Genetic Algorithm Based Models for Pedagogical Examples in Epigenetics

Kellin Rumsey

Department of Mathematics and Statistics

University of New Mexico

Complex Adaptive Systems - Final Project

Abstract—As the movement for an extended evolutionary synthesis gains traction, phenomena such as plasticity and niche construction are becoming increasingly important. Some argue that a gene-centric view ignores key features of evolution, and fails to adequately explain these aforementioned processes. In this paper, we examine some pedagogical examples of plasticity and niche construction, and attempt to capture the dynamics using simple genetic based algorithms. In a drastically simplified version of reality, we find that these epigenetic processes can be roughly explained by the simple genetic approach.

I. Introduction

After spending more than two decades researching and collecting data, Charles Darwin published his book "On the Origin of Species"[1] in 1859 and the idea of *natural selection* was born. For many years after Darwin, naturalists still commonly believed that other frameworks such as Lamarckism better explained the observed complexity in nature. In 1900, Gregor Mendel's work on genetics would be rediscovered, but was seen at first as a rival to natural selection. In the years that followed, the work of R.A. Fisher, J.B.S Haldane and Sewell Wright would produce the mathematical framework required to bring these two theories together to form the *modern synthesis*.[2]

Although modern synthesis has been widely accepted, it has not been immune to criticism over the years. Many have argued that epigenetic's, roughly defined as the study of phenotypic variability not explained by genetics, is not adequately resolved by modern synthesis. Recently, the work of Massimo Pigliucci and others has sparked a revival of the search for an *extended synthesis*.

In Pigliucci's 2007 paper entitled "Do we need an extended evolutionary synthesis?"[3], he gives several examples of processes which he claims are not explained by modern synthesis. In this paper, we will primarily focus on *phenotypic plasticity*, the ability of a genotype to express different phenotypes subject to different environments. We also give an example which is loosely related to *niche construction* where organisms bias their selection by changing their environment.

Background

Genetic algorithms have become an increasingly useful tool for complex optimization problems, often when the target function is a black box[4]. Any genetic algorithm requires a *fitness function* $f(\cdot)$ to be optimized, and a population of solutions. The *genotype* g of an individual in the population is

Fig. 1. *Precis octavia* - The Gaudy Commodore butterfly emerges from it's cocoon with different colors in the dry (left) and wet (right) seasons.

an encoded binary string which represents a candidate solution to the fitness function. In our examples, we will be explicit, but the *phenotype* of an individual can be roughly related to the performance or fitness of the individuals genotype. The population of individual solutions evolves over time via *crossover* and *mutation*. In this paper, we examine some pedagogical examples in epigenetics, using genetic algorithms acting on cellular automata as a simple analogy.

A *cellular automaton* (CA) in two dimensions consists of a discrete grid such that each cell $c_{ij}(t) \in A$ at time t where A is some finite alphabet. The *Moore Neighborhood* with radius 1 of a cell c_{ij} is the set

$$N_{ij} = \{c_{i'j'} \ s.t. \ |i - i'| \le 1 \ \wedge \ |j - j'| \le 1\}$$

The grid changes states over time according to some fixed *rule* which maps $c_{ij}(t) \rightarrow c_{ij}(t + 1) \in A$ based on the cells neighborhood. Taking $A = \{0, 1\}$, there are 2^9 possible configurations of N_{ij} and a rule can be encoded as a binary string of length 2^9. We apply all rules synchronously, and we never apply the rule to cells on the boundary of the grid.

II. Plasticity

Plasticity has been defined many times by many authors. One simple definition given by Agrawal[5] is "the ability of an organism to express different phenotypes depending on the environment". In this section, we focus on modeling two well-known examples and examine the ability of a standard genetic algorithm to evolve solutions.

A. Seasonal Plasticity - Precis octavia

The **gaudy monarch** butterfly exhibits plasticity by emerging with vastly different patterns in the wet and dry season[6].

This is illustrated in Figure 1, where the environment-phenotype relationship is apparent, despite the similar genotype.

We model this by introducing a *seasonal fitness function*. Formally, we consider an environment with $S + 1$ seasons. Further we suppose that season $s \in \{0, 1, \cdots S\}$ has a season dependent fitness function $f_s(\cdot)$. Now the marginal fitness at generation m can be given by

$$f(g, m) = f_{m \% S}(g)$$

Where $\%$ represents the modulo function. For our purposes, we focus on the wet (s=0) and dry (s=1) season case by setting $S = 1$.

In our simple analogy, a commodore larvae is given an initial season-dependent environment represented by some 11×11 CA, denoted $C_s(0)$. The larvae then has $T = 5$ time steps to transform itself into a butterfly. An optimal pattern is defined for each season s and is denoted by $\hat{C}_s$. Fitness is measured by the number of bits which match the current seasons optimal pattern. Figure 2 shows the initial and optimal patterns for the dry and wet seasons.

As a final side note, we can guarantee that at least one optimal solution exists (for each seasonal fitness function), since the optimal pattern seen in Figure 2d is the result of `Algorithm 2` from Peled et. al. [7]. By a symmetry argument, the optimal pattern in Figure 2b must exist as well.

We utilized a relatively simple GA, which behaved as follows.

- A diverse population of 200 individuals is generated randomly. We employ a hierarchical generation technique so that that p, the proportion of 1's in an initial candidate follows a Uniform(0,1) distribution.
- Two (distinct) parents are selected probabilistically, with "average fitness rank" serving as the sampling weights.
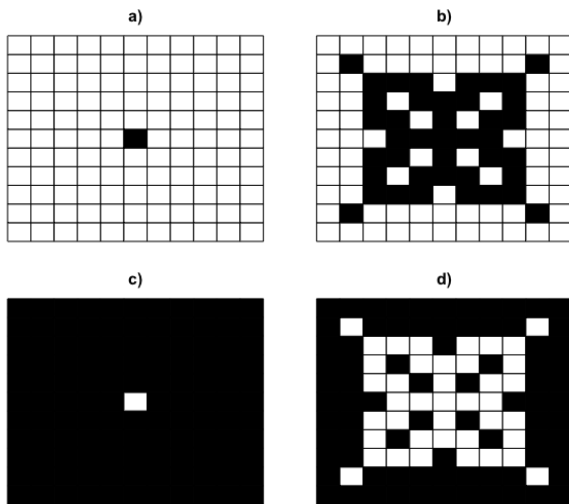

Fig. 2. a) Initial dry season environment $C_0(0)$. b) Optimal dry season pattern $\hat{C}_0$. c) Initial wet season environment. d) Optimal wet season pattern.
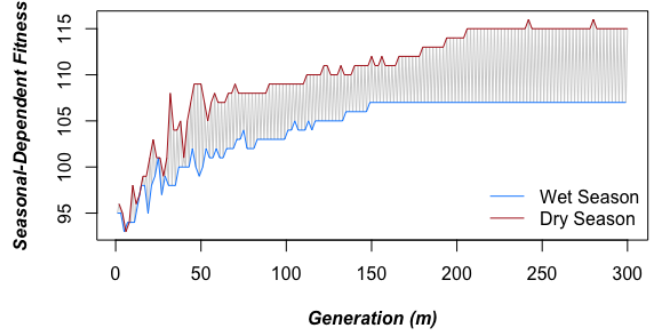

Fig. 3. Best fitness as a function of generation for the first 300 generations for the season-dependent fitness function.

- A single split crossover is used on a reversible permutation of the parent genomes to generate two children.
- Mutation occurs randomly in 5% of children. The number of mutations which occur follows a Zipfian (power-law) distribution so that occasionally a large number of mutations can occur.

If the desire is to achieve an optimal fitness score of $11^2 = 121$, this rather arbitrary problem is quite difficult. The fitness landscape is incredibly irregular, with a multitude of local extrema. Even after varying the parameters of the genetic algorithm, optimality is never reached. Although the phenotypic distance between $C_s(T)$ and $\hat{C}_s$ is fairly small the genotypic distance can be enormous, and there lies the problem[8]. Put another way, the population readily converges to one of the many near-optimal "paths", but this path may be drastically different from the optimal path.

Figure 3 illustrates the best fitness as a function of generation for the first 300 generations. These are the results of a single run using the parameters described above. In each run, there is usually a seasonality effect, where the individuals consistently camouflage better in one season (although occasionally this swaps halfway through). Still, the genotype is able to produce a "near"-optimal solution in each of the two environments.

Figure 4 illustrates the best fitness genotype from generation $m = 300$ operating on the wet-season and dry-season environment. Given different environmental conditions, the genotype is able to produce two separate near optimal phenotypes. In the context of our blatantly over-simplified model, we might interpret this as evidence that standard evolutionary theory can account for this between-generation plasticity.

B. Stochastic Plasticity - Taeniopoda eques

The **western horse lubber grasshopper** displays a different type of plasticity. As shown in Figure 5, the body size of these insects physically changes based on nutrition, which can be directly related to precipitation. Although there may be cyclical effects present, this is an example of plasticity
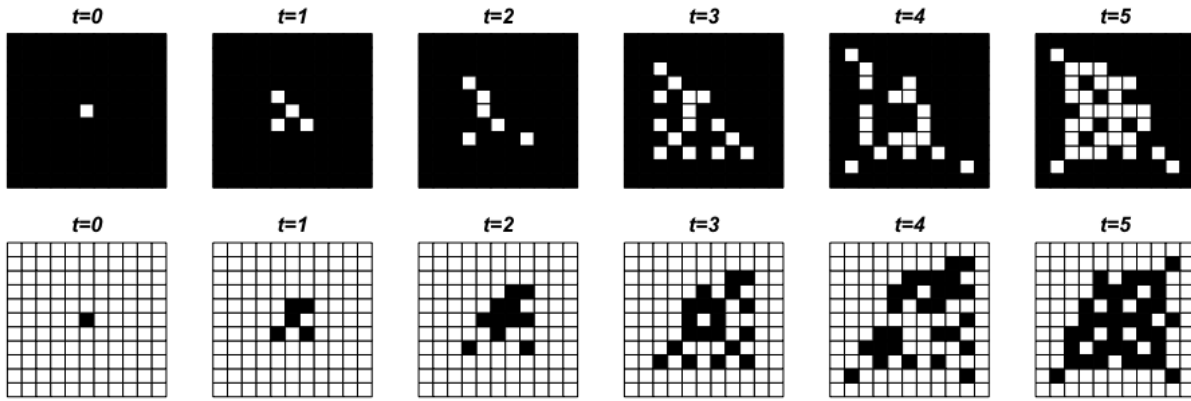
Fig. 4. Seasonal plasticity. The best-fitness monarch butterfly genotype at generation 300 is tested on the wet (top) and dry (bottom) environments.

which is less predictable than the season-dependent version in the previous section. We must define a fitness function which is dependent on a stochastic environmental factor. Formally, we consider a sequence of random variables $\{X_m\}_{m=1}^{\infty}$, with some probability distribution $\pi_m()$, and we give X_m as an argument to the environment-dependent fitness function at generation m.

$$f(g, m) = f_{X_m}(g)$$

This definition is quite general, but we choose to simplify the discussion by taking the sequence $X_m \overset{iid}{\sim} Bernoulli(1/2)$[1]. In the context of our analogy, this is equivalent to assuming that we have dry years and wet years randomly, with equal probability. Again for simplicity, we consider the same problem as we did for the monarch butterflies.

The results are somewhat similar as in the previous case, although we can make two informal observations.

- On average, the genetic algorithm has a more difficult time evolving effective strategies for *both* environments. For instance, if there is a draught leading to a long stretch of dry years, then many of the well-adapted wet year genotypes may die off.
- At any time, the population is in a given state indicating that individuals are better adapted for a particular environment. The population changes states more frequently under the stochastic fitness function than the seasonal one.

Figure 6 shows the best fitness over time for the first 300 generations. The results are qualitatively similar to what was observed for the seasonal case.

The final patterns exhibited by these individuals is shown in Figure 7. Although we do see that the population is more effective in wet years than in dry years, the emergence of two distinct patterns is indeed present. Again, a well-adapted genotype produces a vastly different phenotype given a different environment.

[1] A Bernoulli random variable takes the value 1 with probability p and takes the value 0 otherwise

III. NICHE CONSTRUCTION

Niche construction refers to the process in which an organism modifies it's environment, thereby biasing it's selection. Rather than having an environment which solely dictates fitness, the individual systematically co-directs it's evolution. Nest or habitat building animals such as birds and beavers provide a classic example of niche construction.

Fire resistant pine trees litter the ground with needles and cones, increasing the frequency of forest fires thereby reducing competition for precious resources. Certain types of yeast capable of fermenting fruit influence their own environment, as well as the environment of the fruit flies which they rely on for transportation. These two examples are of particular interest in this section, since they highlight the relationship between niche construction and *coevolution* of individuals.

The following definition is convenient for our purposes, and accounts for this relationship. "Niche construction is the process whereby organisms, through their activities and choices, modify their own and each other's niches"[9]. To formalize this concept, we consider a population $\mathcal{P}$ of individuals distinguished by their genotype g. Now fitness of a genotype depends on the environment, the population at the current generation and the interplay between the two.

$$f(g) = f(g, \mathcal{P})$$

Fig. 5. Best fitness as a function of generation for the first 300 generations for the season-dependent fitness function.
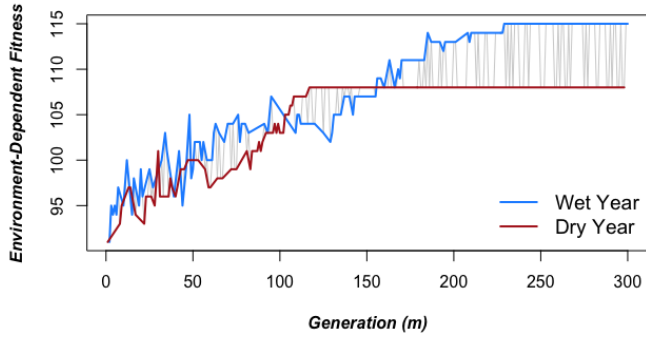
Fig. 6. Best fitness as a function of generation for the first 300 generations for the season-dependent fitness function.
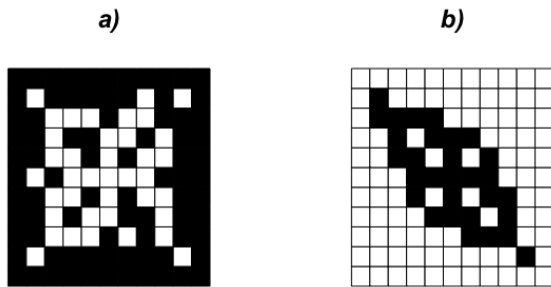

Fig. 7. Stochastic Fitness. a) Evolved pattern for wet environment. b) Evolved pattern for dry environment.

A. Niche construction - Uta stansburiana

The **common side-blotched lizard** is a fascinating evolutionary enigma, with individuals playing a generational and eternal game of rock-paper-scissors[10]. These lizards, pictured in Figure 8, come in three different flavors or *morphs*.

- **Orange-throated Lizards** are ultra-dominant. These lizards maintain large territories with many females. *An Orange-throated lizard can fight off a Blue-throated lizard to win a mate.*
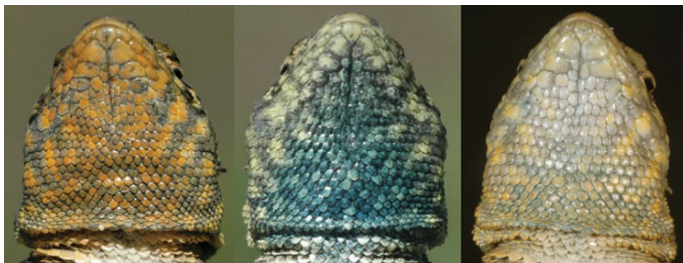

Fig. 8. The three morphs of the common side-blotched lizard: Orange (left), blue (middle) and yellow (right).

- **Blue-throated Lizards** are semi-dominant. These lizards maintain moderate sized territories, with a smaller number of females. *A Blue-throated lizard can fight off a Yellow-throated lizard to win a mate.*
- **Yellow-throated Lizards** are non-dominant. Although these lizards technically don't have a territory, their coloration is similar to a female. *A Yellow-throated lizard can steal a mate from an Orange-throated lizard.*

We attempt to model this unique form of niche construction by using an *agent based model* combined with a genetic algorithm component. We lay out the details of the toy model here.

- Sixty female lizards are randomly placed into the environment ($[0, 20]^2$) onto the lattice points of a latin hypercube. For simplicity, the females never move.
- An initial population of 36 male lizards are randomly placed into the environment, and a morph (throat color) is assigned at random. Each lizard has a territory radius which depends on its morph. Specifically, $R_y = 2$, $R_b = 2.5$ and $R_o = 3$.
- **Fitness.** Based on the position and morphs of the population, we define fitness of a lizard to be the likelihood of that lizard mating. This is influenced by three things.
  - **Range:** More females in a territory *increases* the chances of mating.
  - **Overcrowding:** Fighting over a mate with a lizard of the same type *reduces* chance of mating.
  - **Competition:** Fighting over a mate with a lizard of the rival type *eliminates* the chance of mating.
- **Evolution.** At each generation, a third of the population is chosen survive and reproduce. Each surviving lizard produces two offspring which migrate to a new location according to a Levy walk.[2] The offspring have a small chance of changing morph color.

As is true with many agent based models, our model is heavily dependent on the parameters controlling fitness and evolution. Even so, there is a wide range of parameters which mimic the coevolution of these lizards. Figure 9 is a diagrammatic representation of the lizards and their environment for three different generations. The transition between dominant morphs is not immediate. In generation 40, the yellow-throated lizards have "beat" most of the orange-throated lizards and are the dominant morph. This gives the blue-throated lizards a chance to re-emerge, and they become dominant in generation 48. The emergence of the blue-throated lizards means that yellow-throated lizards have become sparse. This gives the orange-throated lizards a chance to become dominant in generation 56.

We explore these ideas further with Figure 10, which plots the proportion of total population as a time series for each morph. The top panel of Figure 10 is cluttered, and it is hard to be certain of any trends. The bottom panel is a close up view of the bottom-right portion of the top panel, and after

---

[2]The offspring "walk" to a new location via a bivariate t-distribution with 5 degrees of freedom and a scale parameter of 2.
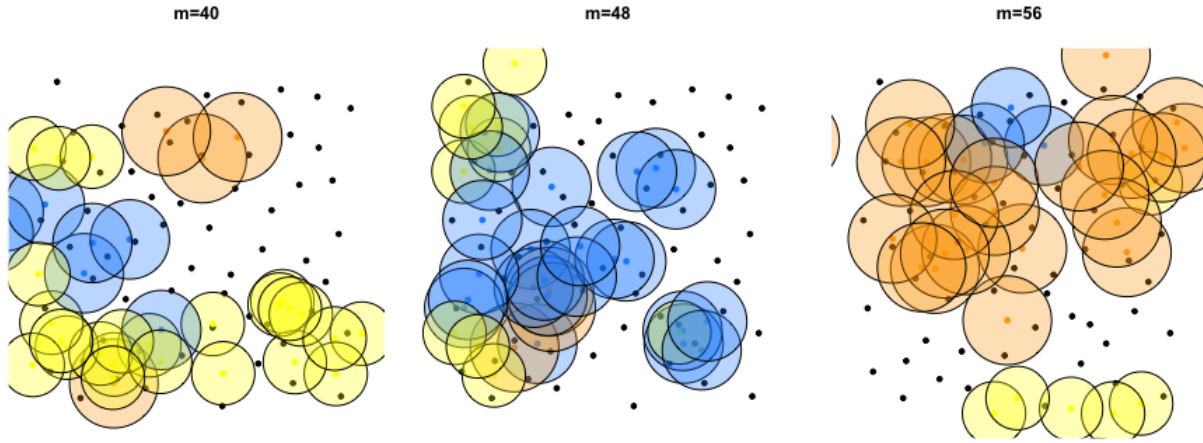
Fig. 9. Simulated lizard populations at generations 40 (yellow majority), 48 (blue majority) and 56 (orange majority). Black dots represent females.

some shading, the pattern "yellow"-"blue"-"orange" becomes obvious. Let's break this down.

1) When the yellow-throated lizards are scant, the orange throated lizards can dominate. Therefore the blue throated lizards will start to rapidly die off.
2) When the blue-throated lizards are scant, the yellow throated lizards can steal mates more easily. Therefore the orange-throated lizards will start to die off.
3) When the orange-throated lizards are scant, the blue-throated lizards can handle their territories again. Therefore the yellow-throated lizards will start to die off.
4) Back to 1.

The model appears to be working as expected, but we now attempt to validate and quantify the behavior of the model via some formal metrics. For brevity, we enumerate the morphs yellow, blue and orange from 1 to 3 respectively.
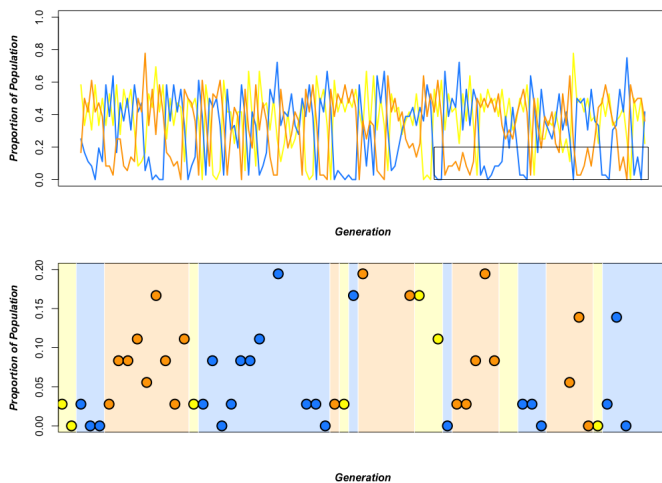

Fig. 10. Lizard population, by morph, across many generations. The bottom panel represents a "close-up" of the black rectangle in the top panel. Shade color represents the current dominant morph.

Dominant Morph Transitions: We say that the dominant morph *transitions* from i to j ($i \rightarrow j$) if morph i is in the majority at time m and morph j is in the majority at time $m + 1$. Let $N_{i \rightarrow j}$ be the number of transitions from i to j.

Under the rules of "LIzard rock-paper-scissors", it is clear that we expect $1 \rightarrow 2$, $2 \rightarrow 3$ and $3 \rightarrow 1$. Therefore the quantities below indicate the ratio of times the population makes the "expected" transition.

$$\frac{N_{1 \rightarrow 2}}{N_{1 \rightarrow 3}} \qquad \frac{N_{2 \rightarrow 3}}{N_{2 \rightarrow 1}} \qquad \frac{N_{3 \rightarrow 1}}{N_{3 \rightarrow 2}}$$

In a formal testing framework, ratios larger than 1 give evidence for the desired/expected model. In simulation, we produce 2000 realizations of the model, each for 1000 generations. Figure 11 shows histograms of the transition ratios given above for these simulations. In all three cases, the distribution of transition ratios is almost always above 1 as desired.

Transfer Entropy: The proportion of morph i in the population over time (generation) defines a stochastic process, denoted X_i. *Transfer entropy* is a directional measure of information transfer from one stochastic process to another[11] (we omit the details for the sake of succinctness). Let $T_{i \rightarrow j}$ be the transfer entropy from X_i to X_j.

Based on Figure 10, we might assume that information flows more directly from $1 \rightarrow 2$ than from $1 \rightarrow 3$ (and similarly for the other morphs). Thus the following statistics give evidence for the desired model behavior if they are positive.

$$T_{1 \rightarrow 2} - T_{1 \rightarrow 3} \qquad T_{2 \rightarrow 3} - T_{2 \rightarrow 1} \qquad T_{3 \rightarrow 1} - T_{3 \rightarrow 2}$$

Figure 12 presents histograms representing the distribution of these statistics for the 2000 simulations. The statistical significance of these results can be argued, but there seems to be a systematic shift of the distributions to the right. In other words, the transfer entropy differences are positive approximately 75% of the time for each of the three cases.
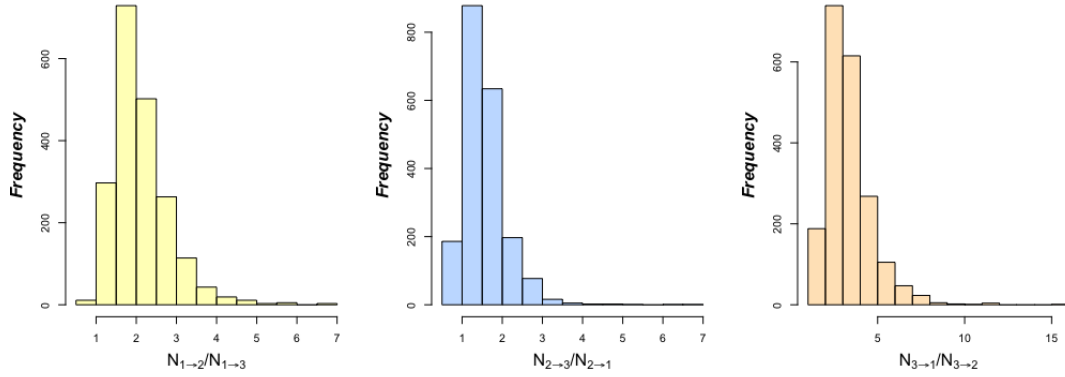
Fig. 11. Sampling distributions for the three transition ratios across $2,000$ simulations.
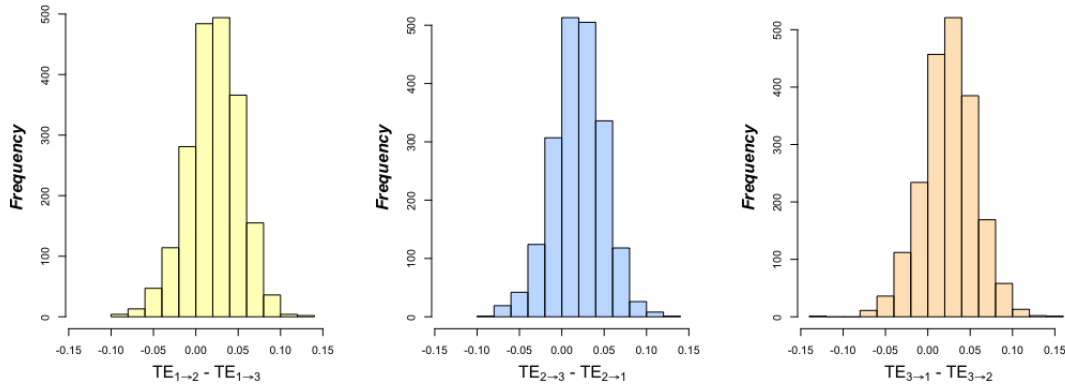

Fig. 12. Sampling distributions for the three transfer entropy differences across $2,000$ simulations.

IV. CONCLUSION

Standard evolutionary theory, largely based on modern synthesis, has been criticized as short-sighted and narrow-minded. Some claim that, for the puzzle that is evolutionary biology, the "gene-centric" view fails to adequately put the pieces together. While plasticity and niche construction are just two small parts of the puzzle, it can be argued that they are corner pieces which must be dealt with.

In the first part of this paper, we examined two pedagogical examples of plasticity in nature. By limiting ourselves to a very simple genetic algorithm, we sought to provide evidence for or against standard evolutionary theory. We showed that an evolved genotype, when exposed to different environmental conditions, is able to exhibit different phenotypes. We must note however, that our model was inordinately simple, and any desire to extrapolate these findings to the real world would be naive.

In the second part of this paper, we model the evolutionary dynamics of the common side-blotched lizard using an agent based model with an evolutionary component. A phenomena that is loosely related to niche construction. We showed that this simple model roughly captures the strange dynamics of these lizards. Here, we must note that the genetic component of the algorithm was minimal, thus it makes even less sense to make any extravagant claims about the gene-centric way of thinking.

Although these models were drastically simplified, the results beg some interesting questions. Under what circumstances can a gene-centric algorithm exhibit plasticity? Can the process of niche construction be modeled with a more classical genetic algorithm? Perhaps these results can be taken as a starting point for more interesting thought experiments in the future.

REFERENCES

[1] Darwin, Charles. On the origin of species, 1859. Routledge, 2004.
[2] Fisher, Ronald Aylmer. The genetical theory of natural selection: a complete variorum edition. Oxford University Press, 1999.
[3] Pigliucci, Massimo. "Do we need an extended evolutionary synthesis?." Evolution 61, no. 12 (2007): 2743-2749.
[4] Forrest, Stephanie Genetic algorithms: principles of natural selection applied to computation SCIENCE13 AUG 1993 : 872-878
[5] Whitman, D. W., and A. A. Agrawal. "What is phenotypic plasticity and why is it important? 2009." Phenotypic Plasticity of Insects (Whitman, DW and Ananthakrishnan, TN, eds): 1-63.

[6] Laland, Kevin, Tobias Uller, Marc Feldman, Kim Sterelny, Gerd B. Mller, Armin Moczek, Eva Jablonka et al. "Does evolutionary theory need a rethink?." Nature 514, no. 7521 (2014): 161.

[7] Peled, Y. Bar, Vikas, K. Mishra, Avishy, Y. Carmi, Computing by nowhere increasing complexity, CoRR 2017

[8] Wagner, Andreas The role of robustness in phenotypic adaptation and innovation Proc. R. Soc. B 2012 279 1249-1258; DOI: 10.1098/rspb.2011.2293. Published 24 February 2012

[9] Laland, Kevin N., John Odling?Smee, and Scott F. Gilbert. "EvoDevo and niche construction: building bridges." Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 310, no. 7 (2008): 549-566.

[10] Tinkle, Donald W. "The life and demography of the side-blotched lizard, Uta stansburiana." (1967).

[11] Schreiber, Thomas. "Measuring information transfer." Physical review letters 85, no. 2 (2000): 461.