Capstone Project Proposal

Matthew Moreno

January 26, 2017

1 Evolutionary Algorithms

At the genesis of modern computing, the 1950s, researchers began to apply advancing computational capabilities to investigate and test models of biological evolution. Very quickly they realized the potential of virtual evolution to achieve other ends, setting into motion a line of research that has since blossomed into the field of evolutionary algorithms (EA) design. These algorithms, which use mechanics inspired by biological evolution to evolve novel solutions to a wide array of problems, share a generally consistent basic methodology. The process begins with a population of randomly generated solutions. In a generation-based loop, an elite subset of the population is selected for their fitness (their quality as a solution), subjected to random changes, and recombined with each other to form the next generation. The cycle repeats for as many iterations as desired, and fitness tends to increase with each iteration. Figure 1 provides a graphical overview of this process.

When discussing evolutionary algorithms (as well as their biological counterpart) an important distinction is drawn between phenotype and genotype. Phenotype refers to the characteristics of an individual that interact with its environment to determine its fitness. In biology, the physical form of an organism (i.e. its body) is the phenotype. In evolutionary algorithms, the phenotype refers to the characteristics of an individual that are evaluated during selection. Genotype refers to information that is used to determine the phenotype that is passed from generation to generation. In biology, a DNA sequence serves as the genotype. Although many different genotypic encodings are employed in evolutionary algorithms, in evolutionary algorithms the genotype ultimately boils down to a collection of digital information. A glossary reviewing this vocabulary, as well as a handful of other terms, is provided in Section 7.

Researchers and engineers have widely demonstrated the ability of EA to attack labor-intensive optimization problems and to discover novel solutions beyond the reach of human ingenuity [Poli et al., 2008]. The intervening half century of EA research has seen diversification of the general evolutionary search process described above and diversification of the contents and format of candidate solutions. Today, evolutionary algorithms serve a dual purpose: a tool for biological inquiry and an algorithm for the automatic design of solutions to problems.

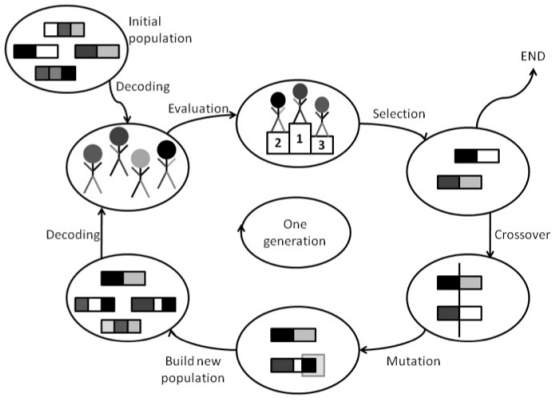

Figure 1: Evolutionary algorithms traditionally operate in a generation-based loop that, over the course of many iterations, gradually refines a population of candidate solutions, which is initialized with randomly-generated individuals, to generate increasingly fit individuals. That is, individuals that provide an increasingly satisfactory solution to a particular problem. At the start of each cycle of the loop, individual solutions are generated from a population of genotypes. These solutions are scored by a fitness function, which measures the performance of the individual as a solution to the problem. Then, the genetic material of fit individuals are mutated and recombined to create the next generation of candidate individuals. Once a predefined stopping criterion is met, usually a maximum number of generations or threshold fitness score, the evolutionary cycle is halted [Prothmann et al., 2009]

.

2 Evolvability

Evolvability is a principal concern to Evolutionary Algorithm researchers and evolutionary biologists alike. Although many competing definitions of evolvability exist in the literature, they primarily fall into one vein: the ability of a population to generate useful variation. Breaking the concept down further, evolvability stems from:

1. potential for generation of heritable phenotypic variation through mutation, and
2. bias towards the generation of potentially useful phenotypic variation through mutation.[1]

Figure 2, which juxtaposes wild-type *Arabidopsis thaliana* with several mutant strains, provides a biologically-motivated illustration of these concepts. The depicted individuals are clustered together in the genotypic space of *Arabidopsis thaliana*. Although the genotypic distance between these phenotypes is small, a relatively broad range of phenotypic variability is represented. The rich diversity of form is hence accessible to evolutionary search allows rapid adaptation to environmental changes or novel adaptation to the existing environment. This comparison illustrates, in particular, the modularity of the system under genotype-phenotype mapping; the mutant strains exhibit significant variability in overall composition of the individual while preserving the general character of existing components. Although not depicted above, the local genotypic neighborhood also maps to a great number of phenotypes indistinguishable from wild-type individuals, many specimens that differ by more calibration-like adjustments to various phenotypic properties, as well as

[1] This might also be conceptualized as a bias against lethal or otherwise deleterious mutation.

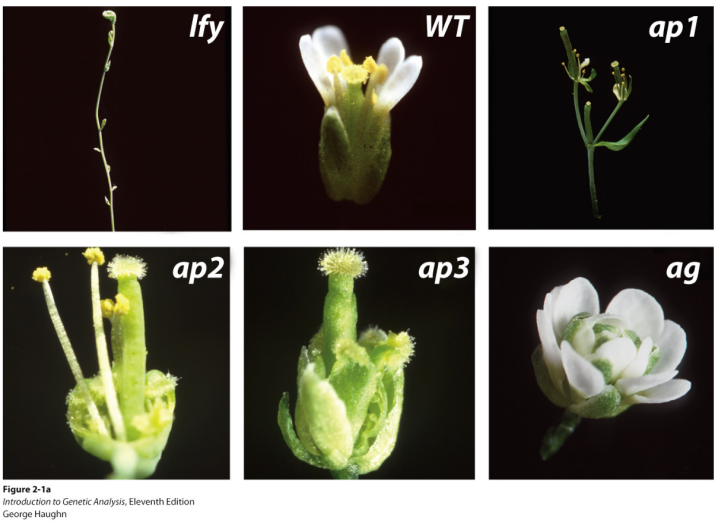

Figure 2: A wild-type specimen is portrayed in comparison with representatives from mutant strains of *Arabidopsis thaliana* [Griffiths et al., 2015].

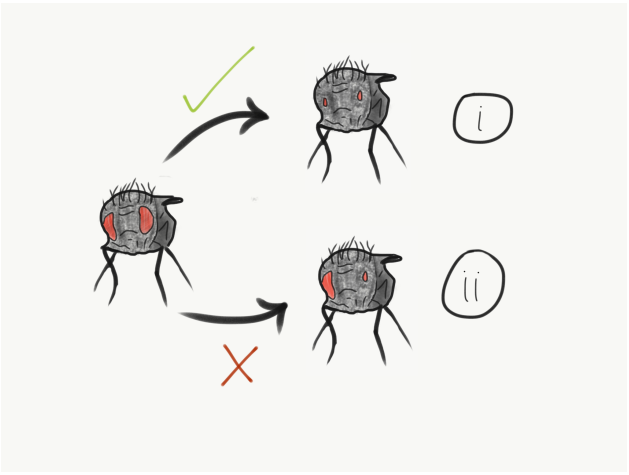

Figure 3: This cartoon depicts an artificial selection experiment on *Drosophilia melangoster*. In treatment i, selection is made for a bilaterally symmetric trait – overall decreased eye size. Because this bilaterally symmetric trait can be produced through heritable variation, artificial selection succeeds. In treatment ii, artificial selection for a non-bilaterally symmetric trait – relatively smaller right eye size – fails because (due to the development process of *Drosophilia*), heritable variation that produces this trait is not readily attainable [Tuinstra et al., 1990]

– doubtlessly – a not-insignificant (but not overwhelming) smattering of fundamentally defective phenotypes. These six specimens are, of course, insufficient evidence to substantiate a rigorous argument to this end, but the apparent preservation of modularity and the supposed moderate incidence of lethality can be cast as canalization, bias towards the generation of potentially useful variation. Figure 3 describes a more concrete biological manifestation of canalization.

Evolvability is a question of interest not just to EA researchers, but also to evolutionary biologists as they push past the modern synthesis and grapple with the extended evolutionary synthesis. The question of evolvability poses a theoretical quandary. The traditional model of static fitness-based selection (i.e. the modern synthesis) posits that selection is performed based on compatibility between an individual's phenotype and the environment but not traits of individual related to the potential to innovate and adapt in evolutionary time.[2] At face value, this modern synthesis fails to explain the emergence of those traits that

[2]Examples of such traits include modularity, canalization, degeneracy.

has been observed in biological evolution – and have been sorely lacking in attempts at digital evolution, how natural selection might "favor properties that may prove useful to a given lineage in the future, but have no present adaptive function" [Pigliucci, 2008]. To resolve this dilemma researchers are considering a number of hypotheses, including but not limited to

- evolutionary selection mechanisms, positing that mechanisms beyond traditional static fitness-based selection (such as divergent selection or a fluctuating selection criteria) might promote evolvability [Mengistu et al., 2016, Kashtan et al., 2007];
- developmental mechanisms, positing that perhaps that indirect encoding of the phenotype adds inherent bias towards regular, modular phenotypes [Clune et al., 2011] or allows for the encoding of biases that canalize mutational effects towards selectively-advantageous ends [Reisinger and Miikkulainen, 2007]; and
- phenotypic plasticity, positing that environmental influence on the phenotype – by altering the trajectory of the developmental process or otherwise inducing phenotypic changes in response to environmental stimulus [Fusco and Minelli, 2010] – might promote evolvability [Moczek et al., 2011].

It seems likely that evolvability stems from a large and diffuse set of contributing factors. The establishment – or rejection – of empirical causal links between theoretical complications of the modern evolutionary synthesis and evolvability is a key research goal in the field; this type of inquiry will help determine how complicated of a model is necessary to account for evolution as observed in biological history and how complicated of a model is necessary to realize digital evolving systems with performance akin to their biological counterpart. Although scientists traditionally strive for simplicity biological systems might also be viewed from the perspective of an engineer, in which complexification in pursuit of performance is more familiar and comfortable [Sterling and Laughlin, 2015, pg 6,7]. Naive efforts to model biological systems at an arbitrary level of detail, however, are also unlikely to be tractable or, ultimately, useful (although perhaps of interest from a certain scientific perspective) [Downing, 2015, pg 354]. Thus, the establishment – or rejection – of empirical causal links between theoretical complications of the modern evolutionary synthesis and evolvability is a key research goal in the fields of evolutionary biology and evolutionary algorithm design. It is hoped that this line of inquiry, into which this project falls, will help determine which elaborations are necessary to account for evolution as observed in biological history and to realize digital evolution with performance akin to its biological counterpart.

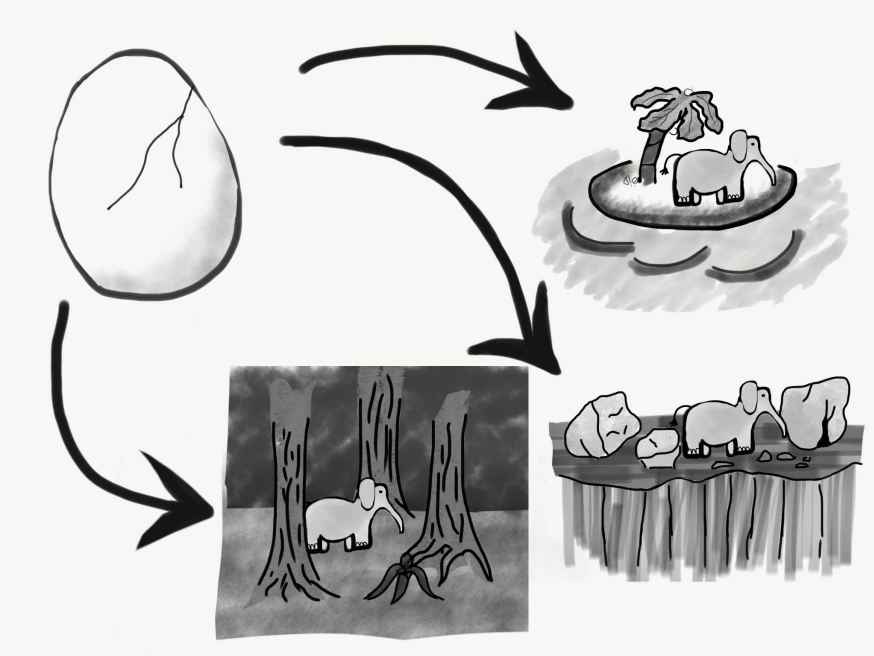

Figure 4: An illustration of direct plasticity, where phenotypic form is stable under environmental perturbation.

3 Phenotpyic Plasticity

3.1 Definition

Plasticity refers to environmental influence on the phenotype. In biology, environmental and genetic influences, together, determine the phenotype. Environmental influences may alter the trajectory of the developmental process or may otherwise induce phenotype changes in response to environmental stimulus [Fusco and Minelli, 2010]. A conceptual distinction, which will be central to this investigation, can be drawn between direct and indirect plasticity. In the first, environmental influence is exerted directly on developmental or physiological processes. In the second, environmental signals prompt responses that are mediated by physiological or developmental systems; that is, cues from the environment are processed more like informational signals than coercive physical influence [Fusco and Minelli, 2010]. Although this distinction might appear nebulous at first blush – how exactly does a signal differ from coercive influence? – it has important implications to the design of the proposed experimental regime. At a fundamental level, successful direct plasticity entails *resistance* to environmental influence on the phenotype while successful indirect plasticity entails strategic *amplification* of environmental influence on the phenotype. Figures 4 and 5 provide a cartoon illustration of this distinction. In Figure 4 the cartoon elephant direct phenotypic plasticity, developing high-fitness phenotypic forms (which, in this example, appear nearly indistinguishable to a casual observer but in general need not be identical) despite variable environmental influence (i.e. diet, temperature, humidity, etc.). In Figure 5 the cartoon plant exhibits indirect plasticity, developing alternate phenotypic forms in response to variable environmental signals (i.e. light and shadow).

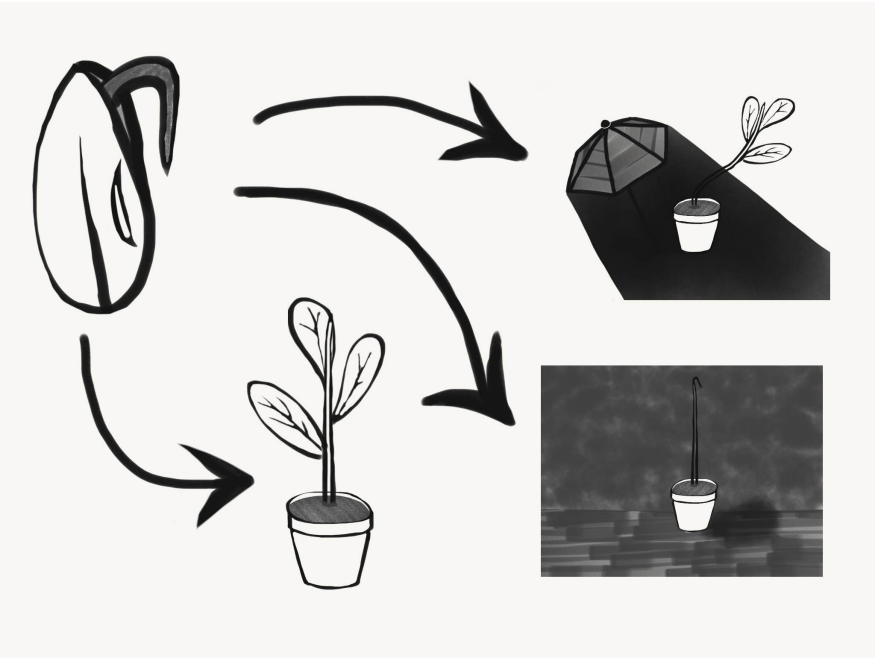

Figure 5: An illustration of indirect plasticity, where alternate phenotypes are expressed based on environmental signals.

3.2 Symbolic Exposition

The goal of this section is to bring in some mathematical notation – albeit, in a fast-and-loose manner – to provide a more concrete definition of direct and indirect phenotypic plasticity as well as flesh out the immediate implications of direct and indirect phenotypic plasticity on evolution. We will use notation that assumes that a genotype $\vec{g} \in G$, a phenotype $\vec{p} \in P$, and an environment $\vec{e} \in E$ are well-modeled as vector quantities. While this model is not particularly biologically plausible, it comes somewhat closer to meaningful resemblance in the case of artificial evolution and – in any case – provides a nice intuition for the argument being made about phenotypic plasticity and evolvability. A concrete example is often useful when thinking through abstract symbolic argumentation. One might picture a microbe swimming in a chemical soup. In this case, $\vec{g} \in G$ would represent the genetic material of the microbe – its DNA. We might imagine $\vec{e} \in E$ as information describing the chemical composition of the microbe's environment. The microbe's phenotype, $\vec{p} \in P$, could be imagined as the set of protein products expressed by the matured microbe.

Begin by considering the relation between a genotype $\vec{g} = \langle g_1, g_2, \ldots, g_n \rangle \in G$, a developmental process f, and a phenotype $\vec{p} = \langle p_1, p_2, \ldots p_n \rangle \in P$,

$$\vec{p} = f(\vec{g})$$

where $g_1, g_2, \ldots, g_n$ each represent a distinct element of the genotype and $p_1, p_2, \ldots p_m$ each represent a distinct element of the phenotype. For a direct-encoded scheme, $m = n$ and the developmental process f is a bijection; each piece of information in the genotype has a one-to-one correspondence to a piece of information (i.e. a characteristic) of the phenotype. (In most direct-encoded schemes, f would typically be the identity

transformation). In our microbe example, this would essentially correspond to a search operating directly on expressed protein concentrations.

For an indirect-encoding scheme, f would instead be a many-to-one function; some phenotype $\vec{p}$ can be achieved through several different genotypic configurations $\vec{g}_1, \vec{g}_2, \ldots, \vec{g}_r$ (i.e. $\vec{p} = f(\vec{g}_1) = f(\vec{g}_2) = \ldots = f(\vec{g}_r)$). Further, there would no longer be a one-to-one causal relationship between genotypic elements $g_1, g_2, \ldots, g_n$ and phenotypic $p_1, p_2, \ldots p_m$; we would no longer necessarily have $n = m$ and perturbing a single genotypic element g_i might result in the perturbation of a set of phenotypic elements $\{p_{q_1}, p_{q_2}, \ldots, p_{q_t}\}$.

Adding environmental influence into our model with the term $\vec{e} = \langle e_1, e_2, \ldots, e_k \rangle \in E$, we obtain the relationship

$$\vec{p} = f(\vec{g}, \vec{e})$$

where $e_1, e_2, \ldots, e_z$ represent a set of environmental characteristics.

In the case of direct plasticity, evolutionary search is pushed towards favoring $\dot{\vec{g}} \in G$ such that $\vec{p}$ is preserved under perturbations to certain characteristics of the environment. Let $\vec{\boldsymbol{n}} \in E$ be a random vector with distribution h_E representing environmental variability,

$$\vec{p} \approx f(\dot{\vec{g}}, \vec{e}) \approx f(\dot{\vec{g}}, \vec{e} + \vec{\boldsymbol{n}}).$$

Including direct plasticity in our model exerts an evolutionary pressure to make the development of the phenotype resistant to certain perturbations characteristic of the distribution of environmental noise h_E. It may be the case that for a certain $\dot{\vec{g}}, \tilde{\vec{g}} \in G$ when $\vec{\boldsymbol{n}} = \vec{0}$ we have $f(\dot{\vec{g}}, \vec{e} + \vec{0}) \approx f(\tilde{\vec{g}}, \vec{e} + \vec{0})$ but that when $\vec{\boldsymbol{n}} = \vec{n} \neq \vec{0}$ we have $\vec{p} \approx f(\dot{\vec{g}}, \vec{e} + \vec{n}) \neq f(\tilde{\vec{g}}, \vec{e} + \vec{n})$. Returning to our microbe example, direct plasticity might correspond to the phenotypic effects of fluctuating ambient temperature during development. Evolution in this case will select for genetic configurations $\dot{\vec{g}}$ that stabilize the development process f across a environmentally-realized distribution of ambient temperatures h_E.

In the case of indirect plasticity, evolutionary search is pushed towards favoring $\hat{\vec{g}} in G$ such that alternate phenotypic forms $\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_s \in P$ are conditionally expressed with the addition of environmental signals $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_s$. Concretely speaking, environmental cue $\vec{e}_i \in E$ signals that selection will be made for individuals expressing $\vec{p}_i \in P$; the environmental cue provides information about how fitness will be determined (which set of phenotypic characteristics $\vec{p}_i \in P$ will result in high fitness). Thus we have selection for $\hat{\vec{g}} in G$

such that,

$$\begin{aligned}
\vec{p}_1 &\approx f(\hat{\vec{g}}, \vec{e} + \vec{e}_1) \\
\vec{p}_2 &\approx f(\hat{\vec{g}}, \vec{e} + \vec{e}_2) \\
&\vdots \\
\vec{p}_s &\approx f(\hat{\vec{g}}, \vec{e} + \vec{e}_s).
\end{aligned}$$

Including indirect plasticity in our model exerts an evolutionary pressure to tune $\hat{\vec{g}}$ such that the developmental process is highly responsive to certain environmental influences (i.e. $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_s$). In our microbe example, indirect plasticity might be as simple as the presence of absence of a chemical metabolite in the microbe's environment. In the presence of the metabolite, expression of biochemical hardware to digest the metabolite confers a phenotypic advantage; in its absence, expressing that biochemical hardware is a useless exercise and, ultimately, a burdensome mistake. The presence of the metabolite itself signals which phenotypic form will enjoy high fitness.

Combining the evolutionary pressures exerted by indirect and direct plasticity, we can write that selective pressures will be directed towards a genotype $\dot{\hat{\vec{g}}}$ such that

$$\begin{aligned}
\vec{p}_1 &\approx f(\dot{\hat{\vec{g}}}, \vec{e} + \vec{e}_1 + \vec{\boldsymbol{n}}) \\
\vec{p}_2 &\approx f(\dot{\hat{\vec{g}}}, \vec{e} + \vec{e}_2 + \vec{\boldsymbol{n}}) \\
&\vdots \\
\vec{p}_s &\approx f(\dot{\hat{\vec{g}}}, \vec{e} + \vec{e}_s + \vec{\boldsymbol{n}}).
\end{aligned}$$

Observe that under $\dot{\hat{\vec{g}}}$ the developmental process f is tuned to be highly responsive to certain perturbations (i.e. $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_s$) but to disregard others (i.e. $\vec{\boldsymbol{n}}$). At the roughest level, one might begin to see a connection between phenotypic plasticity and the concepts of robustness and adaptability.

3.3 Relation to Evolvability

The exact role of phenotypic plasticity in evolution is an issue of active debate in the evolutionary biology community [Pigliucci, 2008]. However, several hypotheses describing how phenotypic plasticity might relate to evolution and evolvability have been put forward. Phenotypic plasticity might serve as a kind of local exploration of the phenotypic search space, allowing for the immediate expression of a phenotype with increased fitness and biasing the evolutionary search towards high-fitness regions of the search space [Downing, 2012]. It is also thought that the homeostatic mechanisms that mediate an organism's interactions with its environment might promote robustness [Moczek et al., 2011]. Researchers have suggested that phe-

notypic modularity might promote plasticity, especially in plants [Schlichting, 1986, De Kroon et al., 2005]. Thus, selection for plasticity might promote modularity. In these ways, plasticity might promote useful variability.

Conditional expression of phenotypic traits through plasticity allows for relaxed selection on the genotypic locus determining those traits. Thus, significant genetic variation can accumulate at that locus in a population. In a process known as genetic accommodation, the environmental influence on when rarely-expressed phenotypic traits are expressed can be diminished or erased through sensitizing mutation; what once was induced via environmental signal can become constitutive. Such processes have been observed experimentally via artificial selection [Moczek et al., 2011].

It is known that plasticity plays an important role in adaptation to unpredictable or variable environment [Fusco and Minelli, 2010]. Finally, plasticity might play a role in concert with indirect genetic encodings. Indirect encodings are biased towards phenotypic regularity [Clune et al., 2011] and plasticity might make available otherwise inaccessible phenotypic forms (i.e. providing a mechanism of irregular refinement of highly regular phenotypic structures generated from indirect encodings).

4 Experimental Proposal

Experiments will be performed using the genetic regulatory model. This model, motivated by the process by which a eukaryotic cellular genotype is transformed into a cellular phenotype, is standard in the field of computational evolutionary biology. Specifically, the two implementations by [Reisinger et al., 2005] and [Wilder and Stanley, 2015] will be used (experiments will be performed using both models). Fitness will be evaluated as the distance between a end-state chemical species expression profile and a target expression profile.

The relation of two experimental conditions – developmental variability and fitness criteria variability – to evolvability will be investigated. The developmental variability experimental condition will be realized through perturbation of the developmental process; in the case of the genetic regulatory model, this may be achieved through arbitrarily fixing the concentration of certain chemical species in the model, altering the initial concentrations of chemical species in the model, or altering the reaction kinetics (i.e. reaction rate) in the model. If time permits, this condition might also be realized by random perturbation directly applied to the genotype or phenotype.

The developmental variability experimental condition will be realized through systematic adjustment of fitness criteria in concert with small environmental perturbations (i.e. environmental signals). These perturbations will be realized in an identical manner to the previous experimental condition, but smaller in magnitude.

Evolvability will be measured through the canalization of the evolving system, that is the rate at which

fitness degrades under a regime of random mutation [Reisinger and Miikkulainen, 2007]. Evolvability will also be assessed through individual evolvability, the mean distance in phenotypic space between the offspring of an individuals, and through the population evolvability, the mean distance in phenotypic space between offspring of generation of individuals [Wilder and Stanley, 2015].

The relation of each of these two experimental conditions to evolvability will be assessed independently. If time permits, combinations of these plastic experimental conditions, variants of these plastic experimental conditions, or a third plastic experimental condition modeling local phenotypic search may be tried. Standard static adaptive selection will be used in these experiments. As time permits, other selective mechanisms may be tested, such as

- fluctuating adaptive selection,
- divergent selection (i.e. novelty search), and
- neutral selection (i.e. genetic drift).

4.1 Genetic Regulatory Network Model A

This model is employed in [Wilder and Stanley, 2015] and was originally developed in [Draghi and Wagner, 2009]. In this model, the genotype is a directed graph with k vertices where k is the number of transcriptional regulators considered in the model. Experiments in [Wilder and Stanley, 2015] used $k = 10$. Edge weights may only take on the values -1, 1, and 0 (i.e. if no edge exists). Thus, the genome may be represented by a $k \times k$ matrix W where W_{ij} represents the edge weight between transcriptional regulators i and j.

An individual's phenotype is the pattern of gene expression induced by the network. Beginning with expression State $S(0) = \vec{1}$, the following update rule is iteratively applied 500 times:

$$S_i(t+1) = \begin{cases} 1 & \text{if } \sum_{j=1}^{k} W_{ji} S_j(t) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

An individual is deemed viable if a fixed point is reached after 500 iterations (i.e. $S(500) = S(501)$). If no fixed point is reached, the individual is deemed inviable and receives a fitness score smaller than those that would be assigned to all possible viable individuals.

Two fitness regimes are employed by [Wilder and Stanley, 2015]. In the first, the adaptive selection regime, the hamming distance d between an individual's gene expression profile and a target profile is is used to determine fitness according to the relationship

$$F = \frac{1}{(1+s)^d}$$

where $s > 0$ is a parameter that tunes the strength of selection. (In the adaptive selection regime, d is taken as $k + 1$ for inviable individuals). The second fitness regime, divergent selection, calculates fitness according to the uniqueness of the individual's gene expression profile in relation to the rest of the profile. This non-objective-based mode of selection is a common trope in the field, as typified by novelty search [Lehman and Stanley,]. Under the divergent selection regime, fitness is calculated as

$$F = \frac{1}{(1+s)^{n_i}}$$

where n_i represents the number of individuals in the population that share a particular individual's gene expression profile. Although not explicitly stated in [Wilder and Stanley, 2015], it seems that inviable individuals would be assigned fitness $\frac{1}{(1+s)^{N+1}}$ where N is the size of the evolving population.

4.2 Genetic Regulatory Network Model B

This second genetic regulatory network model, employed in [Reisinger and Miikkulainen, 2007], is a bit more sophisticated. The genome is a set of if-then rules. The if component of these rules are activated by the current state of regulatory transcription factor concentrations. The chemical identity of each regulatory transcription factor is described by a value $v \in [0, 1]$. We denote the set of all transcription regulatory factors as V. Each gene has one or more associated regulatory factors, each of which has a chemical identity $c \in [-1, 1]$. The chemical match between a regulatory region c and a transcription regulatory factor v is computed as

$$E(c, v) = \exp -T \times (|c| - v)^2.$$

In this expression, the value T is an evolvable tolerance parameter, which controls how likely meaningful connections are to form between promoter and regulator chemical species. The activation level of each gene is given by a fuzzy matching between the set of regulatory transcription factors and the regulatory regions associated with that gene

$$P(G) = \sum_{c \in C} \sum_{v \in V} \text{sign}(c) \times \text{concentration}(v) \times E(c, v)$$

where the chemical identity of each regulatory region is signified by a value and C denotes the collection of all regulatory regions associated with a gene. Note that regulatory regions with $c > 0$ act as promoters while regulatory regions with $c < 0$ act as inhibitors. According to this chemical match, the concentrations

of the products are updated by the quantity

$$\delta \,\text{concentration}(v) = P(G)/\,\text{length}(c).$$

This update cycle is repeated for ten iterations. The concentrations of several certain chemical species, designated specially as environmental transcription factors, at the end of this process is taken as the phenotype of the individual. This GRN encoding, which has a variable number of variable length genes, is recombined during the evolutionary process using the NeuroEvolution of Augmenting Topologes (NEAT) method. Further details, including some parameters used in the model, can be found in [Reisinger et al., 2005].

4.3 Computational Description

Computational experiments will employ the open-source package Distributed Evolutionary Algorithms in Python (DEAP). This package is well suited to this project: it is actively maintained, well documented, and designed to enable "rapid prototyping and testing of ideas." DEAP will provide many components of the evolutionary algorithm right out of the box. Referring to Figure 1, the components that will be left to me to implement on my part will be "Decoding," "Evaluation," and "Selection." Decoding, synonymous with development, refers to the generation of a phenotype from a genotype. Evaluation refers to scoring the fitness of the phenotype. Selection refers to determining which members of a population will contribute genetic material to the next generation. Evaluation and selection will be extremely straightforward to implement.

The decoding will be the primary programming burden associated with the project. Experimental conditions will primarily be imposed through modifications to this aspect of the evolutionary process. The genetic regulatory model described in Section 4.1 will be extremely straightforward to implement, and will be tackled first. The model presented in Section 4.2 will be more challenging to implement. This challenge is two pronged: the decoding process is much more complex and the recombination process relies on NEAT, which will be tricky to interface with DEAP (although packages do exist to this end). This model was originally chosen because of the high profile of the paper in which it appeared [Downing, 2015, p 347]. However, upon further investigation of the NEAT-DEAP issue, I might opt for another model such as [Dwight Kuo et al., 2006] or [Quayle and Bullock, 2006] which share many characteristics of [Reisinger et al., 2005] and were also employed to study evolvability but rely on simpler genetic representation and recombination approaches. A cursory inspection revealed no obvious open source implementations of sophisticated evolving genetic regulatory network models. I would be pleased to make my implementation available upon the conclusion of the project. (This would be a goal of secondary priority, though).

5 Schedule

- Week 1: January 16th
  - Work on proposal, play with DEAP
- Week 2: January 23rd
  - Code and test first model
  - Decide definitively on second model
- Week 3: January 30th
  - Code and test second model (part I)
  - Run first model
- Week 4: February 6th
  - Code and test second model (part II)
  - Analyze data from first model
  - Plan adjustments to experimental protocol (either to pursue interesting questions raised by results or to try to get results)
- Week 5: February 13th
  - Code and test second model (part III)
- Week 6: February 20th
  - Run second model
- Week 7: February 27th
  - Analyze data from second model
- Week 8: March 6th
  - Further experimental tweaks and data collection
- Week 9: March 13th
  - Spring Break
  - Prepare for thesis presentation
  - Complete data collection

- Week 10: March 20th
  - Weds March 22nd: Honors thesis presentation
- Week 11: March 27th
  - Prepare for departmental presentation
  - Complete data analysis
- Week 12: April 3rd
  - Monday April 3rd: Math/CS Department presentation
  - thesis writing ☕☕☕
- Week 13: April 10th
  - more, and more frantic, thesis writing ☕☕☕☕☕
- Week 14: April 17th
  - Mon April 17th: Honors thesis document deadline
  - prepare for Math/CS Day presentation
  - adapt thesis material for Capstone report
- Week 15: April 24th
  - prepare for Math/CS Day presentation
  - adapt thesis material for Capstone report
  - Sat April 29th: Math/CS Day presentation
- Week 16: May 1st
  - odds and ends
  - Reading Period
  - odds and ends
- Week 17: May 8th
  - Finals Week
  - fin!

6 Miscellania

6.1 Challenges and Learning Opportunities

I chose to pursue this project because it directly relates to my own career interests. I hope to study in graduate school. This design of this project builds off of research that I conducted last semester through my thesis unit.

This experimental project poses a significant challenge – experimental inquiry in and of itself is unpredictable, so analysis of experimental results are typically a fraught process. Building a reasonably efficient and highly accurate implementation of the necessary computational models will also be a difficult task. Finally, communicating the results of my project, its evolution-theoretic basis, and its relevance to less biologically-inclined mathematicians and computer scientists as well as a more general audience will also be hard.

These hurdles will provide a valuable learning opportunity. The skills I develop conducting, analyzing, and presenting computational experiments will directly apply to my graduate career where I will be doing nearly the same things for (hopefully) a meager living. I am also genuinely hoping through this project to gain and share new insight that is of active interest to researchers in computer science and in evolutionary biology (as well as myself).

6.2 Deliverable Functionality

This project will not result in deliverable functionality in, perhaps, the traditional sense of these proposals. Although I intend to develop functional computational models of genetic regulatory networks and scripts to perform data analysis, these products will be of little use – or even interest – outside the scope of this project. Instead, my deliverables would best be characterized as my final presentation and report. Although not formally linked to this course, my Honors thesis document, Honors thesis presentation, and Math/CS Departmental presentation will provide additional outlets for deliverable production – and ensure that the communication-based deliverables that are formally associated with the capstone course are well-developed.

I would characterize the "basic" functionality as an inconclusive or negative result. This outcome, which is typical in experimental research, is not project failure. In this case, I will prepare and present a thoughtful analysis of my experimental protocol that attempts to account for the results and recommendations for future work. The attached schedule includes several weeks for experimental adjustment, which will ensure that several attempts can be made at trying to pinpoint what is at play behind the experimental results observed and, even if a positive result is not ultimately obtained, a series of experimental follow-ups that eliminate certain hypothesis and favor others will provide interesting – and meaningful – material to prevent. My "basic" functionality will be more than "Here's what I did. It didn't work."

My stretch functionality would be a positive result and further experimentation and analysis to better characterize/support that result or pursuing questions raised by the positive result. However, in this case, my deliverables will not be affected dramatically; I will present a discussion of my methods, my results, the implications of my results, and recommendations for future work.

7 Glossary

This section reviews a handful of terms that are essential to discussions of evolution and evolutionary algorithms.

7.1 Individual

Individual are the object upon which evolution operates; evolution evaluates and selects on individuals and recombines individuals to form new individuals. In biology, and individual is an individual organism such as a single tree or a single bird. In evolutionary algorithms, an individual is abstracted as a candidate solution to a problem.

7.2 Population

A population is a collection of individuals that compete to transmit their genetic information to the next generation. These individuals are typically highly similar and, in many cases in both biology and evolutionary algorithms, recombine their genetic information to produce offspring.

7.3 Phenotype

Phenotype refers to the characteristics of an individual that interact with its environment to determine its fitness. In biology, the physical form of an organism (i.e. its body) is the phenotype. In evolutionary algorithms, the phenotype refers to the characteristics of an individual that are evaluated during selection.

7.4 Genotype

Genotype refers to information that is used to determine the phenotype that is passed from generation to generation. In biology, a DNA sequence serves as the genotype. Although many different genotypic encodings are employed in evolutionary algorithms, in evolutionary algorithms the genotype ultimately boils down to a collection of digital information.

7.5 Recombination

Recombination refers to the generation of new genetic material from existing genetic material. This can involve combinations of two or more sets of genetic material, as in sexual reproduction, and/or random perturbation of genetic information (i.e. mutations).

7.6 Selection

Selection refers to the determination of which individuals will pass genetic material on to the next generation by creating offspring (and how many offspring they will be able to generate) and which will not.

7.7 Fitness

Fitness refers to the success of an individual at passing its genetic information to the next generation. An individual with high fitness creates many offspring while an individual with low fitness does not. Success at surviving challenges posed by the environment is an important factor in determining fitness. In evolutionary algorithms, the concept of fitness is abstracted to the fitness function where an individual is scored based on its aptitude at performing a certain task.

References

[Clune et al., 2011] Clune, J., Stanley, K. O., Pennock, R. T., and Ofria, C. (2011). On the performance of indirect encoding across the continuum of regularity. *IEEE Transactions on Evolutionary Computation*.

[De Kroon et al., 2005] De Kroon, H., Huber, H., Stuefer, J. F., and Van Groenendael, J. M. (2005). A modular concept of phenotypic plasticity in plants.

[Downing, 2012] Downing, K. L. (2012). Heterochronous Neural Baldwinism. *Artificial Life*, 13.

[Downing, 2015] Downing, K. L. (2015). *Intelligence emerging : adaptivity and search in evolving neural systems*. MIT Press, Palatino.

[Draghi and Wagner, 2009] Draghi, J. and Wagner, G. P. (2009). The evolutionary dynamics of evolvability in a gene network model. *Journal of Evolutionary Biology*.

[Dwight Kuo et al., 2006] Dwight Kuo, P., Banzhaf, W., and Leier, A. (2006). Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence. *Biosystems*, 85(3):177–200.

[Fusco and Minelli, 2010] Fusco, G. and Minelli, A. (2010). Phenotypic plasticity in development and evolution: facts and concepts. *Philosophical Transactions of The Royal Society*, 365:547–556.

[Griffiths et al., 2015] Griffiths, A. J. F., Wessler, S. R., Carroll, S. B., and Doebley, J. F. (2015). *Introduction to genetic analysis*.

[Kashtan et al., 2007] Kashtan, N., Noor, E., and Alon, U. (2007). Varying environments can speed up evolution. *Proceedings of the National Academy of Sciences*, 104(34):13711–13716.

[Lehman and Stanley,] Lehman, J. and Stanley, K. O. Exploiting Open-Endedness to Solve Problems Through the Search for Novelty.

[Mengistu et al., 2016] Mengistu, H., Lehman, J., and Clune, J. (2016). Evolvability Search: Directly Selecting for Evolvability in order to Study and Produce It. *GECCO Proceedings*.

[Moczek et al., 2011] Moczek, A. P., Sultan, S., Foster, S., Ledó N-Rettig, C., Dworkin, I., Nijhout, H. F., Abouheif, E., and Pfennig, D. W. (2011). The role of developmental plasticity in evolutionary innovation. *Proc. R. Soc. B*.

[Pigliucci, 2008] Pigliucci, M. (2008). Is evolvability evolvable? *Nature Reviews Genetics.*, 9(1):75–82.

[Poli et al., 2008] Poli, R., Langdon, W., McPhee, N., and Koza, J. (2008). *A field guide to genetic programming*.

[Prothmann et al., 2009] Prothmann, H., Schmeck, H., and Schmeck, H. (2009). Evolutionary algorithms for traffic signal optimisation. In *Proceedings of mobil.TUM 2009 - International Scientific Conference on Mobility and Transport*. TU München.

[Quayle and Bullock, 2006] Quayle, A. and Bullock, S. (2006). Modelling the evolution of genetic regulatory networks. *Journal of Theoretical Biology*, 238(4):737–753.

[Reisinger and Miikkulainen, 2007] Reisinger, J. and Miikkulainen, R. (2007). Acquiring Evolvability through Adaptive Representations. *GECCO'07 Proceedings*.

[Reisinger et al., 2005] Reisinger, J., Stanley, K. O., and Miikkulainen, R. (2005). Towards an Empirical Measure of Evolvability. *GECCO'05 Proceedings*.

[Schlichting, 1986] Schlichting, C. D. (1986). The Evolution of Phenotypic Plasticity in Plants. *Source: Annual Review of Ecology and Systematics Ann. Rev. Ecol. Syst*, 17(17).

[Sterling and Laughlin, 2015] Sterling, P. and Laughlin, S. (2015). *Principles of neural design*. MIT Press.

[Tuinstra et al., 1990] Tuinstra, E., De Jong, G., and Scharloo, W. (1990). Lack of response to family selection for direction asymmetry in Drosophila melanogaster: left and right are not distinguished during development. *Proc. R. Soc. Lond. B*, 241(1301):146–152.

[Wilder and Stanley, 2015] Wilder, B. and Stanley, K. (2015). Reconciling explanations for the evolution of evolvability. *Adaptive Behavior*, 23(3):171–179.