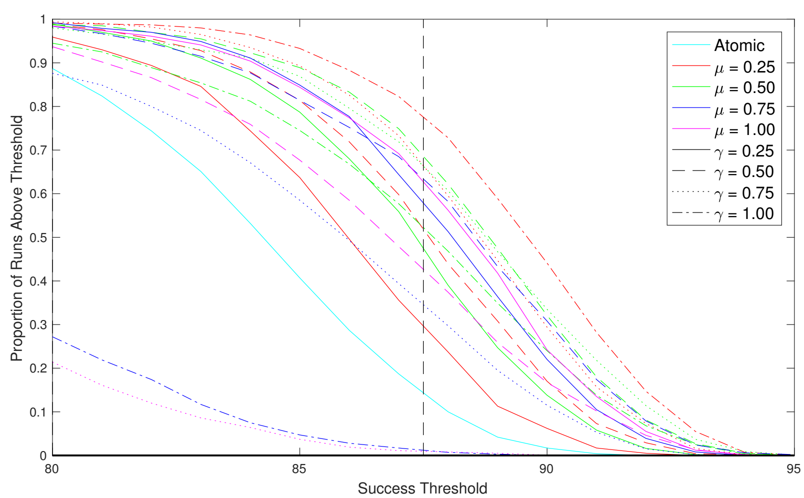

FIGURE 5. The proportion of successful runs above thresholds $[0.8, 0.95]$ shown for each combination of parameters, $[\mu, \gamma]$, compared with the atomic game, and centred about the pooling threshold, 0.875 (dashed vertical line)

($[\mu, \gamma] = [0.50, 0.75]$), they are almost 8 times more successful than in the atomic case, and they are able to achieve this rate of success more efficiently, with 2/3 the number of signals. With about half of the signals, they are able to coordinate more than 3 times as often than in the atomic case ($[\mu, \gamma] = [0.75, 0.75]$).

Note further that the invention of signals captures a notion of communicative development which is *diachronic* rather than synchronic. One charge against the assumptions in the signalling game model is that the sender and receiver start with a fixed number of messages (Hurford, 2012). The correction game from signalling with invention captures a subtle process, which we might take to be more realistic than the signalling game with fixed signals, in the following sense. The sender, might invent a signal for representing a particular state. When the sender invents a new signal, holding everything else fixed, she is *not* just choosing randomly; rather she is creating for the purpose of communicating a particular thing. If the receiver fails to understand, it stands to reason that, again, holding everything fixed, the sender might *insist*.

5. Discussion

In summary, correction is helpful in several cases, though this is not universally true. I repeat the claim made at the outset: this process presupposes little over and above the standard signalling game, as far as cognitive sophistication goes. When the sender tries to communicate that the receiver does X, and the receiver fails to understand the meaning of X, the sender corrects—"no, no no, do X!"—the *meaning* of X is *still* unknown, but the meaning of *no* is known, by the pre-evolved disposition. Thus, the receiver communicates perfect information that the action tried was incorrect, but the receiver still needs to *learn* which act *is* correct.

When the sender has pre-evolved dispositions, correction not only helps to avoid pooling equilibria in the 8×8 case, but it allows the receiver to learn how to coordinate to the sender's fixed disposition very quickly—in the best case, the sender and receiver have surpassed the pooling threshold every time after only 500 plays.

However, these results do not generalise to the signalling game, where the sender and receiver both learn their dispositions at the same time. This should not come as a surprise—early on in the game, when the sender's dispositions are highly variable, it makes little sense for her to insist that the receiver has done something wrong. Even in this case, however, those runs that surpassed the pooling threshold perform no worse than in the atomic case. This highlights that correction is helpful, as long as the sender knows what it is that she is correcting.

Finally, in the signalling game with invention, we saw that correction once again has a significant impact on learning, both in terms of speed and avoiding pooling equilibria, in most cases. The caveat here is that if the sender is too aggressive in trying to correct behaviour, it can be detrimental to learning; however, correction was only detrimental in two cases: when the sender always or almost always corrects the receivers behaviour, and there is no cost to correction. In every other case, correction helps learning. Furthermore, the sender and receiver are able to learn to signal more efficiently in the sense of requiring fewer synonymous signals: often, the end up inventing between 2/3 and 3/4 of the signals invented in the atomic case.

Note also that we interpreted the correction model as the sender correcting the receiver. However, nothing necessitates this interpretation. Indeed, in social animals correction may not be done by the sender, but rather by a conspecific who is a bystander. This may occur in the case of adults correcting juveniles.Thus, there are several different interpretations that are allowed by the generality of this model. Each of the agents in the model may be understood as functional components of a social group or of an individual agent.[18] Further, the agents in the meta-game need not be the same as the agents in the base game. We might imagine an observer watching a signalling interaction, keeping track of what signals and actions are correct given the conventions of the agents in the base-game, and occasionally correcting those actions that are inconsistent with previous behaviour.

5.1. **Relation to Previous Work.** It was assumed in every case that the correction game involved a *pre-evolved* disposition, of which the sender variably takes advantage. We might wonder whether the sender and receiver can co-evolve this disposition as they are learning to signal.

Barrett (2016) examines how a metalanguage might co-evolve with the language it describes. In the first model he describes, the meta-game co-evolves to indicate the success and failure of the base-game agents as they evolve signalling dispositions, in the sense of the atomic signalling game. In the second model he describes, the sender attends to the co-evolving conventional use of expressions in the base game. Thus, the meta-game evolves to track whether the expressions of the base-game are *true* (in a simple pragmatic sense), and so provides a sense in which the base language might be understood to have evolved propositional content (1–2).

The base game that Barrett (2016) describes is an atomic 4-game, where the agents evolve their dispositions under simple reinforcement learning. The meta-game is an atomic 2-game, which takes the success or failure of the sender and receiver in the base-game as input. The meta-game sender and receiver also learn by simple reinforcement. The state of nature, which the sender observes, may be obtained either by looking at the state and act of the base-game to see whether they match or examining whether or not the sender and receiver in the base-game received a payoff, for example. Actions in the meta-game correspond to success or failure, and the meta-game sender and receiver reinforce just in case the meta-game receiver's action matches the meta-game state. Thus, the meta-game agents learn from their observations of the evolving dispositions of the base-game agents. Barrett (2016) reports that on simulation, the meta-game receiver exhibits a cumulative success rate of better than 0.95 on better than 0.99 of the runs of the model, with 1000 runs of 10^6 plays per run.

This may seem unsurprising, given the results of Argiento et al. (2009) for the atomic 2-game. However, what happens here is slightly more subtle: Barrett (2016) points out that even the 2×2 signalling game can get stuck in sup-optimal pooling equilibria when nature is biased.[19] This is relevant because as the base-game evolves, the successes become more frequent. But the input (nature) in the meta-game is just the successes and failures of the base-game; thus, nature in the meta-game is unbiased at the outset, but becomes biased as the base-game evolves. Nature in the meta-game is strongly biased toward success over time.

Even so, the base game is more complicated than the meta-game. Therefore, the meta-game evolves more quickly than the base game. Hence, by the time nature in the meta-game becomes strongly biased toward success, the meta-game sender and

[18]See Barrett et al. (2018).

[19]See the discussion in Skyrms (2010). See Hofbauer and Huttegger (2008) for a proof in the context of the replicator dynamic.

receiver have already evolved a signalling system, which clearly demarcates the two input states. This happens 0.98 of the time, with 1000 runs of 10^6 plays per run.[20]

It was assumed, in this paper, that the sender and receiver in the correction game have already evolved a disposition to successfully communicate some binary "yes/no" signal. In the correction game, they *use* this pre-evolved disposition to try to correct the receiver's action when the sender and receiver miscoordinate. I take the results of Barrett (2016) to be sufficient for an affirmative answer to the question of whether or not the correction meta-game can co-evolve alongside the signalling game. The set up for a co-evolutionary correction game, where the meta-game agents learn the "yes/no" signalling disposition where the input (states of nature) for the meta-game are given by the success of failure to coordinate in the base-game, is almost equivalent to the "true/false" model that Barrett (2016) presents. The main difference in the correction game is that the *output* of the meta-game also affects the dispositions in the base game. However, we can imagine that the meta-game evolves in the atomic period of the correction game. Since the meta-game is significantly less complex than the base-game, the meta-game will evolve faster. Thus, by the time the sender and receiver start to *utilise* their dispositions, they should have already coordinated upon a signalling convention in the meta-game.

However, due to the subtleties described in Barrett (2016), these results will also be limitative: the meta-game will not necessarily evolve for significantly more complex base-games, like the 8×8 signalling game. This is because, in this case, there are 8^{16} (almost 300 trillion) possible combinations of strategies, but only 8! (slightly more that $40,000$) of these are signalling systems. At the outset, it is significantly more probable that the input for the meta-game will be a failure rather than a success (0.875 in the atomic case).

Indeed, when we examine the co-evolution of a 2×2 meta-game that tracks truth and falsity, taking successes and failures from the atomic 8×8 base game as input, only around 20% of the meta-game plays surpass the pooling threshold of 0.75.[21] However, what is typical here is that the sender perfectly partitions the states of nature (success and failure) by the two available signals, and the receiver learns the meaning of the failure signal, but is indifferent between her actions for the success signal. Thus, as successes become more frequent, it is probable that she will eventually learn the meaning of this signal—that is, she does not bias her action toward "failure" for each signal. This is for the atomic 8-game. It remains to be seen whether or how the signalling game with invention affects these meta-game propensities, given that the output of the meta-game feeds back in to the base game.

5.2. **Affirmation and Negation from a Linguistic Perspective.** Why is this particular pre-evolved disposition relevant to the evolution of communicative capacities? Negation is a universal category of human language Dahl (1979)—every

[20]In a second model, the meta-game sender tracks whether the base-game sender used the signal that is *customary*, given what the agents in the base game have been doing. This simple game can be extended to include the co-evolution of a pragmatic sense of *probability*; such a model is discussed further in Barrett (2017).

[21]This is for $25,000$ plays per run and $10,000$ runs—corresponding to the atomic learning period in the model presented in Section 3. However, the results are essentially equivalent when we increase this to 10^7 plays per run.

natural language, at least, to express clausal negation; however, the way that different languages negate varies. In English, and other indo-European languages, sentence negation is frequently realised by the negative participle "not".[22] For example,

(1a) Atlas believes that Sarah is not home.
(1b) Atlas does not believe that Sarah is at home.

In some languages, however, sentence negation is expressed by a negative verb. For example, in Tongan, the negator *ikai* acts as a higher verb which takes the corresponding affirmative clause as its complement, and *ke* is a subjunctive marker, which marks the complement clause as subordinate (Churchward, 1953, 56):

(2a) naˈe ˈalu ˈa siale.
PST *go* ABS Siale
'Siale went.'
(2b) naˈe ˈikai ke ˈalu ˈa siale.
PST NEG SBJN *go* ABS Siale
'Siale did not go.'

However, Miestamo (2007) notes that this type of negation is marginal.

Along with truth-functional negation, a large range of word-formation processes can be used to coin negative meanings. For example, in English, these word-formation processes include *prefixation*, *suffixation*, *compounding* and *conversion*. Morphologically, negation is quite complicated. For example, in English negation may be expressed through a number of *negative derivational affixes*: *de-*, *dis-*, *in-*, *non-*, *un-* and *-less*.

In most languages, negation systematically either precedes or follows the verb. Dryer (1988) studies the placement of the marker of sentential negation in relation to the subject (S), object (O) and verb (V)—three main clausal elements—in a worldwide sample of 345 languages. His results suggest that SOV languages are most commonly either SOV*Neg* or SO*Neg*V. *Neg*SOV and S*Neg*OV languages are infrequent. SVO languages are most commonly S*Neg*VO, and V-initial languages are almost always *Neg*V (i.e. *Neg*VSO or *Neg*VOS). In 70% of the 325 languages surveyed, Dryer (1988) finds that the negation marker is placed before the verb.[23]

It has been claimed that no animal communication system has a notion of negation (Horn, 1989; Jackendoff, 2002). Even so, it is suggested that some variety of *pre-logical* negation might be available in the cognitive representation of higher animals—this is consistent with the view that non-human animal communication systems lack recursion (Hauser et al., 2002); bona fide truth-functional negation in natural language is recursive to the extent that, semantically, it takes an arbitrary proposition, ϕ, and creates a new proposition, $\neg\phi$, where ϕ may itself be a negated proposition.

Negation, in natural languages, is complex for a variety of reasons. First, the logic of affirmation and negation is asymmetric: negations are generally less valuable, less specific, and less informative than affirmations (Plato, 1921). Aristotle (1995) held that affirmations have ontological, epistemological, psychological, and

[22]This is typically referred to as "standard negation". This terminology originates in Payne (1985).
[23]See also the discussion in de Swart (2010).

grammatical priority over negations (996b1416). Further, negations are morphosyntactically more *marked*, and psychologically more difficult to parse (Just and Carpenter, 1971; Horn, 1989). In some sense, negation presupposes affirmation: "the feeling is as if the negation of a proposition had to make it true in a certain sense in order to negate it" (Wittgenstein, 1953, §447).[24] Finally, affirmation usually introduces a proposition into the "discourse model", whereas negation—in its "chief use" (Jespersen, 1917, 4), its "most common use" (Ayer, 1952, 39), its "standard and primary use" (Strawson, 1952, 7)—is directed at a proposition that is already in, or that can be accommodated by, the discourse model.[25]

Proto-languages need not contain propositions nor truth-functions, though these would at least need to emerge somewhere in the transition from proto-language to language. In a review of the relevant literature, Heine and Kuteva (2007) suggest that trained animals are able to develop notions of rejection and refusal, and even of non-existence.[26]

In addition to the omnipresence of negation in natural languages, negation and affirmation may have evolved early on, and so serve as linguistic "fossils" of a one-word stage of the evolution of language, wherein single utterances serve holophrastic purposes and are not integrated into a larger combinatorial system (Jackendoff, 1999).[27] It is irrelevant that no known animal communication system contains a generalised negation; rather, what is important is the *signal* understood as a proto-command of encouragement or negation.

Several such one-word utterances exist in language: Jackendoff (1999) points to sudden, high-affect utterances, such as *ouch!*, *dammit!*, *wow!* and *oboy!*, and suggests that

> These exclamations have no syntax and therefore cannot be integrated into larger syntactic constructions[.] . . . They can remain in the repertoire of the deepest aphasics, apparently coming from the right hemisphere. There also exist situation-specific utterances such as *shh*, *psst*, and some uses of hey that have almost the flavor of primate alarm calls. Though the *ouch* type and the *shh* type both lack syntax, they have different properties. . . . Further single-word utterances include the situation-specific greetings *hello* and *goodbye* and the answers *yes* and *no*.

Hurford (2012) highlights the fact that such one-word phrases (along with pragmatic inference) allow for the possibility of conveying propositional information without the benefit of syntax.

This is specifically the type of linguistic fossil that is suggested by a pre-evolved disposition for correction.

5.3. **Future Work.** There are several variants of this simple correction model that might be of interest. For example, it was supposed that if the sender and receiver abandon an attempt to correct after n repetitions, then they receive a payoff of 0. However, we might suppose that there is a (time/effort) cost for correction, such

---

[24]See also Givón (1978).

[25]See also the discussion in Horn and Wansing (2017).

[26]See also Patterson (1978); Premack and Premack (1983); Herman and Forestell (1985); Savage-Rumbaugh (1986); Pepperberg (1999); Zuberbühler (2002).

[27]See also the discussion in Progovac (2015).

that the payoff is discounted even when the repetitions do not end in a success—i.e., the result is a negative payoff when the sender and receiver attempt to correct action and fail repeatedly. This sort of extension incorporates varying punishment for failure to coordinate in the same way that the correction game incorporates a varying (positive) payoff as a function of the number of repeat attempts made. This amounts to varying a parameter of the underlying model; several other such extensions could be made. For example, we might vary the initial payoff, $u(s, a)$, or we might add a punishment parameter for failure, even when the attempt is not repeated, and then vary the punishment to increase with an increase in repetitions. This is in addition to the usual parameters that might be varied—e.g., the dimension of the game, the underlying dynamic itself, including punishment in general, etc. In this case, we used a single, well-studied dimension and the simplest learning dynamic for illustrative purposes.

There are several questions that arise, with respect to analysis of the model that was presented here. For example, we might look at different choice rules for how the receiver chooses her action in the event of a repetition. It was supposed that the receiver reduces the set of possible actions by abandoning previous actions that resulted in a failure to coordinate. This was taken to be the most parsimonious decision for how the sender and receiver play this modified signalling game: given that we assume that the sender and receiver have already coordinated on a "yes/no" signal, it makes sense that the receiver would "understand" that she should not re-try the action that led to a failure. However, we might relax this assumption by allowing the receiver to randomise over the entire set of possible actions repeatedly—this might be plausible to the extent that individuals might keep trying something that was incorrect even when they are told it is incorrect.

In this case, there is some nonzero probability that the correction cycle will continue indefinitely, for any $\mu > 0$—especially as μ gets arbitrarily close to 1. Note that the probability on a given round that the sender and receiver end up in a loop of repetitions is 0, in the limit.[28] However, for any particular n, the probability that they miscoordinate n times in a row has non-zero probability.

Nonetheless, in the case of getting caught in pooling equilibria, the probability that the sender and receiver get caught in such an infinite loop may increase greatly. Thus, in terms of modelling, we might want to have an upper bound on how many times the sender bothers to try to correct the action of the receiver. Note that, given the modelling assumptions that were made in this paper, there will always be a maximum number of repetitions possible—namely, $(n-1)$ times for the $n \times n$ game.

Finally, we assumed a fixed μ and γ for any particular set of simulations that were run; there may well be an optimal combination of parameters. Another extension would be to see whether or not the sender and receiver can coordinate on such an optimal combination. That is, we might model the game in a way such that the

[28]For this particular case (the 8×8 correction game), assume the probability of repetition is 1; then, for example, at the outset when all dispositions are equiprobable, the probability that the sender and receiver miscoordinate, given a fixed state and signal, is 7/8. Thus, the probability that they miscoordinate n times in a series of n repetitions is $\left(\frac{7}{8}\right)^n$, since each miscoordination is independent of anything that has happened previously. Further,

$$\lim_{n\to\infty} \left(\frac{7}{8}\right)^n = 0$$

.

sender *learns* a probability parameter for attempting to correct μ. This will likely be most effective when μ is very low to start—to allow the sender and receiver to begin moving toward a coordination equilibrium—and then gradually increasing as time goes on. It might be the case that when the sender only tries to correct the receiver's behaviour when it is *salient* to do so.[29] For example, it might be more salient to try to correct when the sender has already clearly differentiated a particular signal, though the meanings of the other signals may still be in flux.

While this is all certainly food for thought, the purpose of this paper was to show, in one particular case, how the composition of games might allow for the more efficient evolution of signalling dispositions. The correction game, as presented, does precisely that.

References

Aldous, David J. (1985). Exchangeability and Related Topics. In Hennequin, P. L., editor, *École d'Été de Probabilités de Saint-Flour XIII 1983*, volume 1117 of *Lecture Notes in Mathematics*, pages 1–198. Springer, Berlin, Heidelberg.

Alexander, Jason McKenzie, Brian Skyrms, and Sandy L. Zabell (2012). Inventing New Signals. *Dynamic Games and Applications*, 2: 129–145.

Argiento, Raffaelle, Robin Pemantle, Brian Skyrms, and Stanislav Volkov (2009). Learning to Signal: Analysis of a Micro-Level Reinforcement Model. *Stochastic Processes and Their Applications*, 119: 373–390.

Aristotle (1995). Metaphysics. In Barnes, Jonathan, editor, *The Complete Works of Aristotle, The Revised Oxford Translation*, volume 2, pages 1552–1728. Princeton University Press, Princeton.

Arthur, Brian W. (1993). On Designing Economic Agents That Behave Like Human Agents. *Journal of Evolutionary Economics*, 3: 1–22.

Ayer, A. J. (1952). Negation. *Journal of Philosophy*, 49: 797–815.

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

Barrett, Jeffrey (2016). On the Evolution of Truth. *Erkenntnis*, 81: 1323–1332.

Barrett, Jeffrey (2017). Truth and Probability in Evolutionary Games. *Journal of Experimental and Theoretical Artificial Intelligence*, 29(1): 219–225.

Barrett, Jeffrey A. and Brian Skyrms (2017). Self-Assembling Games. *The British Journal for the Philosophy of Science*, 68(2): 329–353.

Barrett, Jeffrey A., Brian Skyrms, and Calvin Cochran (2018). Hierarchical Models for the Evolution of Compositional Language. Unpublished Manuscript. January, 2018. PDF File.

Beggs, A. W. (2005). On the Convergence of Reinforcement Learning. *Journal of Economic Theory*, 122: 1–36.

Börgers, Tilman and Rajiv Sarin (1997). Learning through Reinforcement and the Replicator Dynamics. *Journal of Economic Theory*, 74: 235–265.

Börgers, Tilman and Rajiv Sarin (2000). Naive Reinforcement Learning with Endogenous Aspirations. *International Economic Review*, 41: 921–950.

[29]See, e.g., LaCroix (2018b).

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.

Bush, Robert R. and Frederick Mosteller (1955). *Stochastic Models for Learning.* John Wiley & Sons, Oxford.

Churchward, C. Maxwell (1953). *Tongan Grammar.* Oxford University Press, London.

Crawford, Vincent P. and Joel Sobel (1982). Strategic Information Transmission. *Econometrica*, 50(6): 1431–1451.

Dahl, Östen (1979). Typology of Sentence Negation. *Linguistics*, 17: 79–106.

D'Arms, Justin, Robert Batterman, and Kryzyztof Gorny (1998). Game Theoretic Explanations and the Evolution of Justice. *Philosophy of Science*, 65(1): 76–102.

de Swart, Henriëtte (2010). *Expression and Interpretation of Negation: An OT Typology*, volume 77 of *Studies in Natural Language and Linguistic Theory.* Springer, Dordrecht.

Dennett, Daniel C. (1971). Intentional Systems. *The Journal of Philosophy*, 68(4): 87–106.

Dennett, Daniel C. (1987). *The Intentional Stance.* The MIT Press, Cambridge, MA.

Dryer, Matthew S. (1988). Universal of Negative Position. In Hammond, Michael, Edith Moravcsik, and Jessica Wirth, editors, *Studies in Syntactic Typology*, pages 93–124. John Benjamins, Amsterdam.

Erev, Ido and Alvin E. Roth (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *The American Economic Review*, 88(4): 848–881.

Givón, Talmy (1978). Negation in Language: Pragmatics, Function, Ontology. In Cole, P., editor, *Syntax and Semantics 9: Pragmatics*, volume 53, pages 69–112. Academic Press, New York.

Harms, William F. (2004a). *Information and Meaning in Evolutionary Processes.* Cambridge University Press, Cambridge.

Harms, William F. (2004b). Primitive content, translation, and the emergence of meaning in animal communication. In Oller, D. Kimbrough and Ulrike Griebel, editors, *Evolution of Communication Systems: A Comparative Approach*, pages 31–48. MIT Press, Cambridge, MA.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298: 1569–1579.

Heine, Bernd and Tania Kuteva (2007). *The Genesis of Grammar: A Reconstruction.* Oxford University Press, Oxford.

Herman, Louis M. and Paul H. Forestell (1985). Reporting Presence or Absence of Named Objects by a Language-Trained Dolphin. *Neuroscience and Biobehavioral Reviews*, 9: 667–681.

Herrnstein, Richard J. (1970). On the Law of Effect. *Journal of Experimental Analysis of Behavior*, 13: 243–266.

Hockett, Charles F. (1963). The problem of universals in language. In Greenberg, J., editor, *Universals of Language*, pages 1–29. MIT Press, Cambridge, MA.

Hofbauer, Josef and Simon Huttegger (2008). The Feasibility of Communication in Binary Signaling Games. *Journal of Theoretical Biology*, 254: 843–849.

Hopkins, Ed and Martin Posch (2005). Attainability of Boundary Points Under Reinforcement Learning. *Games and Economic Behavior*, 53: 110–125.

Hoppe, F. M (1984). Pólya-Like Urns and the Ewans Sampling Formula. *Journal of Mathematical Biology*, 20: 91–94.

Horn, Larry (1989). *A Natural History of Negation*. University of Chicago Press, Chicago.

Horn, Laurence R. and Heinrich Wansing (2017). Negation. In Zalta, Edward N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 edition.

Hurford, James R. (2012). *Language in the Light of Evolution II: The Origins of Grammar*. Oxford University Press, Oxford.

Huttegger, Simon M. (2007a). Evolution and the Explanation of Meaning. *Philosophy of Science*, 74: 1–27.

Huttegger, Simon M. (2007b). Evolutionary Explanations of Indicatives and Imperatives. *Erkenntnis*, 66: 409–436.

Jackendoff, Ray S. (1999). Possible Stages in the Evolution of the Language Capacity. *Trends in Cognitive Sciences*, 3: 272–279.

Jackendoff, Ray S. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, Oxford.

Jespersen, Otto (1917). *Negation in English and Other Languages*. Høst, Copenhagen.

Just, Marcel Adam and Patricia Ann Carpenter (1971). Comprehension of Negation with Quantification. *Journal of Verbal Learning and Verbal Behavior*, 12: 21–31.

Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4: 237–285.

LaCroix, Travis (2018a). Evolutionary Explanations of Simple Communication: Signalling Games and Their Models. Unpublished Manuscript. November, 2018. PDF File.

LaCroix, Travis (2018b). On Salience and Signaling in Sender-Receiver Games: Partial Pooling, Learning, and Focal Points. *Synthese*. Forthcoming.

Lewis, David (2002/1969). *Convention*. Blackwell, Oxford.

Miestamo, Matti (2007). Negation: An Overview of Typological Research. *Language and Linguistics Compass*, 1(5): 552–570.

Millikan, Ruth Garrett (2005). *Language: A Biological Model*. Oxford University Press, Oxford.

Patterson, Francine (1978). The Gestures of a Gorilla: Language Acquisition in Another Pongid. *Brain and Language*, 5: 72–97.

Pawlowitsch, Christina (2008). Why Evolution Does Not Always Lead to an Optimal Signaling System. *Games and Economic Behavior*, 63(1): 203–226.

Payne, John (1985). Negation. In Shopen, T., editor, *Language Typology and Syntactic Description I: Clause Structure*, pages 197–242. Cambridge University Press, Cambridge.

Pepperberg, Irene Maxine (1999). *The Alex studies*. Harvard University Press, Cambridge, MA.

Pitman, Jim (1995). Exchangeable and Partially Exchangeable Random Partitions. *Probability Theory Related Fields*, 102(2): 145–158.

Plato (1921). Sophist. In *Plato in Twelve Volumes*, volume 7. Harvard University

Press, Cambridge, MA.

Premack, David and Ann James Premack (1983). *The Mind of the Ape.* Norton, New York.

Progovac, Ljiljana (2015). *Evolutionary Syntax.* Oxford University Press, Oxford.

Roth, Alvin and Ido Erev (1995). Learning in Extensive Form Games: Experimental Data and Simple Dynamical Models in the Intermediate Term. *Games and Economic Behavior*, 8: 164–212.

Savage-Rumbaugh, Sue (1986). *Ape Language: From Conditioned Response to Symbol.* Columbia University Press, New York.

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

Skyrms, Brian (2014/1996). *Evolution of the Social Contract.* Cambridge University Press, Cambridge.

Steinert-Threlkeld, Shane (2016). Compositional signaling in a complex world. *Journal of Logic, Language, and Information*, 25(3–4): 379–397.

Strawson, P. F. (1952). *Introduction to Logical Theory.* Methuen, London.

Suppes, Patrick and Richard Atkinson (1960). *Markov Learning Models for Multiperson Interactions.* Stanford University Press, Stanford.

Sutton, Richard S. and Andrew G. Barto (1998). *Reinforcement Learning: An Introduction.* MIT Press, Cambridge.

Taylor, P. and L. Jonker (1978). Evolutionarily Stable Strategies and Game Dynamics. *Mathematical Biosciences*, 40: 145–156.

Thorndike, Edward L. (1905). *The Elements of Psychology.* The Mason Press, Syracuse.

Thorndike, Edward L. (1911). *Animal Intelligence: Experimental Studies.* The Macmillan Company, New York.

Thorndike, Edward L. (1927). The Law of Effect. *American Journal of Psychology*, 39: 212–222.

Wittgenstein, Ludwig (2009/1953). *Philosophical Investigations.* Wiley-Blackwell, Oxford, 4 edition.

Zollman, Kevin J. S. (2011). Separating Directives and Assertions Using Simple Signaling Games. *The Journal of Philosophy*, 108(3): 158–169.

Zuberbühler, Klaus (2002). A Syntactic Rule in Forest Monkey Communication. *Animal Behaviour*, 63(2): 293–299.