Evolving to Generalize

Trading Precision for Speed

Cailin O'Connor

Abstract Biologists and philosophers of biology have argued that learning rules that do not lead organisms to play evolutionarily stable strategies (ESSes) in games will not be stable and thus not evolutionarily successful [21, 12]. This claim, however, stands at odds with the fact that learning generalization—a behavior that cannot lead to ESSes when modeled in games—is observed throughout the animal kingdom [22]. In this paper, I use learning generalization to illustrate how previous analyses of the evolution of learning have gone wrong. It has been widely argued that the function of learning generalization is to allow for swift learning about novel stimuli. I show that in evolutionary game theoretic models learning generalization, despite leading to suboptimal behavior, can indeed speed learning. I further observe that previous analyses of the evolution of learning ignored the *short term* success of learning rules. If one drops this assumption, I argue, it can be shown that learning generalization will be expected to evolve in these models. I also use this analysis to show how ESS methodology can be misleading, and to reject previous justifications about ESS play derived from analyses of learning.

Keywords learning generalization · evolutionary game theory · philosophy of biology

1 Introduction

Stimulus generalization, or learning generalization, is a learning behavior wherein an actor conditioned to one stimulus responds in the same way to perceptually similar

Cailin O'Connor
Department of Logic and Philosophy of Science, University of California, Irvine
Tel.: 1-949-824-3812
E-mail: cailino@uci.edu

stimuli.[1] This type of learning is extremely well documented.[2] It occurs across a wide variety of test subjects—mammals, birds, reptiles, amphibians, insects—across contexts, and across sensory modalities [22, 8]. In evolutionary game theoretic models, however, learning generalization does not lead to the play of what are called 'evolutionarily stable strategies' (ESSes). One point that theorists have generally agreed on is that learning rules that do not lead organisms to play ESSes in games will not be stable and thus not evolutionarily successful (see Maynard-Smith [21] and Harley [12]). Why, this incongruity?

In this paper, I will use the case of learning generalization to investigate how previous analyses of the evolution of learning have gone wrong. I point out that such analyses have largely ignored the short term behavior of learning rules. Learning generalization is standardly thought to be adaptive because it allows actors to quickly learn to respond to novel stimuli [8]. In other words, it is especially useful in the short term. I present evolutionary game theoretic models of learning generalization and show that, indeed, generalizing can be beneficial in these models in that it helps speed learning. Furthermore, if one considers evolutionary models of learning where the short term behavior of learning rules *is* important, it becomes clear that generalization can evolve. This supports the argument that previous analyses ignoring short term learning were misguided. These results further inform game theory. Previous theorists used analyses of learning to argue that ESS behavior should be seen in the real world. The work presented here indicates that such claims are overly hasty. Furthermore, this analysis lends credence to the idea that ESS methodology is often misleading.

The paper will proceed as follows. In section 2, I will discuss previous work on the evolution of learning. In section 3, I will outline the 'approximation game', which appropriately models the class of scenario in which generalization is seen. In section 4, I describe several learning rules where actors generalize to varying degrees. I go on to show that in the long run rules that do not generalize outperform those that do in the approximation game. In section 5, I present simulation results showing that despite the long term success of non-generalized learning, under certain parameter settings higher levels of generalization can do significantly better in the short term. In section 6, I show that in evolutionary game theoretic models where short term learning is important, learning generalization can evolve. I conclude by discussing how this analysis informs game theory and evolutionary game theory.

2 The Evolution of Learning

Harley [12] and Maynard-Smith [21] use evolutionary game theoretic models to show that only certain sorts of learning rules should be expected to evolve. Without going into too much detail, these authors argue that only learning rules that lead to play of evolutionarily stable strategies (ESSes) in games should be expected to persist in an

---

[1] This behavior was documented in the famous 'Little Albert' experiment. Watson and Rayner [33] conditioned a nine month infant to fear a white rat by frightening the child with loud noises whenever he touched the animal. The child subsequently showed similar fear reactions to a number of fuzzy stimuli, including a rabbit and a fur coat.

[2] Thankfully not with regard to infant fear response.

evolutionary setting. An ESS is a strategy in a game that is robust against invasion by other strategies because it garners high payoffs for those using it.[3] The arguments Maynard Smith and Harley give are intuitively straightforward. Suppose that some learning rule does not lead to play of ESSes in games. A rule that *does* lead to play of an ESS will provide a higher payoff for those employing it. Then a learning rule leading to ESSes will be more evolutionarily successful than one that does not, and will be able to invade a population of those using a non-ESS learning rule.

This argument leads to a puzzle, however. Generalized learning cannot lead to play of ESSes in games (as I will show in section 4). How does the observed ubiquity of learning generalization in the natural world square with these results?

The work of Maynard-Smith and Harley, of course, is not the end of the discussion of the evolution of learning. It has been pointed out by Smead [29] that learning rules that take populations to ESSes have no advantage over static behavioral rules where the actor simply adopts ESS play rather than bothering to learn it.[4] Furthermore, most models of the evolution of learning assume that learning will bear a greater cost than non-learning strategies (for cognitive architecture, time required to learn, etc.). This means that non-learning strategies that adopt ESS play will actually receive *higher* payoffs than rules that learn such play and so should be able to invade these learning rules. This point seems to create a worry about learning generally. If learning rules that do not lead to ESSes are unstable, and static behavioral rules can invade learning rules that *do* lead to ESSes, there are no stable learning rules at all (never mind ones that generalize).

The usual response by biologists and philosophers of biology to worries of this sort is to argue that learning rules are primarily useful in situations where the environment exhibits some level of variability.[5] In such environments, the argument goes, non-learning strategies get poor payoffs because the actors cannot respond to changing payoff structures by changing action. Actors that play an ESS in one situation, but cannot deal with changes to the environment, now do poorly against learners that reach this same ESS in the original situation and can re-adapt when necessary.

Something is amiss here, though. The arguments forwarded by Maynard-Smith and Harley explicitly depend on the following assumption: when modeling the evolution of learning one can ignore what happens in the short term. In other words, when associating fitnesses with learning rules, these authors do not consider payoff *while* the actors are learning. Instead, they look only at the payoffs of the long term, stable strategies developed by learners. To date, most game theoretic work on the evolution of learning has shared this assumption.[6]

But, if learning is most effective in a variable environment, to the extent that it should not be expected to evolve otherwise, this assumption is suspect. In a vari-

---

[3] To be specific, an evolutionary stable strategy x_i is one such that if $u(x_i, x_j)$ is the payoff of strategy x_i played against x_j: 1) $u(x_i, x_i) > u(x_i, x_j)$ or 2) $u(x_i, x_i) = u(x_i, x_j)$ and $u(x_i, x_j) > u(x_j, x_j)$ for all $x_j \neq x_i$.

[4] Maynard-Smith [21] was aware of this. Smead and Zollman [31] find something similar. Smead [30] also argues that learning rules that lead to equilibria in many cases should not be expected to evolve.

[5] See, for example, Godfrey-Smith [11], Plotkin and Odling-Smee [25], Johnston [17], Stephens [32], Dunlap and Stephens [6], Shettleworth [28], and Maynard-Smith [21].

[6] There are some exceptions. Zollman and Smead [34], for example, use interim strategies developed by learning rules to determine the fitnesses of actors in an evolutionary model.

able environment, an actor will be changing strategies and so may spend a significant amount of time playing strategies that are not stable, long term outcomes of the learning process. If so, short term behavior should be important to the evolution of learning.[7] In particular, if payoff in the short term matters there should be selection pressure for learning rules that work quickly.

Biologists and psychologists have argued that the function of learning generalization is to allow organisms to quickly learn to respond to novel scenarios [8]. Furthermore, as mentioned, it should not evolve according to Maynard-Smith and Harley. As such, this learning behavior is an excellent case to explore whether the intuitive argument I just gave—that short term learning matters in an evolutionary context—is correct. In the rest of the paper, I will present evolutionary game theoretic models of learning generalization. As I will show, when the short term behavior of learners is incorporated into evolutionary models, generalization will evolve for just the reasons that biologists and psychologists outline. If short term behavior is ignored, on the other hand, generalization will not evolve. These results indicate that the intuitive argument is right, and that ignoring short term behavior of learning rules can lead evolutionary analyses significantly astray.

3 The Approximation Game

Learning generalization occurs when an organism applies behavior that was successful in one scenario to a perceptually similar scenario. What this means is that an appropriate model to explore the evolution of this phenomenon will need to include 'similar' scenarios for the actor to potentially generalize over. In order to do this, I introduce the approximation game.[8] The approximation game involves one actor and occurs in two stages. In the first stage, a state of the world is chosen probabilistically by nature or some exogenous force. In the second stage, the actor observes this state of nature and chooses an act. The state/act combination then determines what sort of payoff the actor receives. In order to model the type of scenario in which generalization evolves, the possible states of the world are assumed to bear similarity relationships to one another. This is done by treating these states as existing in a metric space where distance represents similarity. For example, an approximation game might have three states (1, 2, and 3) existing on a line. If state 1 is closer to state 2 than to state 3, it is assumed that state 1 is more similar to state 2.[9]

---

[7] Smead [29] points out something similar. Empirical observations about, for example, death rates in young birds also confirm the important of learning speed in animals [28].

[8] This model should more properly be called the 'approximation problem' because it is a one-player decision problem rather than a multi-player game. Decision problems, however, are formally idental to one-player games. For this reason, the relevant results on the evolution of games directly bear on decision problems, and results from the problems investigated here can be used to inform evolutionary game theory. For simplicity sake, then, I use the language of game theory, and not decision theory, to describe the model used.

[9] Note that this is similar to the sim-max game, introduced by Jäger [16] to model signaling in situations where states of the world bear similarity relations to one another.

For each state of the world in the approximation game, it is assumed that there is some ideal act which, should the actor choose it, will give a perfect payoff.[10] It is also assumed that acts will receive similar payoffs in similar states. In the previous example, in state 1 the actor would achieve a perfect payoff by choosing act 1. But she would also obtain a good payoff for choosing act 2. Her payoff for choosing act 3 would be less good. One simple way to model this is to determine payoff using a function that takes as input the distance between the state and the act.[11] For the purposes of this paper, unless otherwise specified it will be assumed that the actor's payoffs are strictly decreasing with distance between state and act.

Figure 1 shows the simplest approximation game of interest—the one described above. The central node of the figure represents the starting point of the game, where nature chooses a state (S1, S2, or S3). The probabilities that each state is chosen by nature are fixed at p, q, and 1-p-q. The three decision nodes labeled 'A' for actor represent the possible choices of act in each state (A1, A2, or A3). Payoffs for each state/act combination are shown at the final nodes. It is assumed that $0 < \varepsilon < \delta < 1$ (payoff decreases strictly in distance between state and act, but is always positive). It is also assumed that p and q are strictly positive and $p + q < 1$ (that every state is played with positive probability).
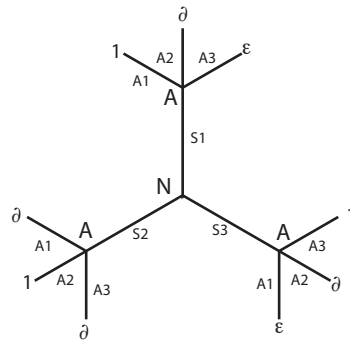

Fig. 1 A 3 state/3 act approximation game with payoffs 1, δ and ε for distance of 0, 1, and 2 between state and act. The game begins with the central node labeled 'N' for nature and continues to the three decision nodes labeled 'A' for actor.

Figure 2 shows some possible state spaces for approximation games. Diagram (a) represents the state space of a game like the one just outlined, i.e., modeled on a line, but with four states. Diagram (b) shows a game with a two dimensional state space.[12] Approximation games with state spaces of any dimensionality are possible, though this paper will only consider the simplest ones—those where states are modeled on a

---

[10] For simplicity sake, acts will always be labeled by the state they are most appropriate for, i.e., act 1 will be the ideal act for state 1 and so forth.

[11] This is a useful way to understand payoff in these games. It is more precise to say that a payoff is defined for each state-act pair, and this payoff is chosen using such a function.

[12] Note that games with state spaces of higher dimensionality can be used to model cases where an actor is responding to states with multiple properties varying along different dimensions.

line. For the purposes of this paper, these spaces are best understood as representing *perceptual* similarity spaces.[13] In other words, the states of the world in the game correspond to perceptual states. This is a useful interpretation of the model as learning generalization happens over perceptually similar states. It also avoids sticky issues around how or whether external states are similar to each other.
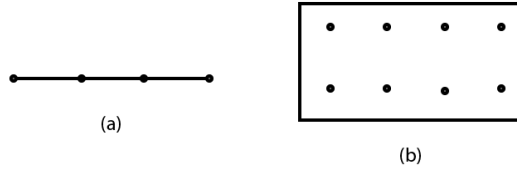

(a)

(b)

Fig. 2 Two examples of state spaces for an approximation game. Diagram (a) shows a game with four states modeled on a line. Diagram (b) shows a game with eight states modeled in a plane.

Most of the approximation games considered in this paper will have a few properties that bear mentioning. First, they will have considerably larger state spaces than the game described above. The reason for this is that in real world learning scenarios, the number of possible states of the world is often extremely large. This is certainly true under the interpretation of the game here—that the actor is responding to perceptual states. Consider, for example, the number of discriminable colors picked out by the human visual system, or the number of distinguishable smells. Furthermore, as I shall show later in the paper, considering games with large state spaces is relevant for understanding why generalized learning might evolve. Second, in the games considered, payoff loss over distance will usually be modeled with a gaussian function. This function is used because it is always positive and strictly decreasing. These attributes make it particularly tractable from a modeling perspective. While this choice may seem arbitrary, the analytic results presented are robust under choice of function as long as it is strictly decreasing.[14] I will call the gaussian just described the 'payoff gaussian' as it determines the degree to which an approximate match of state and act will lead to payoff for the actor.

As noted, for every state of an approximation game there is one ideal act. A strategy for a game defines an act in every possible state.[15] What this means is that there is a single, optimal strategy for every approximation game in which the actor always picks the correct act for the state. The existence of a single optimal strategy is significant from an evolutionary standpoint. Under the replicator dynamics, the most common model of evolutionary change in evolutionary game theory, a population playing the approximation game will evolve to take this strategy in every case. For

---

[13] See Gärdenfors [7] for more on such spaces. See Krantz et al. [18] for how such spaces can be built using experimental data.

[14] O'Connor [23] also found that results in simulations of related signaling games were robust under choice of function for payoff loss modeled as linear, quadratic, or decreasing in steps.

[15] Again, while the term that technically should be used here is 'choice' because this is a one-player problem, I use 'strategy' to avoid confusion. Once again, nothing hangs on this distinction.

this reason, the approximation game would not usually be of much interest to evolutionary game theorists—it is immediately obvious what behavior will be adopted by a population *evolving* to play it. However, as I will argue in the next section, an organism *learning* to respond to this game, and employing generalizing learning, will not develop the optimal strategy.

4 Learning Rules

4.1 Herrnstein Reinforcement Learning and Generalized Reinforcement Learning

In evolutionary game theory, learning dynamics, unlike evolutionary dynamics, are taken to model the emergence of learned individual behaviors over the course of a organism's lifetime, rather than the emergence of evolved population behaviors over the course of evolutionary time. Herrnstein reinforcement learning, first proposed by Roth and Erev [26], is so named in reference to R.J. Herrnstein's psychological work on learning, which motivates the model [14].[16] This learning rule has been widely used in evolutionary game theory because 1) it is psychologically natural, i.e., based on observed learning behavior and 2) it makes minimal assumptions about the cognitive abilities of the actors. This means that behaviors which emerge under this rule can be assumed to be available to cognitively simple animals. In this case, because generalized learning is seen in a wide variety of animals, including those with minimal cognitive abilities [22], Herrnstein learning is an appropriate starting place to model it.

The basic assumption that underlies reinforcement learning rules is that actors will be more likely to repeat successful behavior. In other words, they *reinforce* this behavior. In a simulation of these rules actors engage in a game many times, at each step reinforcing successful behavior and thus improving their strategies. Herrnstein learning can be described using the following analogy. In the context of the approximation game, imagine that for each state of the world the actor has an urn into which is placed one colored ball for each possible act available to her. In the first round of learning, nature selects a state of the world and the actor draws a ball from the urn for that state. The color of the ball determines which act she will take. If the act is successful, the actor returns the drawn ball to her urn and then reinforces her tendency to take that act in that state by adding a ball (or two, or half a ball, etc.) of the same color to that urn. The reinforcement is proportional to the success of the act, i.e., the higher the success the greater the reinforcement. For our purposes, the amount of reinforcement will always be equal to the payoff achieved by the actor in each step of the simulation. At the beginning of a simulation using Herrnstein learning, an actor uses all her acts with equal probability, as she has one of each type of ball in each urn. As play progresses and successful acts are reinforced, the actor becomes increasingly likely to choose these acts. In the limit, the actor's strategy may, under the right cir-

---

[16] This learning rule is also sometimes called 'Roth-Erev' or 'Vanilla' reinforcement learning.

cumstances, converge to a successful one. In other words, she will use this strategy
with probability approaching 1.[17]

Generalized reinforcement learning (GRL) builds on the Herrnstein reinforce-
ment learning model.[18] Under GRL rules, successful acts are reinforced, but they
are also generalized, i.e., reinforced for other, similar states of the world. In other
words, to continue the urn analogy, when an actor draws a colored ball from her urn
for a state and takes a successful act she adds balls of the same color to that urn,
but also adds balls of that color to the urns for similar states. It must be specified,
for these rules, the degree to which generalization occurs. How many other states
are reinforced? How much reinforcement occurs in those states? For the purposes of
this paper, generalization will be determined using a gaussian function. To be clear,
a model of an approximation game evolved using GRL employs two gaussian func-
tions. The payoff gaussian, introduced above, determines the level of payoff based
on how accurate the act chosen is for the state. The second gaussian determines to
what degree this payoff is generalized taking as input the distance between the state
of the world and the state to be reinforced. I will call this second gaussian the 're-
inforcement gaussian'.[19] Figure 3 represents the way these two functions determine
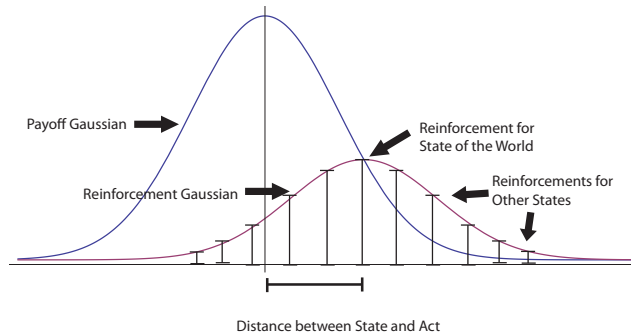reinforcement in an approximation game evolved using a GRL rule.

Fig. 3 A representation of how the payoff and reinforcement gaussians determine payoff in an approxi-
mation game evolved using a GRL rule.

A model of an approximation game evolved using these learning rules will have
five relevant parameters. The first is the size of the state space of the game. The sec-
ond and third are the height and standard deviation of the payoff gaussian. These

---

[17] For more on this and other learning dynamics see Huttegger and Zollman [15]. For extensive work on
Herrnstein reinforcement learning and variations of it in signaling games (which are in some ways similar
to the approximation game) see recent work by Barrett [1, 2, 4].

[18] This learning rule was first outlined by O'Connor [23]. Roth and Erev [26] look at a learning rule that
incorporates a slight amount of generalization in a similar way to GRL. They interpret this aspect of the
learning rules as persistent error.

[19] Ghirlanda and Enqvist [8] argue that generalization is best modeled in many cases by a gaussian
function, suggesting that the choice of a gaussian as the reinforcement function here is a natural one. Fur-
thermore, Shepard [27] argues that the specifics of how an actor learns to generalize may not be particularly
important in determining subsequent behavior.

control the level of payoff for perfect coordination in the approximation game (the height) and the degree to which an actor receives payoff for imperfect action in the game (the standard deviation). The fourth parameter is the standard deviation of the reinforcement gaussian.[20] Variations of this parameter correspond to GRL rules with different degrees of generalization. In models where Herrnstein reinforcement learning is used, this parameter will not apply. It can be noted, though, that Herrnstein learning is a limiting case of GRL as the width of the reinforcement gaussian approaches 0. The fifth relevant parameter will be the length of trial for simulations of these models. This parameter will control the number of times the actor plays the approximation game and updates her strategies.

4.2 Long Term Success

One way to explore the evolution of generalized learning is to compare learning rules with different levels of generalization, like GRL and Herrnstein reinforcement learning, to see if high levels of generalization can outperform lower levels in these models. One method for doing this is to consider convergence outcomes of the models just described. When this is done, however, it becomes clear that in the long term, Herrnstein reinforcement learning can always outperform GRL in the approximation game.

Laslier et al. [19] show that a single actor employing Herrnstein learning in a stationary environment, i.e., where payoffs remain constant, in the long run will always learn to play the act that receives the highest expected payoff.[21] This result can be applied to each state in the approximation game. To do so requires that each state be a stationary environment, which is the case given that the payoffs in the approximation game do not change.[22] It also requires that each state be selected infinitely often as the length of learning goes to infinity, which is also the case as each state in the approximation game has a strictly positive probability. Thus these results indicate that in the long run, for each state in the approximation game, the act of an agent employing Herrnstein reinforcement learning will converge to the optimal one. For the entire game, then, the strategy of the actor will converge to the optimal strategy. In the long run, the actor will take the perfect act in every state in the approximation game if using Herrnstein learning. This result holds for an approximation game of any finite size.

What happens to the strategy of an actor using a GRL rule in the approximation game in the long run? Unlike Herrnstein learning, GRL rules will not converge to the optimal strategy and, in fact, the level of generalization will determine a bound of accuracy which a player will not be able to surpass. This bound of accuracy will

---

[20] The height of the reinforcement gaussian is determined by the level of payoff.

[21] In other words, as the learning time goes to infinity, the probability with which the actor chooses non-optimal acts goes to 0.

[22] The results of Laslier et al. [19] hold for expected payoffs. An expected payoff is an average payoff over possible outcomes weighted by the probabilities of those outcomes. Note that for an approximation game, the payoff for choosing an act in a state is always the same, and so the expected payoff in that state is simply equal to the payoff. These results would apply even if this were not the case, as in a multi-armed bandit problem.

in turn determine a bound on the payoff success an actor can achieve. The intuitive reason for this is that if an actor were able to converge to the perfect act in one state, she would simultaneously prevent convergence in neighboring states by generalizing the same act to them.

One can show this by solving for the consistent, limiting probabilities of acts for a model of the approximation game evolved using a GRL rule. This is done by finding the distribution of reinforcements in a game where the probability of an act being selected in one round of simulation is equal to the probability of it being selected in the next round. Consider a toy model of the approximation game with two states and two acts. Suppose that in each state the payoff for the perfect act is 2 and for the other act is 1. Assume that states of the world are equiprobable.[23] This game is pictured in figure 4 which should be read like figure 1. Also consider a simple form of GRL where successful acts are reinforced in the state of the world by the amount of the payoff and in the other state by that amount multiplied by α where $0 \leq \alpha \leq 1$. In this simple model, α determines the level of generalization. A high α means that success will lead to strong generalization in the other state of the world, a low α will mean that generalization is weak. If α is equal to .1, the consistent, limiting probabilities of this game are such that the actor selects the more successful act in each state with probability 5/6 and the other act with probability 1/6. It is possible (though increasingly difficult) to calculate such limiting probabilities for larger games and more complex generalization rules.
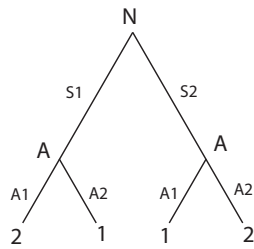

Fig. 4 A 2 state/2 act approximation game with payoffs 2 and 1 for distance of 0 and 1 between state and act. The game begins with the central node labeled 'N' for nature and continues to the two decision nodes labeled 'A' for actor.

One can further explore this phenomenon through simulation. It is easy to show what happens in this toy model at the two bounds of α. If one sets $\alpha = 0$, the learning rule is the same as Herrnstein learning and so converges to perfect behavior. If one sets $\alpha = 1$, the actor fully generalizes. In other words, if she reinforces act 1 in state 1 by .43, she will also reinforce act 1 in state 2 by .43 and so on. This complete generalization of success means that reinforcement levels for the actors will always be identical in the two states of the world. Because actors will not be able to learn to condition their acts on which state has been selected, every attainable strategy (those

---

[23] This degenerate approximation game is not generally an interesting one as it is formally the same as a game with no similarity structure over the payoffs. It is useful, however, as a simple case to consider GRL.

where the probability for each act is the same in both states) will get an expected payoff of 1.5, the same as choosing by chance.

For intermediate levels of α, simulations of the toy model show that the actor eventually reaches a level of accuracy, and thus success, that is bounded by the level of generalization. The lower the generalization, the greater the success. In figure 5 success rates are shown for a simulation of this game for α ranging from 0 to .3 and α equal to 1. In each case, success is calculated by dividing the expected payoff for the actor given her learned strategy by the perfect possible expected payoff (which, in this case, is 2).

$$\text{Success} = \frac{\text{expected payoff given learned strategy}}{\text{perfect possible expected payoff}}$$

Each line represents the success rate of a simulation over time for a different level of generalization. Darker lines represent lower levels of generalization. Rates were averaged over 50 runs of simulation. As should be clear from figure 5, for each level of generalization, the success of the simulation reaches some upper bound and stays there. Note that time is presented logarithmically. The reason for this bound on success has already been laid out. When the actor generalizes, success in one state means that an act will be taken with greater probability in other states where it is less successful.
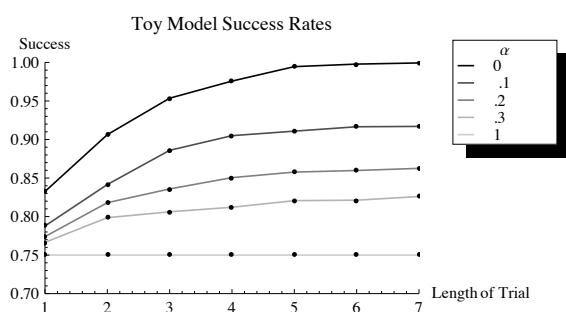

Fig. 5 Success levels for a 2 state/2 act approximation game with various levels of generalization (α). The y-axis tracks success and the x-axis represents of length of the trial where each value x is 10^x runs.

The results from these toy models can be extended to larger approximation games since in every larger game reinforcement in neighboring states will prevent convergence in the same way as it does in a two state model.[24] Thus these results indicate that in the approximation game, over the long run, low levels of generalization will outperform high levels of generalization from a payoff perspective, and in particular Herrnstein learning will outperform any GRL rule. The single optimal strategy

---

[24] To see why this is the case, consider two states of any larger approximation game. Use the reinforcement gaussian for this larger game to define α as above (the proportion of reinforcement on a neighboring state). It has been shown that this smaller system cannot reach an optimal strategy and so the larger system it is a part of cannot either.

provides the highest possible level of payoff in the game, and so learning to use any other strategy will be strictly worse. Importantly, the optimal strategy in an approximation game is always the unique ESS. Therefore, GRL is unable to learn ESSes in this game, while Herrnstein is guaranteed to do so.

Furthermore, although this analysis only addresses approximation games, it may be extended to some other games, including ones with multiple players. O'Connor [23] obtained similar simulation results in sim-max games, which are a variation on the Lewis signaling game where the state space has the same similarity structure as the approximation game. Unlike approximation games and sim-max games, most games do not have several possible states and so it is not possible to evolve them using GRL. For those games that do, though, if an actor generalizes over states she will only be able to achieve optimal behavior if the acts generalized are ideal for all the states they are generalized to. Otherwise, generalized learning will lead to reinforcement of sub-ideal acts and thus to sub-optimal behavior, preventing play of ESSes.

5 Short Term Success and Simulation

As I will outline in this section, there is a tension that can arise between the two desiderata a learning rule should meet—working quickly and developing behavior that obtains the highest possible payoff.[25] While low generalization learning outperforms high generalization learning eventually, the very property that prevents high generalization rules from approaching optimal behavior is the one that allows them to outperform low generalization rules in the short term. I will illustrate this argument using simulation results showing that in trials of the approximation game, high levels of generalization can outperform low levels under certain parameter settings. In particular, high generalization does best when states of the world are numerous, when trials are short, and when the payoff gaussian (modeling how accurate an actor must be to get a good payoff) is wide. This result confirms intuitive arguments about the benefits of learning generalizations.

All the results presented in this section were generated using models where payoff and reinforcement were calculated with gaussian functions. Each trial of a parameter setting was run 50 times and reported results are averages of these. The parameters that varied were the size of the state space, the length of the trial, the standard deviation of the reinforcement gaussian, and the standard deviation of the payoff gaussian.[26] The state spaces considered were of size 100, 200, 300, 400, and 500. The lengths of trial were 1,000, 10,000, 100,000, and 1 million runs. The reinforcement gaussian standard deviations were 5, 10, 15, 20, and none (Herrnstein learning). And lastly, the payoff gaussian standard deviations were 1, 5, 10, 15, and 20.

---

[25] This has been widely observed in other fields. It has been argued in psychology that 'fast and frugal' decision heuristics, which allow actors to make decent strategies quickly and easily, are adaptive, despite the possibility that they lead to irrational or sub-optimal behavior [10, 9]. Generalized learning can be thought of as a learning rule that leads to making decent, if sometimes inaccurate, decisions quickly. In machine learning, much work has been done on learning models that generalize from limited input to make predictions in novel scenarios. Similar tradeoffs between speed and accuracy are found in these models [13].

[26] Height of the payoff gaussian was always 2.

Figure 6 shows the success rates (calculated as they were in the previous section) for one set of these trials—those where the payoff gaussian had a standard deviation of 10. The x-axis of the figure represents the length of trial (ranging from 1,000 runs to 1 million). The z-axis tracks the size of the state space (from 100-500). And the y-axis tracks average success of the trials. Each surface shown represents results for one reinforcement width parameter setting. In other words, each surface corresponds to one learning rule and these rules vary with respect to generalization. The black surface represents the highest levels of generalization (a reinforcement gaussian with a standard deviation of 20) and successively lighter surfaces represent lower and lower levels of generalization.
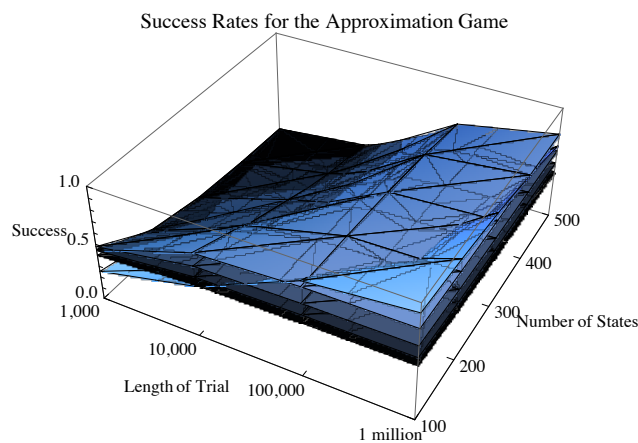

Fig. 6 Average success levels for various parameter settings for an approximation game with a payoff gaussian of standard deviation 10 evolved using GRL and Herrnstein reinforcement learning. Results are averaged over 50 runs of each setting.

As is evident in the figure, each level of generalization considered outperforms the others for some region of parameter space. The rule with the highest level of generalization (black) outperforms the others in the area of parameter space where trials are short and the number of states of the world is large. Herrnstein learning (the lightest surface) performs best in the longest trials and when states of the world are fewer. These results should not be surprising. In a short trial with many states of the world, there is not enough time for the actor to learn ideal actions in each state, so a learning rule that allows success to be generalized does better. When an actor has a long time to learn, more precise strategies can be developed using low generalization rules and so these do better. Similar results were obtained for the other payoff gaussian values with the slight difference that in games where approximate action was successful (wide payoff gaussians), higher generalization could perform better. In extreme cases of games with very narrow payoff gaussians approximate actions do not receive a good payoff. Generalization thus does not help the actor in this case, because only precise strategies will be successful.

Real world learners do not use learning strategies that exactly mimic those used in the models here. In order to strengthen these results, I investigated their robustness across learning rules. Under reinforcement learning with punishment, actors reinforce successful acts for the state of the world (and for similar states under the generalized version), and simultaneously punish, or decrease the reinforcement level for that act in other states.[27] The results of simulations for these rules were highly similar to those presented in the last section. I also explored a learning rule outlined by Barrett [3], which I call Barrett Learning. This rule is in some ways similar to Adjustable Reference Point with Truncation learning introduced by Bereby and Erev [5]. Actors using this rule discount past experience compared to more recent experience. Results were, again, very similar to those presented in this section.

It should be noted that the results presented in this section are not particularly surprising given previous results from machine learning, and previous observations from psychology and biology about the benefits of generalization. As we will see in the next section, however, generating similar results in an evolutionary game theoretic model is useful in that is allows us to discuss the motivating problem presented in section 2: why do previous evolutionary game theoretic analyses of learning predict that rules like GRL should be unable to evolve if generalization is so ubiquitous?

6 Evolving to Generalize

At this point it has been established that high generalization learning can perform well in the approximation game when time is limited and states are numerous despite the fact that only non-generalized learning leads to optimal behavior. How, it will now be asked, do these results inform the *evolution* of learning generalization?

The larger question at hand, remember is whether or not it is problematic to assume that the short term behavior of learning rules does not matter in evolutionary analyses. In order to assess this using the case of learning generalization, let us consider an evolutionary model where the environment for the actor changes regularly, meaning that speed of learning may be evolutionarily relevant. The replicator dynamics are the most commonly used model of the evolutionary process in evolutionary game theory and will be employed here. These dynamics assume that actors using strategies that receive higher payoffs will replicate more successfully than actors using strategies that receive lower payoffs.[28] In populations modeled under these dynamics high payoff strategies tend to proliferate. In the approximation game in particular, because there is only one player, the learning rule that will evolve under the replicator dynamics is simply the one that gets the best payoff.

Consider a model where a population of actors learns to play an approximation game using either Herrnstein learning or various GRL rules. One can think of the ac-

---

[27] There is experimental evidence supporting the use of rules where actors punish or forget strategies, i.e., sometimes decrement their reinforcements. See Bereby and Erev [5], for example.

[28] The replicator equation determines how proportions of strategies in a population change under the replicator dynamics. This equation states that $\dot{x}_i = x_i(f_i(x) - \sum_{j=1}^{n} f_j(x)x_j)$ where x_i is the proportion of a population playing strategy i, $f_i(x)$ is the fitness of type i in the population state x and $\sum_{j=1}^{n} f_j(x)x_j$ is the average population fitness in this state.

tors' strategies as now consisting in which learning rules to adopt. The payoffs associated with each learning rule will be the expected payoffs for the behavioral strategies that these various learning rules develop in simulation. Now suppose that at regular intervals, the population encounters a new approximation game (one where the actors encounter new states and must associate them with new actions). If these intervals of learning are short enough, under the replicator dynamics this population will evolve to use a GRL rule rather than Herrnstein learning. This is the case because, as shown in the previous section, generalizing rules will lead to higher payoffs for the actors over a short timescale. And, as pointed out, for an approximation game the replicator dynamics will always select whichever behavior receives the best payoff. To give an example, suppose that actors in the population play approximation games with 100 states, and that they switch games every 1,000 rounds. If the initial population contains the learning rules considered in the last section (Herrnstein learning and GRL with reinforcement gaussians of widths 5, 10, 15, and 20), GRL with a reinforcement gaussian of width 10 will evolve. In other words, when the environment varies, generalization can evolve.

One might worry that in the model just described actors begin their learning processes anew when the environment changes rather than having to forget currently developed actions. To alleviate this worry, I also considered models of populations in changing environments where actors must forget previously learned strategies when the world changes. I found that under a wide range of parameter settings, generalization evolved.[29]

Furthermore, there is a feature of learning situations that I have not discussed yet which makes generalization relatively more important and more successful in real world scenarios with numerous states. In the approximation game, every possible state of the world has its own ideal act. In reality, though, for highly similar states, it will often be appropriate for an organism to take the same act, in which case generalization will be more effective than the models here predict [28]. To further elucidate this claim, consdier a scenario where a bird is learning to interact with blackberries. Imagine a model of this scenario. The state space of this model would have hundreds (thousands?) of states varying along multiple dimensions of perceptual space—smell, size, color, shape, etc.—but the birds would only have two available acts—eat and not-eat. Generalization, in this case, will only lead to suboptimal behavior for states right at the boundary between edible and inedible berries. For all the other possible states, generalizing will be completely successful. In models of this scenario, Herrnstein learning will still lead to ESS play while GRL will not, but the benefits of Herrnstein learning are only relevant for a small proportion of states, while GRL provides more significant benefits for most of the state space. In other words, the window of time during which GRL is a more successful learning rule is longer, making it more problematic to ignore short term learning behavior in evaluating the evolution of generalization.

In the evolutionary models presented above, the learning rule that evolves strictly outperforms the other learning rules from a payoff perspective, and so satisfies the

---

[29] These results are not presented here as the description of these models is lengthy and the results are unsurprising.

definition of an ESS (if one treats a choice of learning rule as a choice of strategy). It would be strange to say that GRL rules are evolutionarily stable, though. In principle, given this set-up, any learning rule (like GRL) that has not gotten to the optimal outcome in the short time period could be outperformed, and so invaded, by a learning rule that does better in that same time period.[30] This, however, does not really matter. The point is not that a particular rule for generalization will be stable, but rather that this type of stability analysis ignores some of the most evolutionarily relevant features of learning rules, in this case a need for speed. Maynard-Smith [21] and Harley [12] are not wrong in thinking that there should be selection pressure for rules that learn ESSes, but just wrong in thinking that this is the only, or the most important type of selection pressuring bearing on learners.

7 Conclusion

To conclude, I will discuss how the results of this paper inform game theory and evolutionary game theory, but first, a word should be said about the proposed interpretation of the state spaces of approximation games.

I pointed out in section 2 that these state spaces should be thought of as perceptual rather than external because generalization happens over perceptually similar states. Given that similarity is built into the approximation game through the payoff structure, however, this interpretation assumes that perceptually similar states will always get similar payoffs when responded to with similar actions. At first consideration, this assumption may seem problematic. It should be noted, though, that perceptual similarity structures themselves evolve. O'Connor [24] argues that in models of the evolution of perceptual categorization, real world states that actors can respond to in similar ways evolve to be perceptually similar. If this is right, it may be reasonable to assume that perceptual similarity (usually) tracks payoff similarity. This line of thinking points to a way in which the exploration of generalization in this paper is incomplete, though. Generalization happens over perceptual states, and will only be successful if the similarity structure of these perceptual states is arranged so that perceptually similar things can be reacted to similarly. In this way, the evolution of generalization arguably cannot be *fully* understood without also understanding the evolution of perceptual similarity.

I will now return to how this exploration of the evolution of learning generalization informs evolutionary game theory. First, and most importantly, the assumption that the short term performance of learning rules can be ignored in evolutionary analyses is a bad one. This assumption is inconsistent with other assumptions made about

---

[30] In fact the real world behavior of learning discrimination points towards a possibility for such an improved rule. Previous investigations into animal learning indicate that when it is relevant from a payoff perspective for organisms to discriminate between states, they learn to do so [20]. In fact, generalization and discrimination can be seen as two sides of a coin. The former allows animals to extend successful behaviors to possibly relevant scenarios, the second allows animals to trim these behaviors back if they are not applicable [28]. This combination of behaviors could be modeled with a learning rule that combines the best aspects of GRL and Herrnstein reinforcement learning. Learners begin by generalizing, but eventually stop generalizing and develop more precise strategies. In fact, the models developed here help illuminate why learning discrimination is important—it helps organisms avoid sub-optimal behaviors developed when generalizing, and can allow actors to move closer to ESSes.

the evolution of learning, in particular that learning should be expected to evolve in variable environments. It is an assumption that matters, because, as shown here, when the short term success of learning rules is taken into account, evolutionary outcomes are significantly impacted. And, as this paper shows, if this assumption is maintained, evolutionary game theoretic models are unable to account for the evolution of generalization. When the assumption is dropped, on the other hand, evolutionary game theoretic models *can* successfully account for this highly successful real world learning behavior. As such this case illustrates how the long term learning assumption is not just intuitively suspect, but can actually lead an evolutionary analysis significantly astray.

Past investigations into the evolution of learning rules have been used to justify assumptions about equilibrium play in game theory (see, for example, Maynard-Smith [21]). The results here indicate that a better understanding of the evolution of learning does not support this justification. Although there should be selection pressure for learning rules to reach ESSes, there should also be selection pressure for rules that learn quickly. When these desiderata are at odds, as is the case with learning generalization, non-equilibrium behavior should be expected in the real world. Even if real world actors eventually learn to discriminate between relevantly different states, and so mitigate the sub-optimal effects of generalization, while learning progresses (which should be a non-trivial proportion of the time if actors face heterogenous environments) non-equilibrium and thus non-optimal behavior should be expected.

In recent years, the tradition of depending on ESS methodology in evolutionary analyses has come under fire. The results presented here are one more example of a case where a dynamical investigation reveals important insights into evolutionary processes that ESS analysis misses. As discussed, simply identifying which learning rules are evolutionarily stable in the sense that they lead to ESSes misses important differences between the processes that actors employing these rules undergo, and thus misses evolutionarily relevant information. This analysis thus gives further reason to be very careful when applying ESS methodology to complicated evolutionary scenarios.

References

1. Barrett JA (2007) Dynamic partioning and the conventionality of kinds. Philosophy of Science 74(527-546)
2. Barrett JA (2009) The evolution of coding in signaling games. Theory and Decision 67:223–237
3. Barrett JA (2014) Description and the problem of priors. Erkenntnis Doi: 10.1007/s10670-014-9604-2
4. Barrett JA, Zollman K (2008) The role of forgetting in the evolution and learning of language. Journal of Experimental and Theoretical Artificial Intelligence 21(4):293–309

5. Bereby-Meyer Y, Erev I (1998) On learning to become a successful loser: a comparison of alternative abstractions of learning processes in the loss domain. Journal of Mathematical Psychology 42:266–286

6. Dunlap AS, Stephens DW (2009) Components of change in the evolution of learning and unlearned preference. Proceedings of the Royal Society 276:3201–3208

7. Gärdenfors P (2000) Conceptual Spaces: on the geometry of space. MIT Press, Cambridge, MA

8. Ghirlanda S, Enquist M (2003) A century of generalization. Animal Behaviour 66(1):15–36

9. Gigerenzer G, Gaissmaier W (2011) Heuristic decision making. Annual Review of Psychology 62:451–482

10. Gigerenzer G, Selten R (eds) (2001) Bounded rationality: The adaptive toolbox. MIT Press, Cambridge, MA

11. Godfrey-Smith P (2002) Environmental complexity and the evolution of cognition. The evolution of intelligence pp 233–249

12. Harley CB (1981) Learning the evolutionary stable strategy. Journal of Theoretical Biology 89:611–633

13. Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning: Data mining, inference, and prediction. Statistics, Springer

14. Herrnstein R (1970) On the law of effect. Journal of the Experimental Analysis of Behavior 13(2):243–266

15. Huttegger SM, Zollman KJS (2011) Language, Games, and Evolution, Springer-Verlag, Berlin Heidelberg, chap Signaling Games: Dynamics of Evolution and Learning, pp 160–176

16. Jäger G (2007) The evolution of convex categories. Linguistics and Philosophy 30:551–564

17. Johnston TD (1982) The selective costs and benefits of learning: an evolutionary analysis. In: Rosenblatt JS (ed) Advances in the Study of Behavior, vol 12, Academic Press, New York

18. Krantz DH, Luce RD, Suppes P, Tversky A (1971) Foundations of Measurement, vol 1. Dover Publications, Mineola, NY

19. Laslier JF, Topol R, Walliser B (2001) A behavioral learning process in games. Games and Economic Behavior 37(340-366)

20. Mackintosh NJ (1974) The psychology of animal learning. Academic Press

21. Maynard-Smith J (1982) Evolution and the theory of games. Cambridge University Press, Cambridge

22. Mednick SA, Freedman JL (1960) Stimulus generalization. Psychological Bulletin 57(3):169–200

23. O'Connor C (2013) The evolution of vagueness. Erkenntnis

24. O'Connor C (2014) Evolving perceptual categories. Philosophy of Science

25. Plotkin HC, Odling-Smee FJ (1979) Learning, change, and evolution: an enquiry into the teleonomy of learning. In: Rosenblatt JS (ed) Advances in the Study of Behavior, vol 10, Academic Press, New York

26. Roth AE, Erev I (1995) Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. Games and Economic Be-

havior 8:164–212
27. Shepard RN (1987) Toward a universal law of generalization for psychological space. Science 237(4):1317–1323
28. Shettleworth SJ (2009) Cognition, evolution, and behavior. Oxford University Press
29. Smead R (2012) Game theoretic equilibria and the evolution of learning. Journal of Experimental and Theoretical Artificial Intelligence 24(3):301–313
30. Smead R (2013) The role of social interaction in the evolution of learning
31. Smead R, Zollman K (2009) The stability of strategic plasticity, working paper
32. Stephens DW (1991) Change, regularity, and value in the evolution of animal learning. Behavioral Ecology 2:77–89
33. Watson J, Rayner R (1920) Conditioned emotional reactions. Journal of Experimental Psychology 3(1):1–14
34. Zollman K, Smead R (2010) Plasticity and language: an example of the baldwin effect? Philosophical Studies 147(1):7–21