Theme Preservation and the Evolution of Representation

Keki M. Burjorjee
DEMO Lab
Computer Science Department, MS018
Brandeis University
Waltham, MA 02454
kekib@cs.brandeis.edu

Jordan B. Pollack
DEMO Lab
Computer Science Department, MS018
Brandeis University
Waltham, MA 02454
pollack@cs.brandeis.edu

ABSTRACT

In his thesis Toussaint calls for a "general project to develop theories on adaptation processes that account for the adaptation of representations". The theory developed in this paper is a contribution to this project. We first define the simple concept of a genotypic theme and define what it means for mutation operators to be theme preserving and theme altering. We use the idea of theme preservation to develop the concept of subrepresentation. Then we develop a theory that illuminates the behavior of a mutation-only fitness proportional evolutionary algorithm in which mutation preserves genotypic themes with high probability. Our theory shows that such evolutionary algorithms implicitly implement what we call *subrepresentation evolving multithreaded evolution*, i.e. such EAs conduct second-order search over a predetermined set of representations and exploit promising representations for first order evolutionary search. We illuminate subrepresentaiton evolving multithreaded evolution by comparing and contrasting it with the behavior of island model EAs. Our theory is immediately useful in understanding the significance of the low probability with which theme altering type 2 mutations are applied to genotypes of the evolutionary systems in Toussaint's thesis.

Categories and Subject Descriptors

I.2.8 [**Problem Solving, Control Methods, and Search**]: Heuristics; F.2 [**Analysis Of Algorithms and Problem Complexity**]: Miscellaneous

General Terms

Algorithms, Theory

Keywords

Evolutionary algorithms, theme preservation, evolution of representation

1. INTRODUCTION

In [8] Toussaint discusses the difference between adaptation mechanisms in the evolution of biological organisms and the adaptation mechanisms that are commonly employed in neural network research. In a section entitled "Complex adaptation mechanisms on arbitrary representations — or simple adaptation mechanisms on suitable representations" he argues that when compared to highly sophisticated adaptation techniques such as backprop, Rprop, conjugate gradient descent etc., the trial-and-error strategy - at the level of genes - that nature seems to use is quite elementary. He observes that despite the simplicity of the adaptation mechanism that it uses, nature is remarkably successful in generating complex functional changes at the phenotypic level.

For an example of a complex functional change consider the experiment in [2] (also cited in [8]) in which Halder et al. induced the misexpression of a single gene in a Drosophilia Melanogaster called "eyeless" and obtained flies with eyes on their wings, legs and antennae. In [10] Wagner and Altenberg make the following remarks about this experiment.

> "The out-of-place eyes contain the entire eye structures, including cornea, bristles, pigment and photoreceptors, and are electrically responsive to light, prompting Halder et al. to suggest that eyeless is a "master control" for the complex formation of the insect eye.
>
> Why is "eyeless" so remarkable? Because it is a single signal that induces the whole complex process of eye construction, and because this process is carried out almost flawlessly despite it occurring in the wrong tissues of the fly's body. All of the functionally relevant structures stay together in their novel locations."

An eye is a very complex organ made up of many functionally coupled parts. The absence of any of these parts greatly compromises its usefulness. The experiment of Halder et. al. shows that the the genome of the Drosophilia Melanogaster is organized to reflect this fact, i.e. the fly is genetically represented in such a way that there exist genes, such as "eyeless" whose mutants preserve "the relationships between the functionally interdependent parts of the eye, while changing only the eye's relationship to the rest of the fly's body" [10].

How is it that in nature non-complex changes at the genotypic level, produce complex functional changes at the

GECCO'05, June 25–29, 2005, Washington, DC, USA.

phenotypic level (such as the change to the fly phenotype wrought by the misexpression of eyeless)? Echoing the views expressed in [10], Toussaint deduces that in order for this to be so, nature must have found a *suitable* way of representing phenotypes. He further remarks that "Nature developed these [suitable] representations in the course of adaptive processes and it *thus should be a general project to develop theories on adaptation processes that account for the adaptation of representations*." (emphasis added).

This paper is a contribution to the "general project" that Toussaint calls for. We formally analyze mutation-only fitness proportional evolutionary algorithms in which some aspect of the genome is mutated with much higher probability and show that such evolutionary algorithms implicitly implement what we call *subrepresentation evolving multithreaded evolution*, i.e. such EAs conduct second-order search over a predetermined set of representations and exploit promising representations for first order evolutionary search. Our theory is immediately useful in understanding the significance of the low probability with which theme altering type 2 mutations are applied to genotypes of the evolutionary systems in [8].

2. MATHEMATICAL PRELIMINARIES

All sequences in this paper are zero based (the index of the first element is zero) and infinite. Let X be some set. Then we denote some sequence of elements in X by $\{p^n\}_{n\geq 0}$. For any $i \in \mathbb{N}$, we denote the element with index i in $\{p^n\}_{n\geq 0}$ by p^i. For some sets X, Y, and some function $\gamma : X \to Y$, we use the notation $\langle y \rangle_\gamma$ to denote the set $\{x \in X \,|\, \gamma(x) = y\}$. We will drop the subscript γ from this notation when it is clear from the context.

As in [8], for any set X we use the notation Λ^X to denote the set of all probability distributions over X, i.e. Λ^X denotes set $\{f : X \to [0,1] \mid \sum_{x\in X} f(x) = 1\}$.

We extend this notation to denote the set of all 1-parent transmission functions (see [1]) over some set as follows: for any set X, the set of all 1-parent transmission functions $\{f : X \times X \to [0,1] \mid \forall x' \in X, \sum_{x\in X} f(x,x') = 1\}$ is denoted by Λ_1^X. Employing the notation used in [8], we use conditional probability notation to denote a 1-parent transmission function (henceforth transmission function). Thus a transmission function $f(x,x')$ is denoted $f(x|x')$. Transmission functions will be used to modelling the effect of a mutation operator. For some mutation operator $\mathbf{M}$ over some genotype set G we can define a transmission function $M \in \Lambda_1^G$ such that for any $g, g' \in G$, $M(g|g')$ is the probability that $\mathbf{M}$ will yield the genotype g when applied to g'.

3. THEMES

Let G be some set of genotypes and let K be some set of objects that codify "properties" that are possessed by elements of G. If the properties are such that every genotype in G possess one and only one property in K (the properties are mutually exclusive), then we say that K is a theme set of G. We call the properties in K themes, and for any $k \in K$ we call the subset of genotypes that map to k the theme class of k. While we could have expressed these ideas in the terminology of mathematical functions, we believe that the terminology just introduced is better suited to our exposition. The correspondence between function terminology and our terminology is made clear in the following definition.

Definition 3.1. (Theme Map, Theme, Theme Set, Theme Class) *Let X, Y be sets and let $\beta : X \to Y$ be a function. We call β a theme map, call the co-domain Y of β a β-theme set, call an element of Y a β-theme, and call the preimage $\langle y \rangle$ of some $y \in Y$, the β-*theme class *of y.*

Remark 1. *Given the objects defined above, it is easily seen that the set of all β-theme classes form a partition of X*

The idea of a theme class is mathematically identical to the idea of a forma discussed in [4] - each of these objects is simply an equivalence class which belongs to the partition that is induced by some function. However the *application* of this mathematical idea in this paper differs in spirit from its application in [4]. There a forma describes some equivalence class of phenotypes, whereas in this paper a theme class is an equivalence class of genotypes.

The idea of a theme class is also similar to the idea of a schema [3], however there are differences between the two. Firstly, schemata are defined on a case by case basis for different genotypic data-structures. Theme classes on the other hand have been defined here in an abstract data-structure-independent way. Secondly, a genotype may belong to many schemata whereas for a given theme map, a genotype will only belong to one and only one theme class.

3.1 Theme Preservation and Alteration

Let G, K be sets such that G is countable and let $\beta : G \to K$ be some function. For some mutation operator that operates on elements of G, we say that this operator is β-preserving if it leaves the β-themes of of its argument unchanged, i.e. the child produced by the mutation operator will always have the same β-theme as its parent. We say that the mutation operator is β-altering if it *always* changes the β-theme of its argument. Examples of theme-preserving mutation operators for three different kinds of genotype data-structures — bitstrings, S-expressions and L-Systems are given below. The genotype set and theme set of each operator is described in the first two columns of table 1. The 3$^{\text{rd}}$ and 4$^{\text{th}}$ columns of the table schematically show the effect of the three mutation operators on three sample genotypes, and the last column schematically shows the theme that is preserved in each case. We leave it to the reader to think of the theme map, and a theme altering mutation operator in each case.

1. A mutation operator which operates on bitstrings of length ten and only modifies the last six bits of its argument.
2. A mutation operator which takes a S-expression for a polynomial as an argument and changes the tree structure of the S-expression in some way while leaving the values of the nodes unchanged.
3. A mutation operator which takes a seeded L-system[1] over the alphabet $\{a, b, c, d\}$ as its argument and *substitutes* a symbol for another symbol in the seed string

[1]In this paper we call a tuple consisting of 1) an L-system over some alphabet Σ and 2) a string in Σ^*, a seeded L-System. The genotypes in the evolutionary algorithms in [8] are seeded L-Systems

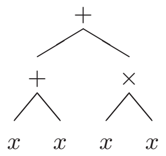
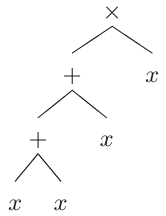
Genotype Set G	Theme set K	Sample Parent	Sample Child	Theme Preserved
Bitstrings in $\{0,1\}^{10}$	Strings in $\{0,1\}^4$??????	1011010111	1011100011	1011??????
S-expressions with binary tree structures that express single variable polynomials, i.e. binary trees with leafs x and internal nodes drawn from $\{+,\times\}$	Any $k \in K$ is a multiset of elements from $\{x,+,\times\}$ s.t. there exists a binary tree structure which uses all the elements in k			The multiset $\{\times,+,+,x,x,x,x\}$
Seeded L-Systems with terminals drawn from the alphabet $\Sigma=\{a,b,c,d\}$	The set of all seeded L-System "skeletons"	$[\langle abbcbd\rangle$ $(a\to b)$ $(b\to bcab)$ $(c\to ac)]$	$[\langle acbdbd\rangle$ $(a\to a)$ $(b\to ccab)$ $(c\to bc)]$	$[\langle$??????$\rangle$ $(a\to$?$)$ $(b\to$????$)$ $(c\to$??$)]$

Table 1: Three different theme-preserving mutation operators

and the right hand sides of the rewrite rules - it does not add or delete rewrite rules, does not change the number of symbols in a seed string or rewrite rule, and does not change the left hand side of the rewrite rules.

Theme preserving and altering mutation can be modelled by placing constraints on transmission functions.

Definition 3.2. (Preserving and Altering Transmission Functions) *Let X, Y be some sets, let $\beta : X \to Y$ be a function, let $M \in \Lambda_1^X$ be a transmission function. We say that M is β-*preserving *if*

$$\forall x, x' \in X, \quad \beta(x) \neq \beta(x') \Rightarrow M(x|x') = 0$$

*and say that M is β-*altering *if*

$$\forall x, x' \in X, \quad \beta(x) = \beta(x') \Rightarrow M(x|x') = 0$$

The following proposition gives us a useful property of a preserving transmission function.

Proposition 3.1. *Let X, Y be some sets, let $\beta : X \to Y$ be a function, let $M \in \Lambda_1^X$ be a β-theme-preserving transmission function. Then,*

$$\forall y \in Y, \forall x' \in \langle y\rangle \qquad \sum_{x\in\langle y\rangle} M(x|x') = 1$$

Proof. By definition of a transmission function we have that $\forall x' \in X$, $\sum_{x\in X} M(x|x') = 1$. But by (def 3.2), $\forall y \in Y$, $\forall x' \in \langle y\rangle$ and $\forall x \notin y$, $M(x|x') = 0$. Hence $\sum_{x\in\langle y\rangle} M(x|x') = 1$ $\square$

4. TRANSMAPS

In this paper we will focus our attention on EAs in which mutation is the only form of variation and selection is fitness proportional. We will call such EAs basic fitness proportional EAs or bfpEAs for short. The representation used by some bfpEA to search some set of objects is determined by 1) an alternate set of software objects called genotypes, 2) a function that maps genotypes to objects in the search space, and 3) the mutation operator that stochastically produces some child genotype given some parent genotype. A transmap, defined below, is our model for a representation.

Definition 4.1. (Transmap) *A transmap is a 4-tuple (G, P, ϕ, M) such that G is a countable set called the genotype set, P is some set called the phenotype set, $\phi : G \to P$ is called the growth map, and $M \in \Lambda_1^G$ is a 1-parent transmission function.*

Let P be some set of objects (e.g. sorting networks, polynomial functions, plant morphologies, etc.) Then, given some representation for P we can construct a transmap D that models this representation as follows: 1) the search space P of the representation is the phenotype set of D, 2) the alternate set of software objects is the genotype set of D, 3) the function that maps the software objects to objects in the search space is the growth map of D, 4) the effect of the mutation operator of the representation is modelled by the transmission function of D.

4.1 Subtransmaps

In this section we show how, a transmap D with a β-preserving transmission function, determines a set of transmaps such that each transmap in this set is in one-to-one correspondence with some β-theme. The transmaps in this set are called subtransmaps of D.

For some function $f : X \to Y$, and some $A \subset X$, the restriction of f to A is denoted $f|_A$. We extend the notion of restriction to functions whose domain is the cross-product of the same set as follows: for a function $g : X \times X \to Y$, the restriction of g to A, denoted $g|_A$ is a function of type $A \times A \to Y$ such that for any $a_1, a_2 \in A$, $g|_A(a_1, a_2) = g(a_1, a_2)$.

Proposition 4.1. *Let G, K be sets, let $\beta : G \to K$ be some function and let $M \in \Lambda_1^G$ be a β-preserving transmission function. Then, for any $k \in K$, $M|_{\langle k\rangle} \in \Lambda^{\langle k\rangle}$*

Proof. *For any k in K, and any $g' \in \langle k\rangle$, $\sum_{g\in G} M(g|g') = 1$. But by* (def 3.2)*, $\forall g \notin \langle k\rangle$, $M(g|g') = 0$. So, $\sum_{g\in\langle k\rangle} M(g|g') = 1$, which implies that $\sum_{g\in\langle k\rangle} M|_{\langle k\rangle} (g|g') = 1$* $\square$

Definition 4.2. ((β, k)-Subtransmap) *Let $D = (G, P, \phi, M)$ be a transmap s.t. for some set K, $\beta : G \to K$ is a function and M is β-preserving. We define the (β, k)-subtransmap of D, denoted $D|_{\langle k\rangle}$, to be the transmap $(\langle k\rangle , P, \phi|_{\langle k\rangle} , M|_{\langle k\rangle})$.*

To see that $(\langle k\rangle, P, \phi|_{\langle k\rangle}, M|_{\langle k\rangle})$ is indeed a transmap note that $M|_{\langle k\rangle} \in \Lambda_1^{\langle k\rangle}$ by proposition 4.1 and $\phi|_{\langle k\rangle}$ is of type $\langle k\rangle \to P$ by definition of restriction.

For some $\beta : G \to K$, suppose R is a representation with a mutation operator that preserves β, suppose B is a bfpEA which uses R and for some theme k, suppose all the genotypes in the initial population of B have the same theme $k \in K$. Then, as the mutation operator of B is β-preserving, all genotypes in all the generations of an evolutionary run of B will have theme k. Therefore, we can define a new representation R_k which is isomorphic to R over $\langle k\rangle$ by "pulling" the theme k out of the genotypes in $\langle k\rangle$ and "pushing" it into the mutation operator and growth function of R_k s.t. when B uses R and starts with a k-themed initial population, its search behavior is the same as if it used R_k and started with an isomorphic initial population in the genotype set of R_k. Let us call R_k a subrepresentation of R. Then a subtransmap, as defined above, is a model of a subrepresentation.

4.2 (β, ω)-Preserving Transmaps

Definition 4.3. (Rate Operator) *Let $M^1, M^2 \in \Lambda_1^G$ be transmission functions. For any $\omega \in [0,1]$ we define the* Rate Operator $\mathcal{R}^\omega : \Lambda_1^G \times \Lambda_1^G \to \Lambda_1^G$ *as follows:*

$$\forall g, g' \in G,\ \mathcal{R}^\omega(M^1, M^2)(g, g') = \omega M^1(g|g') + (1-\omega) M^2(g|g')$$

To see that for any set G and $\omega \in [0,1]$, the range of $\mathcal{R}^\omega$ is indeed Λ_1^G, observe that $\forall M^1, M^2 \in \Lambda_1^G, \forall g' \in G$,

$$\sum_{g\in G} \mathcal{R}^\omega(M^1, M^2)(g|g') = \omega \left(\sum_{g\in G} M^1(g|g')\right) + (1-\omega)\left(\sum_{g\in G} M^2(g|g')\right) = \omega + (1-\omega) = 1 \quad (1)$$

Definition 4.4. ((β,ω)-preserving Transmap) *Let $D = (G, P, \phi, M)$ be a transmap, for some set K, let $\beta : G \to K$ be a function and let $\omega \in [0,1]$. D is said to be (β,ω)-preserving if there exists $M_1, M_2 \in \Lambda_1^G$ such that M^1 is β-preserving, M^2 is β-altering and $M = \mathcal{R}^\omega(M_1, M_2)$.*

Remark 2. *Note that for any (β,ω)-preserving transmap (G, P, ϕ, M), the preserving and altering components of M (M_1 and M_2 in the definition above) are unique. We denote them as M_P and M_A respectively.*

Suppose mutation in some representation is β-preserving with some probability ω and β-altering mutation with probability $(1-\omega)$, then it is easy to see how we could model such a representation using a (β,ω)-preserving transmap.

5. EVOLUTION MACHINES

Definition 5.1. (Evolution Machine). *An evolution machine – which we also call an EM – is a 3-tuple (D, f, s) such that $D = (G, P, \phi, M)$ is a transmap, $f : P \to \mathbb{R}^+$ is called the fitness function and $s \in \Lambda^G$ is called the* initial genotype distribution.

An evolution machine is a collection of all the formal objects needed to model an evolutionary run of a bfpEA. A bfpEA, which was introduced in section 4, is similar to a Simple Genetic Algorithm as defined in [9] in all respects except that 1) it performs fitness proportional selection (an SGA may use other selection methods), 2) its genotypes may be instances of arbitrary datatypes (SGAs use only bitstrings), and 3) mutation is its only variational operator (SGAs also use a recombination operator).

Populations in a bfpEA are modelled as distributions of an EM. In order to define how these distributions change from generation to generation we recall, and in the process extend, the following operators from [9] and [8].

Definition 5.2. (Selection Operator) *Let X be some set and let $f : X \to \mathbb{R}^+$ be some function. We define the* Selection Operator $\mathcal{S}_f : \Lambda^X \to \Lambda^X$ *as follows:*

$$(\mathcal{S}_f p)(x) = \frac{f(x)p(x)}{\sum\limits_{x'\in X} f(x')p(x')}$$

The selection operator is parameterized by a fitness function. It takes a distribution p_X over some set X as its argument and redistributes the probability mass of the distribution over the elements of X in proportion to the fitness of the elements and their probability mass in p_X. In typical usage of $\mathcal{S}$ in the literature, the set X is the genotype set. In this paper $\mathcal{S}$ will also be used to express meta-selection applied to a distribution over a theme set. The precise sense in which we use the phrase meta-selection will become clear later on.

Definition 5.3. (Expected Fitness Operator) *Let X be some set, and $f : X \to \mathbb{R}^+$ be some function. We define the* expected fitness operator $\mathcal{E}_f : \Lambda^X \to \mathbb{R}^+$ *as follows:*

$$\mathcal{E}_f(p) = \sum_{x\in X} f(x)p(x)$$

The expected fitness operator will be useful in defining the theme fitness function later on. It can also be used to express the selection operator more compactly as follows.

Remark 3. *The selection operator can be expressed in terms of the Expected Fitness Operator as follows:*

$$(\mathcal{S}_f p)(x) = \frac{f(x)p(x)}{\mathcal{E}_f(p)}$$

Definition 5.4. (Transmission Operator[2]) *Let X be a set, and let $M \in \Lambda_1^X$ be a transmission function over X. We define the transmission operator $\mathcal{T}_M : \Lambda^X \to \Lambda^X$ as follows:*

$$(\mathcal{T}_M p)(x) = \sum_{x'\in X} M(x|x')p(x')$$

The transmission operator will be used to model the effect of mutation on the genotypes that are selected as parents in each generation of a bfpEA

[2]also called the Mixing Operator in [9] and [8]

Definition 5.5. (Evolution Epoch Operator) *Let $D = (G, P, \phi, T)$ be a transmap, and let $f : P \to \mathbb{R}^+$ be some function. We define the evolution epoch operator $\mathcal{G}_{(D,f)} : \Lambda^G \to \Lambda^G$ as follows:*

$$\mathcal{G}_{(D,f)}(p) = \mathcal{T}_M \circ \mathcal{S}_{f\circ\phi}\, p$$

Given some bfpEA, an evolution epoch operator that is parameterized by the bfpEA's representation and fitness function models the advancement by one generation of a population of genotypes in the bfpEA. In section 6 we will see that when the mutation operator of a bfpEA is theme preserving, this operator can be used to express the advancement of a sub-population of genotypes that share the same theme.

The following definition associates a sequence of genotypic distributions with an EM using the operators we defined above. This sequence is a model of the generations of genotypic populations that are generated by a bfpEA.

Definition 5.6. (Genotype Distribution Sequence of an EM). *Let $E = (D, f, s)$ be some evolution machine. The genotype distribution sequence of E is a sequence $\{p^t\}_t$ of elements in Λ^G s.t. $p^0 = s$ and for any $t \in \mathbb{N}$:*

$$p^{t+1} = \mathcal{G}_{(D,f)} p^t$$

5.1 (β, ω)-preserving EMs

For some β and some ω, if the transmap of an EM E is (β, ω)-preserving then we say that E is (β, ω)-preserving. Consider a bfpEA B, such that mutation of any genotype in B is β-preserving with probability ω and β-altering with probability $(1-\omega)$. Clearly, B can be modelled by a (β, ω)-preserving EM. Hence we call B a (β, ω)-preserving bfpEA.

6. ANALYSIS OF A (β, ω)-PRESERVING EM

The following definition recalls the projection operator described in [9] and [8]. A projection operator projects a distribution over the domain of some function to the range of that function. The projection function is typically used to project distributions over the genotype set to the phenotype set. Here we will also use it to project genotype distributions onto theme sets.

Definition 6.1. (Projection Operator) *Let X, Y be some sets and let $\gamma : X \to Y$ be a function. We define the projection operator, $\Xi_\gamma : \Lambda^X \to \Lambda^Y$ as follows:*

$$(\Xi_\gamma p\ )(y) = \sum_{x \in \langle y \rangle} p(x)$$

We call $\Xi_\gamma p$ the γ-projection of p. To see that the range of Ξ_γ is indeed Λ^Y, i.e. that a projected distribution is also a distribution, note that for any $p \in \Lambda^X$,

$$\sum_{y\in Y} \sum_{x \in \langle y \rangle} p(x) = \sum_{x \in X} p(x) = 1 \tag{2}$$

Given a countable set X, some set Y, a distribution p_X over X and some function-map $\gamma : X \to Y$, then for any element $y \in Y$ such that $(\Xi_\gamma p)(y) > 0$ we can define a new distribution over $\langle y \rangle$ by normalizing the probability mass of elements in $\langle y \rangle$ by the sum of their probability masses. We call this new distribution the γ-conditional distribution of p given k. Formally,

Definition 6.2. (Projection Conditional Distribution) *Let X, Y be sets and let $\gamma : X \to Y$ be a function. Let $p \in \Lambda^X$ be some distribution. For any $y \in Y$ such that $(\Xi_\gamma p)(y) > 0$, we define the γ-conditional distribution of p given y, to be a distribution $q \in \Lambda^{\langle y \rangle}$ s.t.*

$$q(x) = \frac{p(x)}{(\Xi_\gamma p)(y)}$$

Definition 6.3. (Theme Fitness Function) *Let G, K be sets, let $\beta : G \to K$ be a function, let E be an evolution machine with genotype set G, and let $\{p^t_G\}_{t\geq 0}$ be the genotype distribution sequence of E. For any $t \in \mathbb{Z}^+_0$, let p^t_K be the β-projection of p^t_G, and for all $k \in K$ such that $p^t_K(k) > 0$, let $p^t_{\langle k \rangle}$ be the projection conditional distribution of p^t_G given k. Then the β-theme fitness function of E at step t, ${}^\beta\mathfrak{F}^t_E : K \to \mathbb{R}^+$ is as follows:*

$${}^\beta\mathfrak{F}^t_E(k) = \begin{cases} \mathcal{E}_{f\circ\phi|_{\langle k \rangle}}({}^\omega p^t_{\langle k \rangle}) & \text{if } {}^\omega p^t_K(k) > 0 \\ 0 & \text{otherwise} \end{cases}$$

We now analyze a (β, ω)-preserving EM. For some $\omega \in [0, 1]$, let $D^\omega = (G, P, \phi, M)$ be a transmap such that for some set K and some $\beta : G \to K$, D^ω is (β, ω)-preserving. Let $E^\omega = (D^\omega, f, s)$ be a (β, ω)-preserving EM, and let $\{{}^\omega p^t_G\}_{t\geq 0}$ be the genotype distribution sequence of E^ω.

For any $t \in \mathbb{Z}^+_0$, and any $\omega \in [0, 1]$, let ${}^\omega p^t_K \in \Lambda^K$ be the projection of the genotype distribution $\{{}^\omega p^t_G\}_{t\geq 0}$ to the theme set K, i.e.

$${}^\omega p^t_K = \Xi_\beta({}^\omega p^t_G) \tag{3}$$

For any $t \in \mathbb{Z}^+_0$, any $\omega \in [0, 1]$ and any theme $k \in K$ such that ${}^\omega p^t_K(k) > 0$, let ${}^\omega p^t_{\langle k \rangle} \in \Lambda^{\langle k \rangle}$ be the β-conditional distribution of ${}^\omega p^t_G$ given k, i.e.

$${}^\omega p^t_{\langle k \rangle}(g) = \frac{{}^\omega p^t_G(g)}{{}^\omega p^t_K(k)} \tag{4}$$

6.1 Analysis for $\omega = 1$

In this subsection our analysis will focus on the sequences $\{{}^\omega p^t_K\}_{t\geq 0}$ and the family of sequences $\{\{{}^\omega p^t_{\langle k \rangle}\}_{t\geq 0} : k \in K\}$ when ω is 1. When $\omega = 1$ we will drop the superscript 1 that precedes the p in our notion for probability distributions. Thus, for example we will denote ${}^1 p^t_G$ as p^t_G.

For any $t \in \mathbb{Z}^+_0$, $k \in K$, by the definition of projection in (def 6.1) we get

$$p^{t+1}_K(k) = \sum_{g \in \langle k \rangle} p^{t+1}_G(g)$$

By definitions of the genotype distribution sequence of an EM and the evolution epoch operator, (defs 5.6 and 5.5), and using the fact that $\omega = 1$,

$$\begin{aligned} p^{t+1}_K(k) &= \sum_{g\in\langle k \rangle} \sum_{g' \in G} M_P(g|g')(\mathcal{S}_{f\circ g} p^t_G)(g') \\ &= \sum_{g' \in G} (\mathcal{S}_{f\circ\phi} p^t_G)(g') \sum_{g \in \langle k \rangle} M_P(g|g') \end{aligned}$$

For any $k \in K$ any $g \in \langle k \rangle$, and any $g' \notin \langle k \rangle$, by definition of a Preserving Transmission Function in (def 3.2), $M_P(g|g') = 0$. So,

$$p^{t+1}_K(k) = \sum_{g' \in \langle k \rangle} (\mathcal{S}_{f\circ\phi} p^t_G)(g') \sum_{g \in \langle k \rangle} M_P(g|g')$$

By proposition 3.1,

$$p_K^{t+1}(k) = \sum_{g' \in \langle k \rangle} (\mathcal{S}_{f\circ\phi} p_G^t)(g')$$

By definition of the Selection Operator in terms of the Expected Fitness Operator in (remark 3),

$$p_K^{t+1}(k) = \sum_{g' \in \langle k \rangle} \frac{f\circ\phi(g') \cdot p_G^t(g')}{\mathcal{E}_{f\circ\phi}(p_G^t)}$$

We examine the following two cases, **case i:** k such that $p_K^t(k) = 0$. This implies that for all $g \in \langle k \rangle$, $p_G^t(g) = 0$, so using the equation above, $p_K^{t+1}(k) = 0$.

case ii: k such that $p_K^t(k) > 0$. Using (4) in the numerator

$$\begin{aligned} p_K^{t+1}(k) &= \sum_{g' \in \langle k \rangle} \frac{f\circ\phi(g') p_K^t(k) p_{\langle k \rangle}^t(g')}{\mathcal{E}_{f\circ\phi}(p_G^t)} \\ &= \frac{p_K^t(k)\sum_{g' \in \langle k \rangle} f\circ\phi(g') p_{\langle k \rangle}^t(g')}{\mathcal{E}_{f\circ\phi}(p_G^t)} \\ &= \frac{p_K^t(k)\sum_{g' \in \langle k \rangle} f\circ\phi|_{\langle k \rangle}(g') p_{\langle k \rangle}^t(g')}{\mathcal{E}_{f\circ\phi}(p_G^t)} \end{aligned}$$

where the last equation follows from the definition of restriction. Using the Expected Fitness Operator defined in (def 5.3) to express the numerator,

$$p_K^{t+1}(k) = \frac{p_K^t(k)\mathcal{E}_{f\circ\phi|_{\langle k \rangle}}(p_{\langle k \rangle}^t)}{\mathcal{E}_{f\circ\phi}(p_G^t)} \tag{5}$$

By expansion of the Expected Fitness Operator in the denominator using (def 5.3),

$$p_K^{t+1}(k) = \frac{p_K^t(k)\mathcal{E}_{f\circ\phi|_{\langle k \rangle}}(p_{\langle k \rangle}^t)}{\sum_{g' \in G} f\circ\phi(g') p_G^t(g')}$$

Using (4) and (3) in the denominator,

$$\begin{aligned} p_K^{t+1}(k) &= \frac{p_K^t(k)\mathcal{E}_{f\circ\phi|_{\langle k \rangle}}(p_{\langle k \rangle}^t)}{\sum_{k' \in K}\sum_{g' \in \langle k' \rangle} f\circ\phi(g') p_K^t(k') p_{\langle k' \rangle}^t(g')} \\ &= \frac{p_K^t(k)\mathcal{E}_{f\circ\phi|_{\langle k \rangle}}(p_{\langle k \rangle}^t)}{\sum_{k' \in K} p_K^t(k')\sum_{g' \in \langle k' \rangle} f\circ\phi(g') p_{\langle k' \rangle}^t(g')} \end{aligned}$$

Using the Expected Fitness Operator defined in (def 5.3) to express the denominator,

$$p_K^{t+1}(k) = \frac{p_K^t(k)\mathcal{E}_{f\circ\phi|_{\langle k \rangle}}(p_{\langle k \rangle}^t)}{\sum_{k' \in K} p_K^t(k')\mathcal{E}_{f\circ\phi|_{\langle k' \rangle}}(p_{\langle k' \rangle}^t(g'))}$$

Hence for all $k \in K$, using the definition of the Theme Fitness Function in (def 6.3),

$$p_K^{t+1}(k) = \frac{p_K^t(k)\, {}^{\beta}\mathfrak{F}_{E^1}^t(k)}{\sum_{k' \in K} p_K^t(k')\, {}^{\beta}\mathfrak{F}_{E^1}^t(k')}$$

Using the definition of the Selection Operator (def 5.2), we obtain the following result

$$p_K^{t+1} = \mathcal{S}_{\,{}^{\beta}\mathfrak{F}_{E^1}^t}(p_K^t) \tag{6}$$

Equation (6) is our first important result. It shows that the EM E^1 performs meta-selection on the instantiated themes in each generation, i.e. the probability mass of a some instantiated theme in the next generation is proportional to its theme fitness in the current generation. We shall call (6) the *Theme Meta-Selection Equation.*

For any $t \in \mathbb{Z}_0^+$, any $k \in K$ s.t. $p_K^t(k) > 0$, and any $g \in \langle k \rangle$ we now analyze $p_{\langle k \rangle}^t$. By (4),

$$p_{\langle k \rangle}^{t+1}(g) = \frac{p_G^{t+1}(g)}{p_K^{t+1}(k)}$$

Using the definitions of the genotype distribution sequence of an EM and the evolution epoch operator, (defs 5.6 and 5.5) in the numerator,

$$p_{\langle k \rangle}^{t+1}(g) = \frac{\sum_{g' \in G} M_P(g|g')(\mathcal{S}_{f\circ\phi} p_G^t)(g')}{p_K^{t+1}(k)}$$

Expanding the Selection Operator in the numerator using (remark 3),

$$p_{\langle k \rangle}^{t+1}(g) = \frac{\sum_{g' \in G} M_P(g|g') f\circ\phi(g') p_G^t(g')}{\mathcal{E}_{f\circ\phi}(p_G^t) p_K^{t+1}(k)}$$

By definition of a Preserving Transmission Function in (def 3.2),

$$p_{\langle k \rangle}^{t+1}(g) = \frac{\sum_{g' \in \langle k \rangle} M_P(g|g') f\circ\phi(g') p_G^t(g')}{\mathcal{E}_{f\circ\phi}(p_G^t) p_K^{t+1}(k)}$$

Using (4) and (3) in the numerator,

$$\begin{aligned} p_{\langle k \rangle}^{t+1}(g) &= \frac{\sum_{g' \in \langle k \rangle} M_P(g|g') f\circ\phi(g') p_K^t(k) p_{\langle k \rangle}^t(g')}{\mathcal{E}_{f\circ\phi}(p_G^t) p_K^{t+1}(k)} \\ &= \frac{p_K^t(k)\sum_{g' \in \langle k \rangle} M_P(g|g') f\circ\phi(g') p_{\langle k \rangle}^t(g')}{\mathcal{E}_{f\circ\phi}(p_G^t) p_K^{t+1}(k)} \end{aligned}$$

Expanding $p_K^{t+1}(k)$ in the denominator using (5),

$$p_{\langle k \rangle}^{t+1}(g) = \frac{p_K^t(k)\sum_{g' \in \langle k \rangle} M_P(g|g') f\circ\phi(g') p_{\langle k \rangle}^t(g')}{\mathcal{E}_{f\circ\phi}(p_G^t)} \times \frac{\mathcal{E}_{f\circ\phi}(p_G^t)}{p_K^t(k)\mathcal{E}_{f\circ\phi}(p_{\langle k \rangle}^t)}$$

Cancelling terms and consolidating yields

$$\begin{aligned} p_{\langle k \rangle}^{t+1}(g) &= \frac{\sum_{g' \in \langle k \rangle} M_P(g|g') f\circ\phi(g') p_{\langle k \rangle}^t(g')}{\mathcal{E}_{f\circ\phi}(p_{\langle k \rangle}^t)} \\ &= \frac{\sum_{g' \in \langle k \rangle} M_P|_{\langle k \rangle}(g|g') f\circ\phi|_{\langle k \rangle}(g') p_{\langle k \rangle}^t(g')}{\mathcal{E}_{f\circ\phi|_{\langle k \rangle}}(p_{\langle k \rangle}^t)} \end{aligned}$$

where the last equation follows from the definition of restriction. Using the definition of the Selection Operator in (def 5.2),

$$p_{\langle k \rangle}^{t+1}(g) = (\mathcal{T}_{M_P|_{\langle k \rangle}}\mathcal{S}_{f\circ\phi|_{\langle k \rangle}} p_{\langle k \rangle}^t)(g)$$

As M_P is theme preserving, by (prop 4.1), $(\langle k \rangle, P, M_P|_{\langle k \rangle}, \phi|_{\langle k \rangle})$ is a transmap. Hence, using the definition of the Evolution Epoch Operator in (def 5.5),

$$p_{\langle k \rangle}^{t+1} = \mathcal{G}_{((\langle k \rangle, P, M_P|_{\langle k \rangle}, \phi|_{\langle k \rangle}), f)} p_{\langle k \rangle}^t$$

But $(\langle k \rangle, P, M_P|_{\langle k \rangle}, \phi|_{\langle k \rangle}) = (G, P, M_P, \phi)|_{\langle k \rangle} = D^1|_{\langle k \rangle}$. So,

$$p_{\langle k \rangle}^{t+1} = \mathcal{G}_{(D^1|_{\langle k \rangle}, f)} p_{\langle k \rangle}^t \tag{7}$$

Equation (7) is our second important result. Note that for any $k \in K$ such that $p_K^t(k) > 0$, $\{p^t_{\langle k\rangle}\}_{t\geq 0}$ is the genotype distribution sequence of the EM $E_k = (D^1|_{\langle k\rangle}, f, p^0_{\langle k\rangle})$. Let us call such an EM a (β, k)-subEM of E^1, let us call the genotype distribution sequence of E^1 the evolutionary process of E^1 and for any $k \in K$, let us call the genotype distribution sequence of E_k the evolutionary thread of E_k. Then, the evolutionary process of E^1 is composed of the evolutionary threads of subEMs of E^1. For any $k \in K$ such that $p_K^t(k) > 0$, the (β, k)-subEM uses the (β, k)-subtransmap of D^1. Thus there is a correspondence between instantiated themes and evolutionary threads. Each thread can be thought to evolve the non-thematic parts of all genotypes which share some common theme. For these reasons we will call equation (7) the *theme-thread correspondence equation.*

For any theme $k \in K$ such that $p_K^t(k) > 0$, and any timestep $t \in \mathbb{Z}_0^+$, let us associate the theme probability mass of k at timestep t, given by $p_K^t(k)$, with the evolutionary thread of E_k at timestep t. Then, the only way that the evolutionary threads interact is by the transfer of theme probability mass between threads in each generation as described by the theme meta-selection equation (6). Since each evolutionary thread uses a different subtransmap, we will refer to the equations (6) and (7) together as the *subtransmap-selecting multithreaded evolution equations.*

6.2 Analysis for $\omega \approx 1$

In this subsection we will show that the subtransmap-selecting multithreaded evolution equations hold approximately in each generation of E^ω when $\omega \approx 1$. In the appendix we define a metric called the *distribution distance* that measures distances between probability distributions. Using this metric, we prove a theorem that shows that for all $t \in \mathbb{Z}_0^+$, ${}^{\omega}p^t_K$ varies continuously with ω, and for any $\omega^* \in [0,1]$ and for all $k \in K$ such that ${}^{\omega^*}p^t_K(k) > 0$, ${}^{\omega}p^t_{\langle k\rangle}$ is a continuous function of ω at ω^*. Specifically, the distribution distance metric is defined as:

Definition 6.4. (Distribution Distance) *Let X be some countable set. We define the* distribution distance *$d : \Lambda^X \times \Lambda^X \to \mathbb{R}_0^+$ as follows:*

$$d(p, q) = \sum_{x\in X} |p(x) - q(x)|$$

and the continuity theorem that we prove is as follows:

Theorem 6.1. *For all $\omega \in [0, 1]$, let $D^\omega = (G, P, \phi, M)$ be a transmap such that for some set K and some $\beta : G \to K$, D^ω is (β, ω)-preserving. Let $E^\omega = (D^\omega, f, s)$ be a (β, ω)-preserving EM, and let $\{{}^{\omega}p^t_G\}_{t\geq 0}$ be the genotype distribution sequence of E^ω. For all $t \in \mathbb{Z}_0^+$, let ${}^{\omega}p^t_K$ be the β-projection of ${}^{\omega}p^t_G$, then $\forall\omega^* \in [0, 1]$ and for any $\epsilon > 0$,*

$$\exists\delta > 0,\ \forall\omega \in [0, 1], |\omega - \omega^*| < \delta \Rightarrow d({}^{\omega}p^t_K, {}^{\omega^*}p^t_K) < \epsilon$$
$$i.e. \quad \omega \to \omega^* \Rightarrow {}^{\omega}p^t_K \to {}^{\omega^*}p^t_K$$

Furthermore, for any $\omega^* \in [0, 1]$, and for all $k \in K$ such that ${}^{\omega^*}p^t_K(k) > 0$, let ${}^{\omega}p^t_{\langle k\rangle}$ be the β-conditional distribution of ${}^{\omega}p^t_G$ given k, then for and any $\epsilon > 0$,

$$\exists\delta > 0,\ \forall\omega \in [0, 1], |\omega - \omega^*| < \delta \Rightarrow d({}^{\omega}p^t_{\langle k\rangle}, {}^{\omega^*}p^t_{\langle k\rangle}) < \epsilon$$
$$i.e. \quad \omega \to \omega^* \Rightarrow {}^{\omega}p^t_{\langle k\rangle} \to {}^{\omega^*}p^t_{\langle k\rangle}$$

Fix some $\omega \in [0, 1]$ and some $t \in \mathbb{Z}_0^+$. For all $\eta \in [0, 1]$, let $Q^\eta = (D^\eta, f, {}^{\omega}p^t_G)$, and let $\{{}^{\eta}q^\tau_G\}_{\tau\geq 0}$ be the genotype distribution sequence of Q^η. For all $\tau \in \mathbb{Z}_0^+$, let ${}^{\eta}q^\tau_K$ be the β-projection of ${}^{\eta}q^\tau_G$, and for all $k \in K$ such that ${}^{\eta}q^\tau_K(k) > 0$, let ${}^{\eta}q^\tau_{\langle k\rangle}$ be the β-conditional distribution of ${}^{\eta}q^\tau_G$ given k. Note that $\forall\tau \in \mathbb{Z}_0^+$, ${}^{\omega}p^{t+\tau}_G = {}^{\omega}q^\tau_G$. So,

$$ {}^{\omega}p^{t+\tau}_K = {}^{\omega}q^\tau_K, \text{ and } \forall k \text{ s.t. } {}^{\eta}q^\tau_K(k) > 0,\ {}^{\omega}p^{t+\tau}_{\langle k\rangle} = {}^{\omega}q^\tau_{\langle k\rangle} \quad (8)$$

Note that for all $\eta \in [0, 1]$

$$ {}^{\beta}\mathfrak{F}^t_{E^\omega} = {}^{\beta}\mathfrak{F}^0_{Q^\eta},\ {}^{\omega}p^t_K = {}^{\eta}q^0_K, \text{ and}$$
$$\forall k \text{ s.t. } {}^{\eta}q^0_K(k) > 0,\ {}^{\omega}p^t_{\langle k\rangle} = {}^{\eta}q^0_{\langle k\rangle} \quad (9)$$

Using the result we obtained in (6),

$$ {}^{1}q^1_K = \mathcal{S}_{{}^{\beta}\mathfrak{F}^0_{Q^1}}({}^{1}q^0_K)$$

As $\omega \to 1$, using Theorem 1 we get that

$$ {}^{\omega}q^1_K \to \mathcal{S}_{{}^{\beta}\mathfrak{F}^0_{Q^1}}({}^{1}q^0_K)$$

Now, by (9) ${}^{\beta}\mathfrak{F}^0_{Q^1} = {}^{\beta}\mathfrak{F}^t_{E^\omega}$, and ${}^{1}q^0_K = {}^{\omega}p^t_K$, and by (8) ${}^{\omega}q^1_K = {}^{\omega}p^{t+1}_K$. So, making these three substitutions in the equation above, we get that for any $t \in \mathbb{Z}_0^+$, as $\omega \to 1$,

$$ {}^{\omega}p^{t+1}_K \to \mathcal{S}_{{}^{\beta}\mathfrak{F}^t_{E^\omega}}({}^{\omega}p^t_K) \quad (10)$$

Using the result we obtained in (7), for all $k \in K$,

$$ {}^{1}q^1_{\langle k\rangle} = \mathcal{G}_{(D^1|_{\langle k\rangle}, f)}({}^{1}q^0_{\langle k\rangle})$$

By an argument similar to the one used in the derivation of (10) we get that for all $t \in \mathbb{Z}_0^+$ and for all $k \in K$ such that ${}^{\omega}p^t_K(k) > 0$, as $\omega \to 1$,

$$ {}^{\omega}p^{t+1}_{\langle k\rangle} \to \mathcal{G}_{(D^1|_{\langle k\rangle}, f)}({}^{\omega}p^t_{\langle k\rangle}) \quad (11)$$

The results that we have proved in this and the previous subsection are stated succinctly below in the central theorem of this paper.

Theorem 6.2 (Theme Preservation Theorem). *For all $\omega \in [0, 1]$, let $D^\omega = (G, P, \phi, M)$ be a transmap such that for some set K and some $\beta : G \to K$, D^ω is (β, ω)-preserving. Let $E^\omega = (D^\omega, f, s)$ be a (β, ω)-preserving EM, and let $\{{}^{\omega}p^t_G\}_{t\geq 0}$ be the genotype distribution sequence of E^ω. For all $t \in \mathbb{Z}_0^+$, let ${}^{\omega}p^t_K$ be the β-projection of ${}^{\omega}p^t_G$. For all $k \in K$ such that ${}^{\omega}p^t_K(k) > 0$, let ${}^{\omega}p^t_{\langle k\rangle}$ be the β-conditional distribution of ${}^{\omega}p^t_G$ given k. Then, for all $t \in \mathbb{Z}_0^+$, as $\omega \to 1$,*

$$ {}^{\omega}p^{t+1}_K \to \mathcal{S}_{{}^{\beta}\mathfrak{F}^t_{E^\omega}}({}^{\omega}p^t_K)$$

and for all $k \in K$ such that ${}^{\omega}p^t_K(k) > 0$, as $\omega \to 1$,

$$ {}^{\omega}p^{t+1}_{\langle k\rangle} \to \mathcal{G}_{(D^1|_{\langle k\rangle}, f)}({}^{\omega}p^t_{\langle k\rangle})$$

In both cases, equality follows when $\omega = 1$.

A corollary of the theme preservation theorem is that the subtransmap-selecting multithreaded evolution equations hold approximately in each generation of E^ω when $\omega \approx 1$.

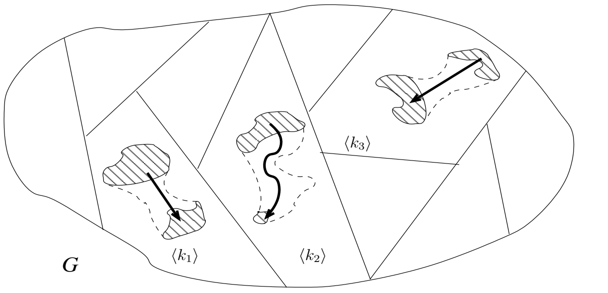


Figure 1: Schematic depiction of the behavior of a $(\beta, 1)$-preserving bfpEA

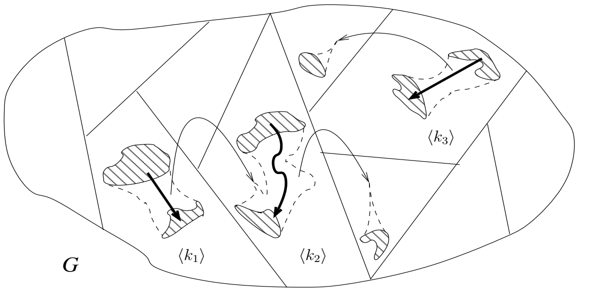


Figure 2: Schematic depiction of the behavior of a (β, ω)-preserving bfpEA when $\omega \approx 1$

7. DISCUSSION

In the previous section we analyzed the behavior of a (β, ω)-preserving EM which is a model for a (β, ω)-preserving bfpEA. In this section we interpret these theoretical results to infer qualitative and quantitative aspects about the behavior of a (β, ω)-preserving bfpEA. Let B^ω be a (β, ω)-preserving bfpEA and let E^ω be a (β, ω)-preserving EM that models B^ω. We will first focus on the behavior of B^ω when $\omega = 1$. This behavior is shown schematically in (fig. 1). The function β partitions the genotype set into theme classes. A shaded region within some theme class $\langle k \rangle$ depicts a subpopulation of k-themed genotypes at some time t [3] The initial population of B^1 is comprised of the union of subpopulations in different theme classes. Some themes may not be present at all in the initial population of B^1. This is depicted by the empty theme classes in the figure.

The sizes of the subpopulations of B^1 vary from generation to generation as evolution proceeds. This is depicted as a change in the sizes of the shaded regions depicting subpopulations. Let n be the fixed total population size of B^1. For any $t \in \mathbb{Z}_0^+$, let F_k^t be the total fitness of all the individuals in some k-themed subpopulation at time t. Then by the theme-meta selection equation the size of the k-themed sub-population in generation $t+1$, denoted by n_k^{t+1}, is given by $n_k^{t+1} \approx n \frac{F_k^t}{\sum_{k' \in K} F_{k'}^t}$.

From the theme-thread correspondence equation we can infer that apart from the change in the size of the subpopulations in each generation (as described above), the evolution of any k-themed subpopulation proceeds independently of the evolution of other subpopulations. If R is the representation that is used by B^1, then we can think of each k-themed subpopulation as evolving within a separate variable-population-size bfpEA that uses a (β, k)-subrepresentation of R; let us call this a (β, k)-sub-bfpEA of B^1.

For any theme k and any generation t, the value F_k^t can be thought of as the fitness of the (β, k)-sub-bfpEA in generation t. In each generation $t \in \mathbb{Z}_0^+$, the population size of the (β, k)-sub-bfpEA is given by n_k^t. If one thinks of the population size of a sub-bfpEA in some generation as the amount of search resources allocated to its representation in that generation, then in each generation B^1 reallocates its search resources amongst the subrepresentations of its sub-bfpEAs in proportion to the fitness of the populations of the sub-bfpEAs. We call this behavior *Subrepresentation selecting multithreaded evolution* (SSME).

Note that when $\omega = 1$, if some theme k' is not instantiated by some genotype of the initial generation of B^ω, then k' will never be present in any of the genotypes in any subsequent generations. Thus the (β, k')-subrepresentation of D^ω will never be explored. So, in each generation, B^ω will only perform subrepresentation selection on the subrepresentations that are already present in the initial generation.

When $\omega \approx 1$ we say that the bfpEA B^ω is *mostly-theme-preserving*. By the corollary to theorem 6.2, the subtransmap selecting multithreaded evolution equations approximately hold for E^ω. Therefore, we infer that B^ω *approximately* implements subrepresentation selecting multithreaded evolution. The key qualitative difference between the SSME behavior of B^1 and the behavior of B^ω is the result of the small number of theme altering mutations in each generation of B^ω. Figuratively speaking, the child genotypes produced by such mutation jump over the theme class partitions as shown in (fig. 2) and land in theme classes that are different from those of their parents. Thus in each generation of B^ω new themes may be instantiated which were not present in previous generations. When $\omega \approx 1$ themes correspond approximately to the subrepresentations used by sub-bfpEAs, so another way of saying the above is that in each generation sub-bfpEAs with new subrepresentations may be generated by B^ω. As B^ω approximately implements SSME, these new subrepresentations will be subject to subrepresentation selection in subsequent generations. Since B^ω generates new subrepresentations in addition to approximately performing SSME, we call its behavior *Subrepresentation evolving multithreaded evolution* (SEME).

8. RELATED WORK

Our study of (β, ω)-preserving bfpEAs was inspired by a desire to understand the significance of the different frequencies with which theme altering and theme preserving mutation is applied to genotypes in the self adapting evolutionary algorithms described in [7] and [8]. The genotypes in these algorithms are seeded L-systems like those described in example 3 of section 3.1. These genotypes are subject to two kinds of mutation - type 1 and type 2. Type 1 mutation preserves genotypic themes that are similar to those shown in the last column and last row of table 1; type 2

[3]While interpreting this figure and (fig. 2) the reader should bear in mind that while subpopulations are depicted as sets in truth they are multi-sets. We nevertheless use the size of a shaded area to depict the size of some sub-population some theme class.

mutation alters these themes. In the experimental systems in [7] and [8], parent genotypes are subjected to type 1 mutation much more frequently than type 2 mutation. These experimental systems are not quite bfpEAs because they use (μ, λ)-selection [5]. Nevertheless, the work in this paper explains the role that the low probability of theme alteration plays in the behavior of these systems, namely that the low probability of theme altering mutation induces evolutionary behavior that is qualitatively similar to SEME.

The experimental systems in [7] and [8] were constructed in order to study the effects of neutral mutations. In addition to being theme altering, type 2 mutations are also neutral (with high probability), i.e. they preserve phenotypes. It is instructive to understand the effect of neutral theme altering mutation using the theory developed in this paper. When theme altering mutation is *not* neutral it amounts to a "blind" jump" from a parent genotype in some subrepresentation to a child genotype in a different subrepresentation. When it *is* neutral the jump is not blind i.e. the phenotype of the parent genotype is preserved.

It is also instructive to compare and contrast the SEME behavior of a mostly-theme-preserving bfpEA with the behavior of a mutation-only island model . An island model (briefly described in [11]) maintains several subpopulations with fixed and typically equal sizes. These subpopulations evolve mostly independently of each other; they are said to be located on different islands to evoke the notion of separation. Periodically, a fraction of the individuals on each island will migrate to other islands. Mostly-theme-preserving bfpEAs and island models are analogous in that the themes of the bfpEA correspond with the islands in an island model, theme preserving mutation corresponds to regular mutation on each of the islands, and theme altering mutation corresponds to migrations between the islands. The differences between a mostly-theme-preserving bfpEA and an island model are as follows: firstly, for any theme k, the subpopulation of k-themed genotypes is variable, and is determined in each generation by theme meta-selection, whereas the subpopulation size of each island is typically fixed. Secondly, the evolutionary thread corresponding to each theme uses a *different* representation, whereas (with the exception of the island models studied in [6]) the representation used by each island is the same. And thirdly, a mostly-theme-preserving bfpEA may discover new themes which were not previously instantiated in the population and may thus spawn new evolutionary threads, whereas the number of islands in an island model is fixed.

9. CONCLUSION

In this paper, we introduced the abstract yet simple concept of a genotypic theme and distinguished it from the concepts of a forma [4], and a scheme [3]. We used this idea of a theme, and the ideas of theme preserving and theme altering mutation to define a (β, ω)-preserving EM which preserves genotypic β-themes with probability ω and alters them with probability $(1 - \omega)$. Our analysis of a (β, ω)-preserving EM in section 6 yielded the Theme Preservation Theorem which is the central theoretical result of this paper. We interpreted this result in section 7 to gain insight into the behaviour of a (β, ω)-preserving bfpEA when $\omega \approx 1$

The fascinating conclusions that we reached about a (β, ω)-preserving bfpEA for the case when $\omega \approx 1$ are as follows: Firstly, genotypic themes correspond to separate evolutionary threads (with distinct subrepresentations) in which the non-thematic parts of genotypes evolve. Secondly, in each generation the (β, ω)-preserving bfpEA performs selection of subrepresentation by reallocating its search resources to the subrepresentations of evolutionary threads with fitter populations. And thirdly, in each generation new evolutionary threads with new subrepresentations may be spawned. These new subrepresentations are then subject to subrepresentation selection in subsequent generations. Thus a (β, ω)-preserving bfpEA performs second-order search over the a space of subrepresentations and in each generation exploits promising subrepresentations for first-order evolutionary search.

10. ACKNOWLEDGEMENTS

We thank the members of the DEMO lab, especially Anthony Bucci and Shivakumar Vishwanathan, for very valuable feedback on the content and form of this paper. We also thank the reviewers for their helpful comments

11. REFERENCES

[1] Altenberg, L. The evolution of evolvability in genetic programming. In *Advances in Genetic Programming*, K. E. Kinnear, Jr., Ed. MIT Press, 1994, ch. 3, pp. 47–74.

[2] Halder, G., Callaerts, P., and Gehring, W. Induction of ectopic eyes by targeted expression of the eyeless gene in drosophilia. *Science 267* (1995), 1788–1792.

[3] Holland, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan, 1975.

[4] Radcliffe, N. J. The algebra of genetic algorithms. *Annals of Mathematics and Artificial Intelligence 10* (1994), 339–384.

[5] Schwefel, H.-P. *Evolution and Optimum Seeking*. Sixth-Generation Computer Technology Series. John Wiley and Sons, New York NY, 1995.

[6] Skolicki, Z., and De Jong, K. Improving evolutionary algorithms with multi-representation island models. In *Late Breaking Papers at the 2004 Genetic and Evolutionary Computation Conference* (Seattle, Washington, USA, 26 July 2004), M. Keijzer, Ed.

[7] Toussaint, M. Demonstrating the evolution of complex genetic representations: An evolution of artificial plants. In *Genetic and Evolutionary Computation – GECCO-2003* (Berlin, 2003), E. Cantú-Paz, J. A. Foster, K. Deb, D. Davis, R. Roy, U.-M. O'Reilly, H.-G. Beyer, R. Standish, G. Kendall, S. Wilson, M. Harman, J. Wegener, D. Dasgupta, M. A. Potter, A. C. Schultz, K. Dowsland, N. Jonoska, and J. Miller, Eds., Springer-Verlag, pp. 86–97.

[8] Toussaint, M. *The Evolution of Genetic Representations and Modular Neural Adaptation*. PhD thesis, Institut fr Neuroinformatik, Ruhr-Universit-Bochum, Germany, 2003.

[9] Vose, M. D. *The simple genetic algorithm: foundations and theory*. MIT Press, Cambridge, MA, 1999.

[10] Wagner, G. P., and Altenberg, L. Complex adaptations and the evolution of evolvability. *Evolution* (1996), In press.

[11] Whitley, D. A genetic algorithm tutorial. *Statistics and Computing 4* (1994), 65–85.

APPENDIX

Definition 1. (Distribution Distance) Let X be some countable set. We define the *distribution distance* $d : \Lambda^X \times \Lambda^X \rightarrow \mathbb{R}_0^+$ as follows:

$$d(p, q) = \sum_{x \in X} |p(x) - q(x)|$$

Proposition 1. *Let X be some countable set. Then the distribution distance $d : \Lambda^X \times \Lambda^X \rightarrow \mathbb{R}_0^+$ is always bounded from above by 2, i.e. for any $p, q \in \Lambda^X$, $d(p, q) \leq 2$*

Proof. See that

$$\begin{aligned}
d(p, q) &= \sum_{x \in X} |p(x) - q(x)| \\
&= \left(\sum_{\substack{\{x \in X | p(x) \\ > q(x)\}}} p(x) - q(x) \right) + \left(\sum_{\substack{\{x \in X | q(x) \\ > p(x)\}}} q(x) - p(x) \right) \\
&\leq \left(\sum_{\substack{\{x \in X | p(x) \\ > q(x)\}}} p(x) - 0 \right) + \left(\sum_{\substack{\{x \in X | q(x) \\ > p(x)\}}} q(x) - 0 \right) \\
&\leq \left(\sum_{x \in X} p(x) \right) + \left(\sum_{x \in X} q(x) \right) \\
&= 2
\end{aligned}$$

Observe that the summands in (def 1) are non-negative, and from the proposition above the summation in (def 1) is bounded. Hence for any set X, d is always convergent. The next proposition states that d is a metric.

Proposition 2. *For any countable set X, the distribution distance d is a metric over the set Λ^X.*

Proof. For any $p, q, r \in \Lambda^X$, it is easily seen that $d(p, q) \geq 0$, $d(p, p) = 0$, $d(p, q) = 0 \Rightarrow p = q$, and $d(p, q) = d(q, p)$. We now prove the triangle inequality. See that $\forall x \in X$, $|p(x) - q(x)| + |q(x) - r(x)| \geq |p(x) - r(x)|$ because the absolute value of differences is simply the 1-d Euclidean metric. Hence $\sum_{x \in X} |p(x) - q(x)| + |q(x) - r(x)| \geq \sum_{x \in X} |p(x) - r(x)|$, so $(\sum_{x \in X} |p(x) - q(x)|) + (\sum_{x \in X} |q(x) - r(x)|) \geq \sum_{x \in X} |p(x) - r(x)|$, which gives us our result.

Lemma 1. *For any $x^* \in [0, 1]$, let $f : [0, 1] \rightarrow [0, 1]$ be a function such that f is continuous at x^* and $f(x^*) > 0$. Then $\forall \epsilon > 0$, $\exists \delta > 0$ such that*

$$|x - x^*| < \delta \Rightarrow |f(x) - f(x^*)| < \epsilon f(x^*)$$

Proof. Note that $\forall x^* \in [0, 1]$ s.t. $f(x^*) > 0$, and $\forall \epsilon > 0$, $|f(x) - f(x^*)| < \epsilon f(x^*)$ when $x = x^*$. Also see that $f(x^*)$ is constant w.r.t. x. As f is continuous at x^*, $f(x) - f(x^*)$ is continuous at x^*, which implies that $|f(x) - f(x^*)|$ is continuous at x^*. Hence $\exists \delta > 0$ s.t. $|x - x^*| < \delta \Rightarrow |f(x) - f(x^*)| < \epsilon . f(x^*)$.

Lemma 2. *For any $\omega \in [0, 1]$, any countable set X and any parameterized distribution $p_\omega : \Lambda^X \rightarrow \Lambda^X$, let us define the function $q_x : [0, 1] \rightarrow [0, 1]$ as follows: $\forall \omega \in [0, 1], \forall x \in X \quad q_x(\omega) = p_\omega(x)$. Let $\omega^* \in [0, 1]$ be such that $\forall x \in X$, q_x is continuous at ω^*, then $\forall \epsilon > 0, \exists \delta$ such that $\forall \omega \in [0, 1]$*

$$|\omega - \omega^*| < \delta \Rightarrow d(p_\omega, p_{\omega^*}) < \epsilon$$

Proof.

$$\begin{aligned}
& \sum_{x \in X} p_{\omega^*}(x) - p_\omega(x) = 0 \\
& \Rightarrow \sum_{\substack{x \in X \text{ s.t.} \\ p_{\omega^*}(x) > \\ p_\omega(x)}} p_{\omega^*}(x) - p_\omega(x) - \sum_{\substack{x \in X \text{ s.t.} \\ p_\omega(x) > \\ p_{\omega^*}(x)}} p_\omega(x) - p_{\omega^*}(x) = 0 \\
& \Rightarrow \sum_{\substack{x \in X \text{ s.t.} \\ p_{\omega^*}(x) > \\ p_\omega(x)}} p_{\omega^*}(x) - p_\omega(x) = \sum_{\substack{x \in X \text{ s.t.} \\ p_\omega(x) > \\ p_{\omega^*}(x)}} p_\omega(x) - p_{\omega^*}(x)
\end{aligned} \tag{12}$$

Now, since

$$\sum_{\substack{x \in X \text{ s.t.} \\ p_{\omega^*}(x) > \\ p_\omega(x)}} p_{\omega^*}(x) - p_\omega(x) = \sum_{\substack{x \in X \text{ s.t.} \\ p_{\omega^*}(x) > \\ p_\omega(x)}} |p_{\omega^*}(x) - p_\omega(x)| \tag{13}$$

and

$$\sum_{\substack{x \in X \text{ s.t.} \\ p_\omega(x) > \\ p_{\omega^*}(x)}} p_\omega(x) - p_{\omega^*}(x) = \sum_{\substack{x \in X \text{ s.t.} \\ p_\omega(x) > \\ p_{\omega^*}(x)}} |p_\omega(x) - p_{\omega^*}(x)|$$

therefore,

$$\sum_{\substack{x \in X \text{ s.t.} \\ p_{\omega^*}(x) > \\ p_\omega(x)}} |p_{\omega^*}(x) - p_\omega(x)| = \sum_{\substack{x \in X \text{ s.t.} \\ p_\omega(x) > \\ p_{\omega^*}(x)}} |p_\omega(x) - p_{\omega^*}(x)| \tag{14}$$

By lemma 1, $\forall \epsilon \in [0, 1]$, $\exists \delta > 0$ s.t. $\forall \omega$ s.t. $|\omega - \omega^*| < \delta$,

$$\begin{aligned}
& \sum_{\substack{x \in X \text{ s.t.} \\ p_{\omega^*}(x) > \\ p_\omega(x)}} |q_x(\omega^*) - q_x(\omega)| < \sum_{\substack{x \in X \text{ s.t.} \\ p_{\omega^*}(x) > \\ p_\omega(x)}} \frac{\epsilon}{2} . q_x(\omega^*) \\
& \Rightarrow \sum_{\substack{x \in X \text{ s.t.} \\ p_\omega^*(x) > \\ p_\omega(x)}} |q_x(\omega^*) - q_x(\omega)| < \frac{\epsilon}{2} \\
& \Rightarrow \sum_{\substack{x \in X \text{ s.t.} \\ p_{\omega^*}(x) > \\ p_\omega(x)}} |p_{\omega^*}(x) - p_\omega(x)| < \frac{\epsilon}{2}
\end{aligned} \tag{15}$$

and by (14) and (15) $\forall \omega$ s.t. $|\omega - \omega^*| < \delta$,

$$\sum_{\substack{x \in X \text{ s.t.} \\ p_\omega(x) > \\ p_{\omega^*}(x)}} |p_\omega(x) - p_{\omega^*}(x)| < \frac{\epsilon}{2}$$

so, $\forall \omega$ s.t. $|\omega - \omega^*| < \delta$

$$\sum_{x \in X} |p_{\omega^*}(x) - p_\omega(x)| < \epsilon$$

THEOREM 1. *For all $\omega \in [0,1]$, let $D^\omega = (G, P, \phi, M)$ be a transmap such that for some set K and some $\beta : G \to K$, D^ω is (β,ω)-preserving. Let $E^\omega = (D^\omega, f, s)$ be a (β,ω)-preserving EM, and let $\{{}^{\omega}p^t_G\}_{t\geq 0}$ be the genotype distribution sequence of E^ω. For all $t \in \mathbb{Z}^+_0$, let ${}^{\omega}p^t_K$ be the β-projection of ${}^{\omega}p^t_G$, then $\forall \omega^* \in [0,1]$ and for any $\epsilon > 0$,*

$$\exists \delta > 0,\ \forall \omega \in [0,1], |\omega - \omega^*| < \delta \Rightarrow d({}^{\omega}p^t_K,\ {}^{\omega^*}p^t_K) < \epsilon$$
$$\textit{i.e.} \quad \omega \to \omega^* \Rightarrow {}^{\omega}p^t_K \to {}^{\omega^*}p^t_K$$

Furthermore, for any $\omega^* \in [0,1]$, and for all $k \in K$ such that ${}^{\omega^*}p^t_K(k) > 0$, let ${}^{\omega}p^t_{\langle k\rangle}$ be the β-conditional distribution of ${}^{\omega}p^t_G$ given k, then for and any $\epsilon > 0$,

$$\exists \delta > 0,\ \forall \omega \in [0,1], |\omega - \omega^*| < \delta \Rightarrow d({}^{\omega}p^t_{\langle k\rangle},\ {}^{\omega^*}p^t_{\langle k\rangle}) < \epsilon$$
$$\textit{i.e.} \quad \omega \to \omega^* \Rightarrow {}^{\omega}p^t_{\langle k\rangle} \to {}^{\omega^*}p^t_{\langle k\rangle}$$

CLAIM 1. *For all $t \in \mathbb{Z}^+_0$ and all $g \in G$, let ${}^{g}q^t_G : [0,1] \to [0,1]$ be defined as follows: ${}^{g}q^t_G(\omega) = {}^{\omega}p^t_G(g)$. Then ${}^{g}q^t_G$ is continuous.*

We prove claim 1 by induction on t. For all $g \in G$, ${}^{g}q^0_G(\omega)$ is a constant function of ω, hence ${}^{g}q^0_G$ is continuous. This proves our base case. For some $t \in \mathbb{Z}^+_0$, we now assume that for all $g \in G$, ${}^{g}q^t_G$ is continuous and prove that for all $g \in G$, ${}^{g}q^{t+1}_G$ is continuous. $\forall g \in G$,

$$\begin{aligned}{}^{\omega}p^{t+1}_G(g) &= \sum_{g' \in G} (\omega M_P(g|g') + (1-\omega) M_A(g|g')) \quad \times \\ &\qquad \mathcal{S}_{f\circ\phi}\, {}^{\omega}p^t_G(g') \\ &= \sum_{g' \in G} (\omega M_P(g|g') + (1-\omega) M_A(g|g')) \quad \times \\ &\qquad \frac{f\circ\phi(g')\, {}^{\omega}p^t(g')}{\sum_{g'' \in G} f\circ\phi(g'')\, {}^{\omega}p^t(g'')}\end{aligned}$$

So,

$$\begin{aligned}{}^{g}q^{t+1}_G(\omega) &= \sum_{g' \in G} (\omega M_P(g|g') + (1-\omega) M_A(g|g')) \quad \times \\ &\qquad \frac{f\circ\phi(g')\, {}^{g'}q^t_G(\omega)}{\sum_{g'' \in G} f\circ\phi(g'')\, {}^{g'}q^t_G(\omega)}\end{aligned}$$

Note that the right side of this equation is the sum, product and non-zero quotient of constant or continuous functions w.r.t ω (we are using our inductive hypothesis in this statement). Hence, ${}^{g}q^{t+1}_G$ is continuous. This completes the proof of claim 1.

CLAIM 2. *For all $t \in \mathbb{Z}^+_0$, all $k \in K$, let ${}^{k}q^t_K : [0,1] \to [0,1]$ be defined as follows: ${}^{k}q^t_K(\omega) = {}^{\omega}p^t_K(k)$. Then, ${}^{k}q^t_K$ is continuous*

The proof of claim 2 is as follows: For any $t \in \mathbb{Z}^+_0$ and any $k \in K$,

$$\begin{aligned}{}^{k}q^t_K(\omega) &= \sum_{g \in \langle k\rangle} {}^{\omega}p^t_G(g) \\ &= \sum_{g \in \langle k\rangle} {}^{g}q^t_G(\omega)\end{aligned}$$

By claim 1, the right hand side of this equation is the sum of continuous functions w.r.t ω. Hence ${}^{k}q^t_K$ is continuous. This completes the proof of claim 2. By claim 2 and lemma 2, for all $t \in \mathbb{Z}^+_0$, any $\omega^* \in [0,1]$ and for any $\epsilon > 0$,

$$\exists \delta > 0 \text{ s.t. } \forall \omega \in [0,1], \quad |\omega - \omega^*| < \delta \Rightarrow d({}^{\omega}p^t_K,\ {}^{\omega^*}p^t_K) < \epsilon$$
$$\text{i.e.} \quad \omega \to \omega^* \Rightarrow {}^{\omega}p^t_K \to {}^{\omega^*}p^t_K$$

CLAIM 3. *For all $t \in \mathbb{Z}^+_0$, all $\omega^* \in [0,1]$, all $k \in K$ such that ${}^{\omega^*}p^t(k) > 0$, and all $g \in \langle k\rangle$, let ${}^{g}q^t_{\langle k\rangle} : [0,1] \to [0,1]$ be defined as follows ${}^{g}q^t_{\langle k\rangle}(\omega) = {}^{\omega}p^t_{\langle k\rangle}(g)$. Then ${}^{g}q^t_{\langle k\rangle}$ is continuous at ω^*.*

The proof of claim 3 is as follows: for all $t \in \mathbb{Z}^+_0$, all $\omega^* \in [0,1]$ and all $k \in K$ such that ${}^{\omega^*}p^t(k) > 0$, ${}^{k}q^t_K(\omega^*) = {}^{\omega^*}p^t(k)$ so, ${}^{k}q^t_K(\omega^*) > 0$. Furthermore, for all $g \in \langle k\rangle$

$$\begin{aligned}{}^{g}q^t_{\langle k\rangle}(\omega^*) &= \frac{{}^{\omega^*}p^t_G(g)}{{}^{\omega^*}p^t_K(k)} \\ &= \frac{{}^{g}q^t_G(\omega^*)}{{}^{k}q^t_K(\omega^*)}\end{aligned}$$

By claim 1, and claim 2, the right hand side of this equation is the division of two functions that are continuous at ω^*. Also, by our earlier comment, the denominator is non-zero. Hence, ${}^{g}q^t_{\langle k\rangle}$ is continuous at ω^*. This completes the proof of claim 3. By claim 3 and lemma 2, for any $\omega^* \in [0,1]$, for all $k \in K$ such that ${}^{\omega^*}p^t_K(k) > 0$, and for and any $\epsilon > 0$,

$$\exists \delta > 0,\ \forall \omega \in [0,1], |\omega - \omega^*| < \delta \Rightarrow d({}^{\omega}p^t_{\langle k\rangle},\ {}^{\omega^*}p^t_{\langle k\rangle}) < \epsilon$$
$$\text{i.e.} \quad \omega \to \omega^* \Rightarrow {}^{\omega}p^t_{\langle k\rangle} \to {}^{\omega^*}p^t_{\langle k\rangle}$$