The Stability of Strategic Plasticity

Rory Smead
University of California, Irvine

Kevin J.S. Zollman *
Carnegie Mellon University

April 23, 2008

Abstract

Recent research into the evolution of higher cognition has piqued an interest in the effect of natural selection on the ability of creatures to respond to their environment (phenotypic plasticity). It is believed that environmental variation is required for plasticity to evolve in cases where the ability to be plastic is costly. We investigate one form of environmental variation: frequency dependent selection – where an individual's fitness is determined by the proportions of other types in the population. Using tools in game theory, we investigate a few models of plasticity and outline the cases where selection would be expected to maintain it. Ultimately we conclude that frequency dependent selection is likely insufficient to maintain plasticity given reasonable assumptions about its costs. This result is very similar to one aspect of the well-discussed Baldwin effect, where plasticity is first selected for and then later selected against. We show how in these models one would expect plasticity to grow in the population and then be later reduced. Ultimately we conclude that if one is to account for the evolution of learning in this way, one must appeal to a very particular sort of external environmental variation.

*The authors would like to thank Kevin Kelly for providing the central insights which led to the development of several of the proofs. Helpful comments were also provided by Brian Skyrms, Steve Frank, Kyle Stanford, and the participants of the Skyrmsfest Workshop in Laguna Beach.

1 Introduction

Humans and many other animals are behaviorally adaptive, they modify their behavior in response to the environment. This flexibility is often placed in contrast to more rigidly specified behaviors which are not capable of changing across environmental circumstances. The evolution of this ability to modify one's behavior to the environment, known as plasticity, has been the source of some discussion.[1]

It is generally believed that plasticity will be selectively advantageous when there is environmental variation. Environmental variation, very broadly conceived, can come in many forms. It may come from a source external to the population, such as seasonal or ecological change. Or, it may be due to changes in the population itself (as in frequency-dependent selection). Variation of the first kind has been investigated in some detail (Godfrey-Smith, 1996, 2002; Sterelny, 2003; Ancel, 1999) and it can be shown that behavioral plasticity will be advantageous when there is such external variation.

One might suppose that frequency dependent selection is merely a special case of environmental variation. There are several reasons to suppose that frequency-dependent selection alone may be sufficient to provide such selective pressure. First, if we imagine that individuals interact each other and there is variation within the population, then each interaction may be different than the previous one. Second, as the population evolves over time, the distribution of types in the population changes and thus the sorts of interactions an individual is likely to have will change as well. These considerations might lead one to suppose that frequency dependent selection represents an extreme version of external environmental variation.

This paper investigates this possibility using evolutionary game theory.[2] We find that populations of plastic individuals will not be maintained by evolution, and often plasticity will be eliminated entirely. This result suggests that a particular sort of environmental variation is necessary to sustain plasticity – not any sort of variation will do. More specifically, the variation caused by frequency dependent selection is often insufficient to result in plasticity in strategic situations that can be modeled as games. Further-

[1] Behavioral plasticity can be distinguished from developmental plasticity where the way an organism develops varies with its environment. For the purposes of this paper, we will be focusing on behavioral plasticity.

[2] The evolution of plasticity in games has been investigated already in a different context by Harley (1981).

more, when we examine the evolutionary dynamics in these settings, we find that plasticity is often initially selected for only to be reduced later on. This is similar to one aspect of the Baldwin effect, where acquired characteristics (in this case, social behaviors) in the population are replaced by similar hereditary characteristics.[3]

We begin by providing the necessary background in game theory in section 2. After these preliminaries, we present three models that include a plastic phenotype (an adaptive strategy). While both include a plastic strategy, they differ in the method by which costs are associated with that phenotype. We analyze these models and present our primary results in sections 3 and 4. Finally section 5 concludes.

2 Game Theory

A game is a mathematical object which specifies a set of players $P = \{1, \ldots, n\}$, a set of strategies for each player S_i (where $i \in P$), and a payoff function for each player, $\pi_i : S_1 \times S_2 \times ... \times S_n \to \mathbb{R}^n$. In keeping with biological game theory, we will here treat the payoff as representing a strategy's fitness.

Here we will restrict ourselves to considering one class of games, the finite symmetric two-player game. A game is symmetric just in case each player has the same set of strategies ($S_i = S_j$ for all i, j) and the payoff function depends only on the strategy chosen not the identity of the player. We can then represent the payoff function as a single function $\pi(x, y)$ which represents the fitness of playing strategy x against strategy y.

Since we are restricting ourselves to considering only finite games, S has only finitely many members. We can, however, extend the strategy space to include mixed strategies (strategies which involve choosing different elements of S with some probability). Now an agents strategy is represented by a probability distribution over S and we can then consider the expected utility from each strategy played against another, represented by $u(x, y)$.

Given any game, we can construct a best response correspondence for that game. For some given (possibly mixed) strategy, x, $B(x)$ represents the

[3] Ancel (1999) refers to this as the "Simpson-Baldwin effect" from the formulation of the Baldwin effect by Simpson (1953). We will follow Ancel and use the phrase "Simpson-Baldwin effect" when discussing the relationship between our results and the Baldwin effect in section 4.

set of strategies which do best against x. Formally,

$$B(x) = \{y | u(y, x) \geq u(y', x) \forall y'\} \tag{1}$$

A set of strategies s^* is a Nash equilibrium if and only if each strategy in that set is a best response to the others. Formally, s^* is a Nash equilibrium when $s^*_i \in B(s^*_{-i})$ for all players i (where s^*_i represents player i's strategy in s^* and s^*_{-i} represents the other player's strategy). We will represent the set of all Nash equilibria in a game by NE.

From an evolutionary perspective Nash equilibria are insufficient to represent stable states of populations since there are some Nash equilibria which one would expect to be invaded by mutation. As a result, a slightly more limiting solution concept was devised for this purpose, the Evolutionarily Stable Strategy (Maynard Smith and Price, 1973; Maynard Smith, 1982).

Definition 1. *A strategy set s^* is an evolutionarily stable strategy (ESS) if and only the following two conditions are met*

- $u(s^*, s^*) \geq u(s, s^*)$ *for all alternative strategies* s ***and***
- *If* $u(s^*, s^*) = u(s, s^*)$*, then* $u(s^*, s) > u(s, s)$.

This more restrictive solution concept captures the idea that a population is stable to invasion by mutation. Evolutionarily stable strategies can be either pure or mixed strategies. Traditionally a mixed strategy represents directly the strategy of an individual who randomizes their behavior, but this interpretation can be implausible in some contexts. Instead one can interpret a mixed strategy as a population state, where a certain proportion of a population is playing each strategy. Here uninvadability represents a sort of asymptotic stability of the population under mutation (Weibull, 1995).

It is important to note that in any mixed strategy equilibrium s^*, if two pure strategies, s and s', are both played with positive probability it must be the case that $u(s, s^*) = u(s', s^*)$. (Otherwise one would have an incentive to play the better of the two.) As a result, a mixed strategy equilibrium is an ESS only if $u(s^*, s) > u(s, s)$ for all s that are present in the mixture.

There are two features of equilibria which will be of use in the discussion. First, is the notion of a symmetric equilibrium. A Nash equilibrium is symmetric when both players are playing the same strategy. Only symmetric Nash equilibria can be evolutionarily stable (although not all are). Second, a Nash equilibrium (or ESS), N, is pareto superior to another Nash equilibrium, O, if $u(N_1, N_2) > u(O_1, O_2)$ and $u(N_2, N_1) > u(O_2, O_1)$ – e.g. both players do better in N than in O.

Although somewhat limited, games with only two strategies (so called 2×2 games) are often used to capture many different types of behavior. When possible we try to present general results, but for some propositions we must restrict ourselves to 2×2 games. While there are four payoffs in a symmetric 2×2 game, we can represent all such games with two parameters without losing best response and dominance relations (Weibull, 1995). This characterization leaves us with only three classes of games which are characterized by different equilibria. Coordination games have two pure strategy Nash equilibria which are symmetric. Hawk-Dove games have two asymmetric pure strategy Nash equilibria, and Prisoner's dilemmas have only one Nash equilibrium.

3 The reduction of strategic plasticity

With these preliminaries in hand we can return to considering the possibility of strategic plasticity. Suppose there is a population of players and individuals in this population are randomly paired to play a repeated game against one another. There are phenotypes which correspond to each of the pure strategies in the underlying game and one phenotype which corresponds to "learning" (our plastic strategy). The learning phenotype is capable of adapting itself to do well against the individual with which it is paired. Traditionally it is assumed that plasticity is costly. Perhaps maintaining or developing the ability to adapt is costly, or alternatively there is an intrinsic cost associated with making errors in the process of learning.[4] Our first two models follow this first suggestion that learning has some exogenous cost, that is there is some basic fitness cost to being able to adapt which is not a feature of the game or the learning process. In the third model we will capture the suggestion that the cost of learning is the result of errors in the learning process where learners occasionally fail to play the best response.

Ultimately we find that populations of learners can only rarely be main-

[4] Ancel (1999) describes the following examples "In bacteria, for example, plasticity requires genetic machinery whose costs are in terms of increased replication time. Learning-based plasticity may entail energetic costs of searching RNA molecules with multiple configurations may trade accuracy for the potential for variable binding" (199).

tained by evolution, and that in many cases, learning will be totally eliminated by the evolutionary process.

3.1 Exogenous cost

Our model will take a base game whose strategies represent phenotypes in a population and extend it by including a new phenotype $\mathcal{L}$, which represents the plastic individual. We will assume that the plastic individual, or learner, is capable of learning the best response against others and learns to play a Nash equilibrium against itself.[5] We will pick members of a population to play the game repeatedly against each other, and assume that the game is repeated sufficiently often that the errors generated by the process of learning are removed (an assumption that will be relaxed in section 3.2). However, we will assume that there is some fixed fitness cost which represents the costs imposed by developing, maintaining, or implementing an adaptive strategy. These costs are taken to be fixed regardless of the underlying game being played.

Formally, we will represent these assumptions by constructing a new game, G^L by defining the utility function u^L based on an underlying game G. We will restrict ourselves to considering cases where every strategy in G has a unique best response.[6] The payoff function $u^L(\cdot)$ is defined as follows:

1. $u^L(x, y) = \pi^G(x, y)$ if $x, y \in S$
2. $u^L(\mathcal{L}, y) = \pi^G(B(y), y) - c$
3. $u^L(x, \mathcal{L}) = \pi^G(x, B(x))$ if $x \in S$
4. $u^L(\mathcal{L}, \mathcal{L}) = \sum_{s \in NE} \alpha(s) \left(\frac{1}{2}u^G(s_1, s_2) + \frac{1}{2}u^G(s_2, s_1) - c\right)$

$c > 0$ represents the cost of learning. So that the cost of learning does not effect the results too substantially, we will restrict c be smaller than the difference between any two payoffs in the game. In condition 4, we assume that there is a probability distribution, α over the possible Nash equilibria of

[5]This has already excluded a set of possible learning rules which do not necessarily converge to Nash equilibria. Extending this model to those cases, while interesting, is beyond the scope of this paper.

[6]While this may seem a substantive assumption, we are only excluding an very small set of games. Any game that features a strategy which does not have a unique best response is not robust to epsilon perturbations of the payoffs.

the game which represents the probability that learning will converge to that equilibrium. The remainder of that function represents the idea that, given one converges to a Nash equilibrium, it is random which role one plays.

u^L has not been fully specified at this point. We have not defined $u^L(s, \mathcal{L})$ for a mixed strategy s. If there is a unique best response to the mixture we will let $u^L(s, \mathcal{L}) = u^G(s, B(s))$. But if s is a part of a mixed strategy Nash equilibrium then any mixture over the pure strategies in s is a best response to s. This is a more complicated situation that is discussed in more detail below. For the time being we will leave this case underspecified.[7]

In our investigation of strategic plasticity, we will first consider whether or not learning could be sustained in the population. We find that it is, a population composed of only $\mathcal{L}$ players cannot be stable for many base games G.

Proposition 1. *For all games G without a pareto dominant mixed strategy Nash equilibrium, $\mathcal{L}$ is not an ESS of the game G^L.*

Proof. We will show that $\mathcal{L}$ is not an ESS by finding a strategy s such that $u^L(s, \mathcal{L}) > u^L(\mathcal{L}, \mathcal{L})$. Let $s = \max_{s'} u^G(s', B(s'))$, i.e. the strategy which does best against its best response among all strategies in G. Consider the Nash equilibrium N of G which has the highest payoff to one of the players, $u^G(s, B(s)) \geq u^G(N_i, N_{-i})$, by definition of s. By hypothesis, s is not a mixed strategy. By condition 4, $u^G(N_i, N_{-i}) > u^L(\mathcal{L}, \mathcal{L})$. So, as a result $u^L(s, \mathcal{L}) > u^L(\mathcal{L}, \mathcal{L})$. $\square$

No symmetric 2×2 game can have a pareto dominant mixed strategy Nash equilibrium and so $\mathcal{L}$ cannot be an ESS for any 2×2 base game. However, there are larger games that satisfy this condition. As an example consider Rock-Paper-Scissors, pictured in figure 1. In this game the unique Nash equilibrium is a mixed equilibrium where each strategy is played with equal probability. In this equilibrium the payoff to both players is 0. Consider the extension of this game to include $\mathcal{L}$. The payoff to a population of learners is $-c$. But any invading pure strategy receives -1, since this is the payoff

[7] The extended game G^L has a an odd feature: for some mixed strategies m of G^L, $u^L(m, \mathcal{L})$ will not be a weighted average of $u^L(s, \mathcal{L})$ for each s in m. This occurs because $\mathcal{L}$'s response to a mixture is different than its response to a pure type. It becomes important, then, that we be careful to distinguish a monomorphic population of mixing types and a polymorphic population of non-mixing types. When necessary mention will be made of the intended interpretation.

of that strategy against its best response. So, no pure strategy can invade. However, the mixed strategy which plays *Rock*, *Paper*, and *Scissors* with equal probability does better against the population (0) than the population does against itself, ($-c$).

	Rock	*Paper*	*Scissors*
Rock	$0, 0$	$-1, 1$	$1, -1$
Paper	$1, -1$	$0, 0$	$-1, 1$
Scissors	$-1, 1$	$1, -1$	$0, 0$

Figure 1: Rock-Paper-Scissors

Whether proposition 1 holds in these cases depends critically on what learning does when it encounters a mixture that constitutes a mixed strategy Nash equilibrium. It is a feature of such mixtures that they make the opponent indifferent between several pure strategies, and by extension indifferent between all mixtures over those pure strategies. Different plausible learning rules will behave in different ways, and thus may effect the stability of $\mathcal{L}$.

	Rock	*Paper*	*Scissors*
Rock	$0, 0$	$-1, 2$	$1, -1$
Paper	$2, -1$	$0, 0$	$-1, 1$
Scissors	$-1, 1$	$1, -1$	$0, 0$

Figure 2: Modified Rock-Paper-Scissors

For example consider the modified Rock-Paper-Scissor game in figure 2. Suppose one player uses the strategy to play *Rock* with probability $\frac{1}{4}$, *Paper* with probability $\frac{1}{3}$, and *Scissors* with probability $\frac{5}{12}$. This renders any strategy by the opponent a best response. We have not specified what a learner will do in this circumstance. Suppose the learner plays *Paper* with probability one. This yields an expected payoff of $-\frac{1}{12}$ for the mixed player. The learners are able to coordinate on the mixed strategy equilibrium with themselves, and so as a result achieve a higher payoff than the mixed players achieve against the learners. This renders $\mathcal{L}$ an ESS.

Of course, this would be a rather bizarre learning rule: one that is able to coordinate on the strategy which would harm its opponent the most, but was otherwise equal to itself. Suppose instead that learners randomly chose a strategy from the available ones.[8] Now, in the modified Rock-Paper-Scissors game, the mixture can invade a population of learners.

Ultimately this shows that in games of this sort, where there are pareto dominant mixed strategy Nash equilibria, the choices we make for the learning rule will have important effects on the stability of learning. It is possible to make learning stable, but it does require some specific choices about the response of the learning rule when all strategies are of equal value.

This caveat to the side, proposition 1 does not spell total doom for strategic plasticity. In some games, plasticity can be maintained in some proportion in a population. Consider the game Hawk-Dove pictured in figure 3. This game, common in evolutionary biology, models animal conflicts where individuals can choose to either escalate the conflict ($Hawk$) or back down ($Dove$).

	$Hawk$	$Dove$
$Hawk$	$0, 0$	$3, 1$
$Dove$	$1, 3$	$2, 2$

Figure 3: Hawk-Dove

The only evolutionarily stable strategy of this base game is a mixed strategy with some proportion of both Hawks and Doves. Suppose a polymorphic population of both types which corresponds to this mixture; this mixture is not stable in the extended game (with $\mathcal{L}$) since $\mathcal{L}$ does better against the mixed population than the population does against itself. $\mathcal{L}$ outperforms the pure strategy types because it learns to play the best response to each pure strategy; it plays $Dove$ against $Hawk$ and $Hawk$ against $Dove$. While learning invades this mixture, we know from proposition 1 that it cannot be an ESS itself. Instead there is a stable mixture where $\mathcal{L}$ and $Hawk$ are both represented. In this population, some proportion of individuals are plastic while some proportion are not. Natural selection will sustain this polymorphism, and so as a result we do not have complete elimination of plasticity.

Hawk-Dove represents one of three classes of symmetric 2x2 games. The others are represented by games which are dominance solvable (like the Prisoner's Dilemma) and coordination games (like the Stag Hunt). In these other two cases learning cannot be sustained in any proportion. In the first class

[8] For instance this would be the case for Herrnstein reinforcement learning (cf. Roth and Erev, 1995).

(dominance solvable games) one pure strategy is the best response to all others, since learning will learn to play that strategy but incurs some cost, it fares strictly worse in all scenarios than does the pure strategy itself.

In coordination games the situation is rather more complex, but we can again show that learning cannot be sustained. These games feature two pure strategies, both of which are best responses to themselves. We can show that there cannot be a mixed strategy ESS which contains both learning and any strategy which is a best response to itself.

Proposition 2. *Suppose a symmetric two player game G, a strategy s of G such that s is a best response to itself, and mixed strategy Nash equilibrium s^* of G^L which includes s and $\mathcal{L}$. s^* is not an ESS.*

Proof. Suppose s is a best response to itself, s is played with non-zero probability in s^*, and that s^* is an ESS. Since s^* is a mixed strategy Nash equilibrium in which s is represented $u(s, s^*) = u(s^*, s^*)$. By hypothesis s^* is an ESS, so it must be the case that $u(s^*, s) > u(s, s)$. Expanded this says,

$$\sum_{i \in S - \{\mathcal{L}, s\}} p_i \pi(i, s) + p_s \pi(s, s) + p_{\mathcal{L}}[\pi(s, s) - c] > \pi(s, s) \tag{2}$$

Where p_i represents the probability i is played in s^*. Reducing this equation gives us,

$$\sum_{i \in S - \{\mathcal{L}, s\}} p_i \pi(i, s) - p_{\mathcal{L}} c > (1 - p_s - p_{\mathcal{L}}) \pi(s, s) \tag{3}$$

Which is equivalent to,

$$\sum_{i \in S - \{\mathcal{L}, s\}} \frac{p_i}{1 - p_s - p_{\mathcal{L}}} (\pi(i, s) - \pi(s, s)) - \frac{p_{\mathcal{L}} c}{1 - p_s - p_{\mathcal{L}}} > 0 \tag{4}$$

Since s is a best response to itself, $\pi(i, s) - \pi(s, s) \leq 0$, and as a result equation 4 cannot be satisfied. $\square$

This shows that Hawk-Dove like games represent a sort of special case for sustaining learning. In any game where a strategy is a best response to itself, learning can only be sustained in populations where that strategy is absent.

So far we have assumed that the only costs imposed by learners was on themselves. This coincides with understanding costs as being caused by maintenance or development of the ability to learn, However, we might open the door for $\mathcal{L}$ as an ESS by dropping this assumption. We will generalize the notion of cost by allowing the costs to be imposed on all interactions. We will introduce cost function $c : S \times S \to \mathbb{R}$ which will represent the cost for one strategy against another.

We are now generating a different game with the following payoff function:

1. $u^L(x, y) = \pi^G(x, y) - c(x, y)$ if $x, y \in S$
2. $u^L(\mathcal{L}, y) = \pi^G(B(y), y) - c(\mathcal{L}, y)$
3. $u^L(x, \mathcal{L}) = \pi^G(x, B(x)) - c(x, \mathcal{L})$ if $x \in S$
4. $u^L(\mathcal{L}, \mathcal{L}) = \sum_{s \in NE} \alpha(s) \left(\frac{1}{2}u^G(s_1, s_2) + \frac{1}{2}u^G(s_2, s_1) - c(\mathcal{L}, \mathcal{L})\right)$

Given this generic representations of costs, we may ask: what is required to make $\mathcal{L}$ an ESS? We will first consider cases where the cost imposed is relatively constant across all strategies.

The following two propositions demonstrate that learners must learn to play the socially optimal and equitable Nash equilibrium (if one exists), in order for $\mathcal{L}$ to be an ESS.[9]

Proposition 3. *Suppose there is no pareto dominant mixed strategy Nash equilibrium in G. If c is such that $|c(x, y) - c(w, z)|$ is sufficiently small for all w, x, y, z and $\mathcal{L}$ is an ESS, then $\alpha(N) \approx 1$ for the socially optimal Nash equilibrium N.*

Proof. Let $N \in NE$ be the Nash equilibrium that is best for one of its players, that is for N_i, $u(N_i, N_{-i}) \geq u(N'_j, N'_{-j})$ for all $N' \in NE$. By the definition of ESS, $u^L(\mathcal{L}, \mathcal{L}) \geq u^L(N_i, \mathcal{L})$. Since N is a Nash equilibrium N_{-i} is a best response to N_i, so $u^L(N_i, \mathcal{L}) = u^G(N_i, N_{-i}) - c(N_i, \mathcal{L}) \leq u^L(\mathcal{L}, \mathcal{L})$. Substituting this yields,

$$\sum_{s \in NE} \alpha(s) \left(\frac{1}{2}u^G(s_1, s_2) + \frac{1}{2}u^G(s_2, s_1) - c(\mathcal{L}, \mathcal{L})\right) \geq u^G(N_i, N_{-i}) - c(N_i, \mathcal{L}) \tag{5}$$

Since we are assuming that costs are relatively constant, we will assume they are equal. As a result the average on the left side of 5 must be greater than

[9] A Nash equilibrium is socially optimal if the sum of payoff to both players is larger (or as large) in that Nash equilibrium than in any other.

or equal to the payoff to the best possible payoff in a Nash equilibrium. This can only be obtained when the socially optimal Nash equilibrium is the best possible payoff for any strategy in equilibrium and α assigns the socially optimal Nash equilibrium sufficiently high probability. $\square$

When costs are relatively uniform, this proposition entails that the function α assigns near probability 1 to one Nash equilibrium. In particular, it must assign probability 1 to the socially optimal Nash equilibrium – i.e., the equilibrium which is best for both players considered together. This restricts the class of possible learning rules which are ESS to a very small set. In addition to this restriction on learning rules there is also a restriction on the underlying game.

Proposition 4. *Suppose there is no pareto dominant mixed strategy Nash equilibrium in G. If c is such that $|c(x,y)-c(w,z)|$ is sufficiently small for all w,x,y,z and $\mathcal{L}$ is an ESS, then the Nash equilibrium N played by $\mathcal{L}$ against itself is such that $u^G(N_1,N_2) \approx u^G(N_2,N_1)$.*

Proof. By proposition 3, $u^L(\mathcal{L},\mathcal{L}) = \frac{1}{2}u^G(N_1,N_2) + \frac{1}{2}u^G(N_2,N_1) - c(\mathcal{L},\mathcal{L})$. Let $u^G(N_1,N_2) \geq u^G(N_2,N_1)$. Since $\mathcal{L}$ is an ESS, it must be the case that:

$$\frac{1}{2}u^G(N_1,N_2) + \frac{1}{2}u^G(N_2,N_1) - c(\mathcal{L},\mathcal{L}) \geq u^G(N_1,N_2) - c(N_1,\mathcal{L}) \tag{6}$$

Again, ignoring c, this can only be satisfied when $u^G(N_1,N_2) = u^G(N_2,N_1)$. $\square$

These two propositions together entail significant restrictions both on the game and on the types of learning that may be stable. The game must feature a symmetric, socially optimal Nash equilibrium and the learning phenotype must learn to play that equilibrium almost all of the time.

But what if the costs imposed are not uniform? Could learning be sustained in all games? The following proposition follows immediately from the definition of ESS.

Proposition 5. *Suppose there is no pareto dominant mixed strategy Nash equilibrium in G and let s be a strategy of G which does the best against its best response. If $\mathcal{L}$ is an ESS, then either $c(\mathcal{L},\mathcal{L}) < c(s,\mathcal{L})$ or $c(\mathcal{L},\mathcal{L}) = c(s,\mathcal{L})$ and $c(\mathcal{L},s) < c(s,s)$.*

For $\mathcal{L}$ to be an ESS it must be more costly to be a non-learner than a learner in a world of learners. The key to the evolutionary stability of learning is the cost of not being a learner when interacting with learners. (This is not to say, however, that it is intrinsically more costly to be a non-learner than a learner – although such a case would satisfy the definition.) There are reasons one may think this may be the case (i.e. that $c(\mathcal{L}, \mathcal{L}) \leq c(s, \mathcal{L})$). For instance, the cost to learners may come from the time spent playing non-best-responses while exploring the game. But, this exploration does not only affect the learner, but also the learner's opponent. Thus, depending on the specifics of the game this exploration may be more costly to those interacting with learners than the learners themselves.[10] In section 3.2, we will consider an idealized version of these "endogenous" costs.

3.2 Endogenous Costs of Learning

If we imagine that the $c(x, y)$ is generated endogenously (by $\mathcal{L}$ playing other possible strategies while learning), we can incorporate it directly into the utility function of the players. We will express the endogenous cost of learning as the difference between the standard payoff of the game G^L and some weighted average of the payoffs received when the learners "explore" other strategies. This means that $c(s, s') = 0$ for any $s, s' \in S$ since neither player will be changing strategies and will simply receive the payoff $u^L(s, s')$ and we only have to examine cases where $\mathcal{L}$ is involved.

For sake of tractability, we will focus on symmetric 2×2 games and restrict ourselves to pure-strategies and learners that always learn to play pure strategy Nash equilibria (this is because it is not clear how how "deviations" from mixed strategies ought to be calculated). We will ask the following question: can $\mathcal{L}$ be an ESS when we express $c(x, y)$ endogenously? Let ϵ be the total proportion of stage-game plays where $\mathcal{L}$ deviates from their normal behavior s_i^*.[11] Let $0 < \epsilon < 0.5$. The endogenous cost can now be expressed

[10] Another reason might be that several different games are being played and it is difficult to distinguish them. This would be a form of "external" variation similar to a change in payoff structure.

[11] For tractability reasons, we assuming that ϵ is the same regardless of the opponent, which is not necessarily true of all learning rules. For instance, Herrnstein reinforcement learning may have a higher ϵ against itself than against pure strategies. A detailed look at Herrnstein reinforcement learning in this evolutionary setting is the topic of a future paper.

as part of the utility functions:

1. $u^L(\mathcal{L}, s) = (1-\epsilon)\pi^G(s_i^*, s) + \epsilon\pi^G(-s_i^*, s)$
2. $u^L(s, \mathcal{L}) = (1-\epsilon)\pi^G(s, s_i^*) + \epsilon\pi^G(s, -s_i^*)$
3. $u^L(\mathcal{L}, \mathcal{L}) = \frac{(1-\epsilon)^2}{2}(\pi^G(s_i^*, s_j^*) + \pi^G(s_j^*, s_i^*)) + (\epsilon - \epsilon^2)\pi^G(s_i^*, -s_j^*) + (\epsilon - \epsilon^2)\pi^G(-s_i^*, s_j^*) + \frac{\epsilon^2}{2}(\pi^G(-s_i^*, -s_j^*) + \pi^G(-s_j^*, -s_i^*))$

When costs to learning are made endogenous in this way, there is an example of a 2×2 game where $\mathcal{L}$ is an ESS: a subset of coordination games (figure 4). To see this, let $a = 2$, $x = 2 - \delta$ and $b = 1$ where $\delta < \epsilon$ and each are close to 0. Further, let $(\mathcal{L}, \mathcal{L})$ result in "attempting" to play socially optimal NE (A, A). To see $\mathcal{L}$ is an ESS, we need only consider $u(\mathcal{L}, \mathcal{L})$ and $u(s, \mathcal{L})$ for $s = A, B$. With these values, letting $\delta \to 0$: $u^L(\mathcal{L}, \mathcal{L}) \approx (2 - 2\epsilon) + \epsilon^2$; $u^L(A, \mathcal{L}) \approx 2 - 2\epsilon$; $u^L(B, \mathcal{L}) \approx 1 + \epsilon$. Thus, $u^L(\mathcal{L}, \mathcal{L}) > u^L(A, \mathcal{L}) > u^L(B, \mathcal{L})$ and hence $\mathcal{L}$ is an ESS.

	A	B
A	a	0
B	x	b

Figure 4: A coordination game $(a > x, a > b > 0)$.

We can vary a, b, x, and ϵ to create a class of examples. $\mathcal{L}$ will be an ESS in a similar coordination game any time the following condition is satisfied:

$$a - x < \frac{b\epsilon}{1-\epsilon} \tag{7}$$

There are several things that are unusual about this set of examples where $\mathcal{L}$ is an ESS, however. First, there is a sense in which $\mathcal{L}$ is not *really* best responding to itself in these cases. The mixed strategy NE of this game can be expressed by playing A with the following probability:

$$p(A) = \frac{b}{a - x + b} \tag{8}$$

So, if $p(A)$ is less than this, the best response to that mixed strategy is the pure strategy B. Note that for $\mathcal{L}$ to be an ESS, we need $(1-\epsilon) < \frac{b}{a-x+b}$

which means that the exploration rate of $\mathcal{L}$ must be high enough that the best response to the *actual* behavior of $\mathcal{L}$ is B not A as the learners are "attempting" to play. Thus, there is a sense in which for $\mathcal{L}$ to be an ESS in this case, learners need to best respond to what other learners are "trying" to do and not their actual behavior.

Second, there is no other class of 2×2 game where $\mathcal{L}$ is an ESS.

Proposition 6. *If G is a 2×2 symmetric game of a form other than the coordination game above and we consider only pure strategies (and pure strategy equilibria) then L is not an ESS of G^L.*

Proof in appendix A.

3.3 Summary

Ultimately we have shown that plasticity can be sustained in only very rare circumstances. When there is an exogenous cost to learning that is higher than other strategies, a limited type of learning can only be an ESS in a very restricted set of games. When the exogenous costs are approximately equal to learners and those interacting with them, only specific learners in specific games can be ESS. The only case where learning can generally be an ESS is if the costs are greater to non-learners than to learners. While this is not impossible, it would require some argument to presume this is a widespread phenomenon.

With respect to endogenous costs modeled as errors in the game we found that learning cannot be sustained except in a very narrow class of 2×2 games for learners of a particular accuracy. Overall there are not many cases where learning can be sustained and we should expect a reduction of the frequency of plastic individuals. However, thus far we have only been considering the equilibria of these games, not the evolutionary dynamics. And, as a result, we have only considered stability. But we might wonder if plasticity is ever favored by natural selection – should we expect the frequency of plasticity to increase in a population out of equilibrium? We will now turn to this second consideration.

4 The Simpson-Baldwin effect

This history of the growth and later reduction of plasticity is one aspect of what is known as the Baldwin effect (Baldwin, 1896).[12] This aspect has come to be called the Simpson-Baldwin effect, following its later explication by Simpson (1953), "Characters individually acquired by members of a group of organisms may eventually, under the influence of selection, be reinforced or replaced by similar hereditary characters" (110).

In his discussion of the Simpson-Baldwin effect, Peter Godfrey-Smith (2003) poses an important challenge: Why should a population first go through the learning stage if there is a genotype that will encode the beneficial behavior? That is, why should we expect the initial increase in phenotypic plasticity, so that it could be later reduced? He considers a few answers to this question and concludes that the only plausible one is where the presence of plastic individuals *creates* the environment in which the less plastic individuals are superior. He cites Terrance Deacon's (1997) model for the evolution of language, whereby individuals who are plastic enough to acquire language create an environment where those who have "hard-wired" linguistic abilities thrive. Were it not for the presence of the plastic individuals, and thus the presence of other linguistically competent agents, the hard-wired types would have had no one to talk to and thus not be superior to non-linguistic types.

This is exactly the type of strategic situation considered by game theory. But so far we have only discussed how evolution would eliminate or reduce the number of adaptive individuals – the second component of the Simpson-Baldwin effect. The first component, that natural selection would first select for plasticity, is not uniformly common to our models. This is largely due to the fact that we have considered only equilibria and not dynamics. While we know that learning cannot be an ESS in our game, we do not know whether learning might increase in frequency or not.

Since learning plays the best response to every other strategy, it will often be the case that it initially does better than some arbitrary mixture. This is not assured because for some mixtures that are sufficiently close

[12] Discussions of the Baldwin effect often involve looking at how evolution affects an organism's norm of reaction and admit of a wide range of plasticity. Because we are focusing on behavioral plasticity in simple games, our discussion here will differ from the tradition and we will continue to assume individuals are either plastic (using the adaptive strategy $\mathcal{L}$ from above) or not.

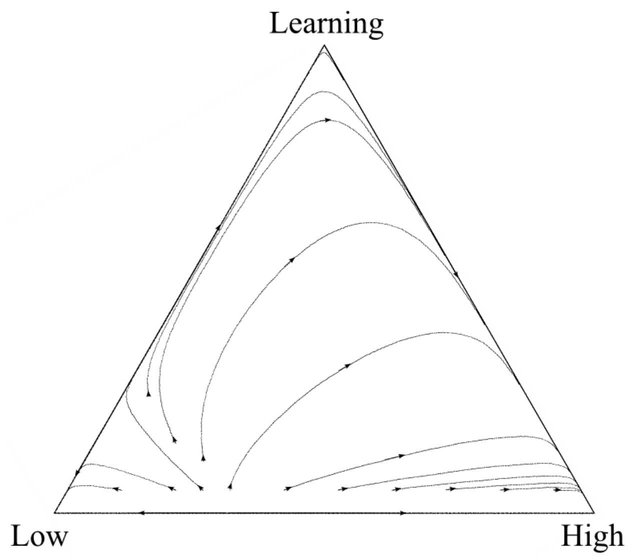

Figure 5: The dynamics of a coordination game

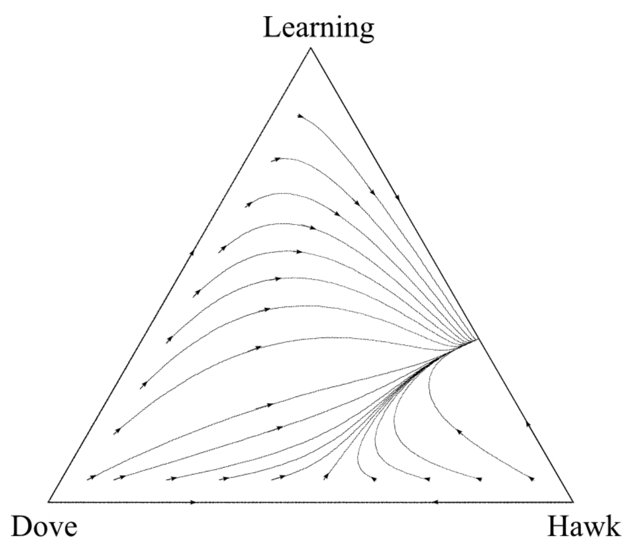

Figure 6: The dynamics of a Hawk-Dove game

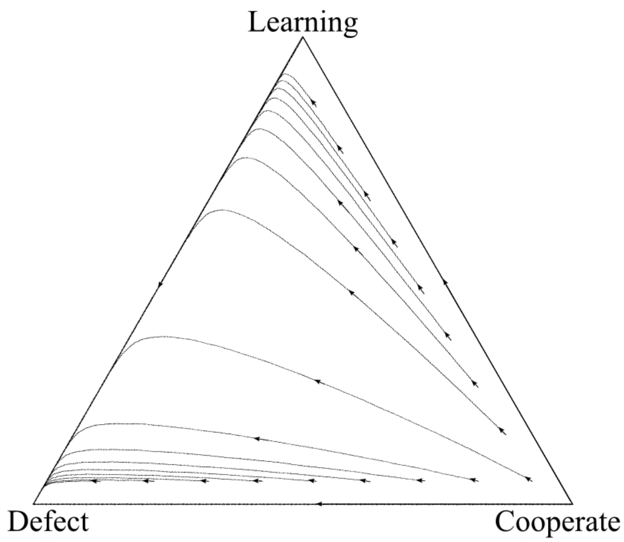

Figure 7: The dynamics of a prisoner's dilemma game

to an equilibrium the impact of the cost to learning, c, will be sufficient to make learning worse than the population. However, in many cases of interest it will be the case that learning is initially rewarded and then later harmed. Figures 5-7 show the three simplices which result from considering the three classes of symmetric 2x2 games.[13] In each of these cases we find that learning is initially benefited from most starting points – this is represented by moving toward the top of the figure. In all three cases we see that once learning represents a significant portion of the population, another strategy can invade.

Figure 5 (a coordination game) and figure 7 (a prisoner's dilemma) both represent striking examples of the Simpson-Baldwin effect. Many trajectories follow a path that initially provides substantial gains for learning, but once there the population is then invaded by a superior pure strategy. This occurs for precisely the reason suggested by Deacon and Godfrey-Smith, learning is initially superior to the other pure strategies. But as learning grows to take over the population its presence changes the fitness of the other strategies

[13]Lines in these simplices represent the trajectory of the discrete time replicator dynamics for one of three canonical games: a pure coordination game with a pareto dominant equilibrium, a prisoner's dilemma, and a Hawk-Dove game. In these three cases the payoffs range from 0 to 3 and the cost of learning is 0.1.

such that one of them is superior. In these cases we have a circumstance which conforms exactly to the process described by Baldwin.

5 Conclusion

We have shown that frequency dependent selection, conceived of as game theoretic interaction, often cannot sustain populations of plastic individuals and can sustain some plasticity only under specific conditions. This demonstrates an important limitation to the sort of environmental complexity that can provide the selective pressure for the evolution of complex traits such as cognition: for the evolution of such traits, we will need more than simple social interaction.

Instead, we assert that the explanation for the evolution of these traits must be found outside of strategic interaction and is instead found in external environmental variation. This variation can effect strategic situations – a game whose payoff function changes over time, for instance – but the source of the variation must be external to features of the population itself. This has significantly limited the types of environmental variation that can be appealed to in order to explain plasticity. In addition, our results demonstrate that the Simpson-Baldwin effect may be more widespread in strategic situations than previously supposed.

A Proof of proposition 6

Proof. Since G is a 2×2 game we know that there is only one possible deviation from any NE. Thus, we only need to consider pure strategies NE in 3 cases (see figure 8).

	A	B
A	a	y
B	x	b

Figure 8: Generic 2×2 Game.

1. *Strictly Dominance Solvable:* If G is strictly dominance solvable, then let s be the strictly dominant strategy. Without loss of generality, let $x = 0$ and A be the strictly dominant strategy ($a > 0$ and $y > b$). We need to show that $u(A, \mathcal{L}) > u(\mathcal{L}, \mathcal{L})$ or:

$$(1-\epsilon)a + \epsilon y > (1-\epsilon)^2 a + (\epsilon - \epsilon^2)x + \epsilon^2 b \tag{9}$$

   And, since $y > b$ and $\epsilon < 0.5$ this is necessarily satisfied.

2. *Hawk Dove:* Without loss of generality, let $a = 0$, $b \geq 0$, $x > a$ and $y > b$. We need to show that either A or B can invade $\mathcal{L}$.

   (i) If $x \geq y$, then let $s = B$:

$$u(\mathcal{L}, \mathcal{L}) = \frac{1}{2}[(1-\epsilon)^2 x + \epsilon^2 y + (1-\epsilon)^2 y + \epsilon^2 x] + (\epsilon - \epsilon^2)b \tag{10}$$

   and

$$u(B, \mathcal{L}) = (1-\epsilon)x + \epsilon b \tag{11}$$

   Hence, $u(\mathcal{L}, \mathcal{L}) < u(B, \mathcal{L})$ and $\mathcal{L}$ is not an ESS.

   (ii) If $y > x$ then $\mathcal{L}$ will resist invasion of B only if $u(\mathcal{L}, \mathcal{L}) \geq u(B, \mathcal{L})$, expanded these are:

$$u(B, \mathcal{L}) = (1-\epsilon)^2 x + (\epsilon - \epsilon^2)x + (\epsilon - \epsilon^2)b + \epsilon^2 b \tag{12}$$

and

$$u(\mathcal{L},\mathcal{L}) = (1-\epsilon)^2\frac{(y+x)}{2} + (\epsilon-\epsilon^2)b + (\epsilon-\epsilon^2)0 + \epsilon^2\frac{(y+x)}{2} \tag{13}$$

which gives us that $\mathcal{L}$ will be an ESS only if:

$$b \leq \frac{y-x}{2\epsilon^2} - \frac{y}{\epsilon} + x + y \tag{14}$$

Considering strategy A, we can state that will resist invasion of A only if $u(\mathcal{L},\mathcal{L}) \geq u(A,\mathcal{L})$, expanded $u(A,\mathcal{L})$ is:

$$u(A,\mathcal{L}) = (1-\epsilon)^2 y + (\epsilon-\epsilon^2)y + (\epsilon-\epsilon^2)0 + \epsilon^2 0 \tag{15}$$

When this is taken in conjunction with $u(\mathcal{L},\mathcal{L})$ gives us that $\mathcal{L}$ will be an ESS only if:

$$b \geq \frac{\frac{y-x}{2\epsilon} + x - \epsilon(y+x)}{1-\epsilon} \tag{16}$$

These two inequalities cannot be satisfied when $y > x > 0$ and $\epsilon < 0.5$.

3. *Coordination:* Without loss of generality, let $y = 0$, $a > x > 0$, and $b \geq a$, (coordination games other than those discussed above). We just need to show that $u(B,\mathcal{L}) > u(\mathcal{L},\mathcal{L})$. Expanding these gives us:

$$u(B,\mathcal{L}) = (1-\epsilon)^2 b + (\epsilon-\epsilon^2)b + (\epsilon-\epsilon^2)x + \epsilon^2 x \tag{17}$$

and if $(\mathcal{L},\mathcal{L})$ "attempts" to play the superior pure-strategy Nash equilibrium (B,B) (their highest possible payoff) then we have:

$$u(\mathcal{L},\mathcal{L}) = (1-\epsilon)^2 b + (\epsilon-\epsilon^2)x + (\epsilon-\epsilon^2)0 + \epsilon^2 a \tag{18}$$

Thus, we need $(\epsilon-\epsilon^2)b + \epsilon^2 x > \epsilon^2 a$ which is satisfied due to $\epsilon < 0.5$ and $b \geq a$ and hence, $\mathcal{L}$ is not an ESS.

References

Ancel, L. W. (1999). A quantitative model of the Simpson-Baldwin effect. *Journal of Theoretical Biology 196*, 197–209.

Baldwin, J. (1896). A new factor in evolution. *The American Naturalist 30*, 441–445.

Deacon, T. W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. Norton.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge University Press.

Godfrey-Smith, P. (2002). Environmental complexity, signal detection, and the evolution of cognition. In G. M. Bekoff and C. Allen (Eds.), *The Cognitive Animal*, Chapter 18, pp. 135–141.

Godfrey-Smith, P. (2003). Between baldwin skepticisim and baldwin boosterism. In B. H. Weber and D. J. Depew (Eds.), *Evolution and Learning: The Baldwin Effect Reconsidered*. MIT Press.

Harley, C. B. (1981). Learning the evolutionarility stable strategy. *Journal of Theoretical Biology 89*, 611–633.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Maynard Smith, J. and G. Price (1973). The logic of animal conflict. *Nature 146*, 15–18.

Roth, A. E. and I. Erev (1995). Learning in extensive-form games: Experimental data and simple dynamics models in the intermediate term. *Games and Economic Behavior 8*, 164–212.

Simpson, G. G. (1953, June). The baldwin effect. *Evolution 7*(2), 110–117.

Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Blackwell.

Weibull, J. (1995). *Evolutionary Game Theory*. Cambridge: MIT Press.