# Evolution with culture[†]

**María-Elena Solano**

**Abstract**  I study a simple variant of Holland's Genetic Algorithm (GA) equipped with a fixed-rate 'Boltzmann' selection rule $p_{\text{select}} \propto e^{-\eta(1-\text{Fitness})}$, meant to capture the 'viral' spread of adaptive innovations characteristic of cultural (*aka* memetic) evolution. Using the recent groundbreaking reduction by Chastain *et al.* (2014) and Meir & Parkes (2015) of evolution to coordination game dynamics with multiplicative weights updates, I show that, for a fixed (optimal) value of $\eta$, the time-averaged expected fitness of populations evolved under this variant follows $O(1/\sqrt{T})$ *no-regret* dynamics —whereas GA 'merely' achieves $O(1/T)$ *diminishing regret* dynamics, ultimately bounded by the selection strength of the fitness landscape. Experiments on a challenging *Ms. Pac-Man* grammatical evolution (GE) task, however, provide mixed evidence for these theoretical results.

## 1  Introduction

What makes *cultural* (*aka* memetic) evolution different? I claim the answer is '*virality*': unlike genes, highly adaptive memes quickly spread person-to-person throughout society in a contagion-like manner. As a result of this 'accelerated' spread, I argue, highly adaptive memeplexes 'survive' the exponential halving induced by recombination. As Dawkins put it (1976, §11):

> [I]f you contribute to the world's culture, if you have a good idea, compose a tune, invent a sparking plug, write a poem, it may live on, intact, long after your genes have dissolved in the common pool.

To model this claim quantitatively, consider the following simple variant of Holland's Genetic Algorithm (GA) (1975): instead of selecting chromosomes $\mathbf{c} \in \mathcal{C}$ using a fitness-proportional rule

$$p_{\text{select}} \propto \text{Fitness}(\mathbf{c}), \tag{1}$$

select according to the exponentially-weighted (*aka* 'Boltzmann') rule

$$p_{\text{select}} \propto e^{-\eta(1-\text{Fitness}(\mathbf{c}))}, \tag{2}$$

for some fixed $\eta$. Henceforth, I use the term 'MA' to refer to this variant[1].

The main result of this paper is that, for a given choice of $\eta$, this modification alone improves the worst-case performance of evolution on 'jagged' fitness landscapes, in terms of the (time-averaged)

---

[1]Short for 'memetic algorithm'. This is not to suggest, however, that this is a *novel* variant in any way —*cf.* §1.1.

fitness *regret* after $T$ generations:

> **Theorem 1**. *Let $s = \max\{\text{Fitness}(\cdot)\} - \min\{\text{Fitness}(\cdot)\}$ be the selection strength of the fitness landscape. Then with high probability, after $T > 0$ generations, GA's time-averaged expected fitness regret is at most*
> $$s^2 + O(1/T).$$

> **Theorem 2**. *Let $\eta = 4(\ln|\mathcal{C}|/(1 - \tilde{f}_{min}))^{1/2}$ and $\tilde{f}_{min} \leq \text{Fitness}(\cdot) \in [0, 1]$. Then with high probability, after $T > 0$ generations, MA's time-averaged expected fitness regret is at most*
> $$O\left(\sqrt{(1 - \tilde{f}_{\min})/T}\right) + O(1/T).$$

In other words, whereas GA 'merely' follows $O(1/T)$ *diminishing regret* dynamics ultimately bounded by the *selection strength* of the fitness landscape $s$, MA follows $O(1/\sqrt{T})$ *no-regret* dynamics —or up to $O(1/T)$ for $\tilde{f}_{\min} \to 1$. Hence, for 'jagged' landscapes with high $s$, MA will likely outperform GA. On the other hand, on 'flat' landscapes with low $s$, GA will converge faster than MA.

Note the resemblance with the canonical regret bounds of the celebrated Multiplicative Weights (MW) class of algorithms (*cf.* Arora, Hazan & Kale 2012). Indeed, Theorems 1 and 2 derive directly —with minor differences— from the recent groundbreaking reduction by Chastain *et al.* (2014) and Meir & Parkes (2015) of evolution to a *coordination game* between genes with MW updates, plus Freund & Schapire (1997)'s regret bounds for their exponentially-weighted variant of MW (*aka* 'Hedge'). The (short) proof is presented in §3.

Also note the bound in Theorem 2 is 'opportunistic', in the sense it relies on setting $\eta$ to its (approximate) optimal value $\eta^* \approx \Theta(\ln|\mathcal{C}|)$. Nevertheless, I argue this value admits a parsimonious interpretation: namely, the rate needed to (roughly) maximize the entropy of the resulting marginal allele densities subject to the information encoded in the fitness landscape. This (heuristic) claim is made more precise in §3.5.

Another implication of the connection to MW is that these bounds are tight up to additive constants (*cf.* Arora, Hazan & Kale 2012, §4) —so they should hold well in practice. Yet as it turns out, they do not. This claim is verified empirically in §4.

Lastly, I note these regret bounds are largely *impractical* for evolutionary computing (EC) applications, where the main concern is to maximize the *best* fitness of the population, not the mean. Furthermore, they only hold with high probability due to the distorting effect of mutation. These and other open problems are discussed in §5.

## 1.1 Related work

Exponentially-weighted (*aka* 'Boltzmann') selection rules have been studied extensively in the EC literature (*cf.* Mitchell 1995, §5.4). Goldman (1990) described a novel (approximate) reduction of GA to simulated annealing by implementing a tournament scheme with exponentially-weighted selection probabilities and a time-varying rate $\eta = \eta_t$, roughly corresponding to the (inverse of the) 'temperature'

of the search. De la Maza & Tidor (1991) showed that GA with time-annealed Boltzmann selection outperforms 'canonical' GA on a set of benchmark tasks, and suggested superior entropy control as the likely source of the advantage. In the context of statistical mechanics, Prügel-Bennett & Shapiro (1994) reduced GA with fixed-rate Boltzmann selection —as in (2)— and Gaussian initial fitness distribution to Derrida's Random Energy Model (1981), a tractable class of spin-glass dynamics with stable thermal equilibrium —and hence a unique optimal choice of $\eta = \eta^*$.

The contagion-like, accelerated dynamics of cultural transmission *vis-à-vis* genetic transmission has been studied extensively in the broader sociobiology literature (*cf.* Boyd, Richerson & Henrich 2011, pp.431-432). Cavalli-Sforza & Feldman (1981, §1.9) study a simple model of frequency-proportionate cultural transmission that results in S-shaped, logistic-like trait frequencies. Boyd & Richerson (1983, §5) describe a stylized dual-inheritance model of population dynamics in which fitness-proportionate cultural transmission leads to an exponential increase in trait frequency, in direct proportion to the variance of the fitness landscape.

In a different line of research, Kanade (2010, §5.4) reduced a *parallel* variant of Feldman's *correlational statistical query* (CSQ) learning (Feldman 2008, §3) with polynomially many parallel steps to an extension of Valiant's evolvability (2009) with recombination and polylog number of generations. Interestingly, under this reduction, evolvability with recombination and *weak* selection can only simulate a somewhat weaker variant of CSQ that requires polylog overhead.

## 2 Preliminaries

Throughout this paper, I use $|\cdot|$ to denote set cardinality, and $\diamond$ to denote vector concatenation. Otherwise, I mostly follow the notation in Meir & Parkes (2015) (where applicable).

### 2.1 Setup

Consider an initial *population* $\mathcal{O}^{(t=0)}$ or finite multiset of *chromosomes* $\mathbf{c} = (a_1, \ldots, a_m) \in \mathcal{C}$ with *population densities* $p_{\mathbf{c}}^{(t=0)} \sim P^{(t=0)}$, each comprising $m$ *genes* (or *memes*) $a_i \in \mathcal{A}$ drawn from a finite set of $n$ integer-valued *alleles* $\mathcal{A} = \{1, ..., n\}$. Given a bounded *fitness function* $f : \mathcal{C} \to [0, 1]$, we want to *evolve* the population: that is, given an *evolutionary algorithm* EA, to iteratively produce a chain of populations

$$\mathcal{O}^{(0)} \xrightarrow{\text{EA}} \mathcal{O}^{(1)} \xrightarrow{\text{EA}} \ldots \xrightarrow{\text{EA}} \mathcal{O}^{(T)}$$

such that, after $T$ *generations*, the *expected fitness* of chromosomes in the last population $\mathcal{O}^{(T)}$

$$\mathbb{E}_{\mathbf{c} \sim P^{(T)}}[f(\mathbf{c})] = \sum_{\mathbf{c} \in \mathcal{C}} f(\mathbf{c})\, p_{\mathbf{c}}^{(T)} \tag{3}$$

$$= \frac{1}{|\mathcal{O}^{(T)}|} \sum_{\mathbf{c} \in \mathcal{O}^{(T)}} f(\mathbf{c}) \tag{4}$$

is as high as 'possible'. In an adversarial environment, a desirable goal would be to bound the 'instan-

taneous' fitness *regret* (*cf.* Cesa-Bianchi and Lugosi 2006, §I)

$$\mathbb{E}_{\mathbf{c}\sim P^{(T)}}[f(\mathbf{c})] - \mathbb{E}[\check{f}^{(T)}], \tag{5}$$

where $\check{f}^{(T)}$ is the highest *achievable* fitness by the EA after $T$ steps.

In this paper, however, I focus on the more opaque —and, in a sense, weaker— notion of *time-averaged* fitness regret

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\mathbf{c}\sim P^{(t)}}[f(\mathbf{c})] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\check{f}^{(T)}], \tag{6}$$

which, although much less interpretable than (5), is much more common in the online learning literature —and indeed, is the usual form of the canonical bounds I am drawing from. The extension of Theorems 1 and 2 to the much more natural notion of 'instantaneous' regret is discussed further in §6.

### 2.2 Evolutionary Algorithms

Let me formally state Holland's 'canonical' GA:

**Algorithm 1 (GA) (Holland 1975)**
*Input:* Fitness function $f : \mathcal{C} \to \mathbb{R}_+$, initial population $\mathcal{O}^{(0)}$ of chromosomes drawn from $\mathcal{C}$,
    number of generations $T$, mutation and crossover rates $p_m, p_r \in [0, 1]$.
  1. For $t = 1, \ldots, T$:
    i. Let $\mathcal{O}^{(t)}$ = new population of $|\mathcal{O}^{(t-1)}|$ chromosomes obtained by repeating steps (a)-(c):
      a. *Select* two chromosomes from $\mathcal{O}^{(t-1)}$, with probability $f(\cdot)/\sum_{\mathbf{c}\in\mathcal{O}^{(t-1)}} f(\mathbf{c})$.
      b. *Cross* the two resulting chromosomes, with probability $p_r$.
      c. *Mutate* the two resulting chromosomes, with probability $p_m$.
  2. Output $\mathcal{O}^{(T)}$.

—and its 'memetic' variant, MA:

**Algorithm 2 (MA)**
*Input:* Fitness function $f : \mathcal{C} \to [0, 1]$, initial population $\mathcal{O}^{(0)}$ of chromosomes drawn from $\mathcal{C}$,
    number of generations $T$, mutation and crossover rates $p_m, p_r \in [0, 1]$, free parameter $\eta$.
  1. For $t = 1, \ldots, T$:
    i. Let $\mathcal{O}^{(t)}$ = new population of $|\mathcal{O}^{(t-1)}|$ chromosomes obtained by repeating steps (a)-(c):
      a. *Select* two chromosomes from $\mathcal{O}^{(t-1)}$, with probability $e^{-\eta(1-f(\cdot))}/\sum_{\mathbf{c}\in\mathcal{O}^{(t-1)}} e^{-\eta(1-f(\mathbf{c}))}$.
      b. *Cross* the two resulting chromosomes, with probability $p_r$.
      c. *Mutate* the two resulting chromosomes, with probability $p_m$.
  2. Output $\mathcal{O}^{(T)}$.

Note the only difference between GA and MA —other than normalization— is in the *Select* step, parameterized by $\eta$. Also note the *Cross* step denotes *single*-point crossover, as in Holland's original GA.

Let me also state the following 'generalized' variant of GA (henceforth GA$_{\text{na}}$) with *no* mutation and *ar-*

*bitrary* homologous regions —which corresponds to the multi-gene version of the evolutionary model studied in Meir & Parkes (2015, §A):

**Algorithm 3 (GA$_{na}$)**
*Input:* Fitness function $f : \mathcal{C} \to \mathbb{R}_+$, initial population $\mathcal{O}^{(0)}$ of chromosomes drawn from $\mathcal{C}$,
    number of generations $T$, crossover rate $p_r \in [0, 1]$.
1. For $t = 1, \ldots, T$:
    i. Let $\mathcal{O}^{(t)} =$ new population of $|\mathcal{O}^{(t-1)}|$ chromosomes obtained by repeating steps (a)-(c):
    a. *Select* two chromosomes from $\mathcal{O}^{(t-1)}$, with probability $f(\cdot)/\sum_{\mathbf{c} \in \mathcal{O}^{(t-1)}} f(\mathbf{c})$.
    b. *Cross* the two chromosomes (on a random partition of $\{1, \ldots, m\}$), with probability $p_r$.
2. Output $\mathcal{O}^{(T)}$.

Note that single-point crossover corresponds to the case where the recombinable homologous regions are limited to the pairs $\{\emptyset, \{1, \ldots, m\}\}$, $\{\{1\}, \{2, \ldots, m\}\}$, $\{\{1, 2\}, \{3, \ldots, m\}\}$, and so forth.

## 2.3 Regret Bounds for Coordination Games with Multiplicative Weights Updates

Consider a repeated *coordination* game between $m$ players $l \in \{1, ..., m\}$, $n$ actions $a \in \{1, ..., n\}$, and shared payoff $\Delta(\mathbf{c}) = \Delta((a_1, \ldots, a_m)) \in [0, 1]$; where at time $t > 0$ each player $l$ *updates* their (initially uniform) mixed strategies $X_l^{(t)}$ according to the exponentially-weighted *multiplicative weights* (MW) update rule (*aka* 'Hedge') (Freund & Schapire 1996):

$$X_l^{(t)}(a) \propto X_l^{(t-1)}(a)\mathbb{E}[e^{-\eta(1-\Delta(\mathbf{c}))} \mid a_l = a] \tag{7}$$

—so that, intuitively, the *weight* of each pure strategy $a$ is updated *multiplicatively* in proportion to its payoff $\Delta(\mathbf{c})$, by an exponential amount controlled by $\eta$.

Surprisingly, using this update rule with $\eta \approx \Theta(\ln n^m)$ ensures that after $T$ rounds, the average payoff regret *vis-à-vis* that of the *best* mixed strategies $\check{X}_l$

$$\frac{1}{T}\sum_{t=1}^{T} \Delta(\mathbf{c}|a_l \sim X_l^{(t)}) - \frac{1}{T}\sum_{t=1}^{T} \Delta(\mathbf{c}|a_l \sim \check{X}_l) \tag{8}$$

converges to *zero* at a worst-case rate of roughly $O(1/\sqrt{T})$, or as high as $O(1/T)$ for $\tilde{\Delta}_{min} \to 1$, where $\tilde{\Delta}_{min} \leq \Delta(\cdot)$ is a prior lower-bound on the instantaneous shared payoff:

**Lemma 1**. *Let* $\eta = 4(\ln n^m/(1 - \tilde{\Delta}_{min}))^{1/2}$ *and* $\tilde{\Delta}_{min} \leq \Delta(\cdot) \in [0, 1]$. *Then after $T$ rounds, the average payoff regret under the Hedge update rule as described above is at most*

$$O\left(\sqrt{(1 - \tilde{\Delta}_{\min})/T}\right) + O(1/T).$$

**Proof**: Follows immediately from Lemma 4 in Freund & Schapire (1996) —with gains instead of losses— applied to Corollary 4 in Freund & Schapire (1999) (extended to the $m$ players case), and refined by the constant factor in (Stoltz & Lugosi 2005) to ensure convergence to a correlated equilibrium by also bounding 'internal regret' (*cf.* Cesa-Bianchi *et al.* 2006, §4.4).

The resemblance with the bound in Theorem 2 is not accidental: MA's population dynamics *reduce* (with high probability) to a coordination game with Hedge update dynamics, so the payoff bound in Lemma 1 also implies a bound on the time-averaged expected fitness regret of populations evolved under MA. This claim will be made precise in the next section.

## 3 Analysis

In this section I restate and sketch the main results of this paper, Theorems 1 and 2. At a high level, they follow directly from Meir & Parkes (2015) and Chastain *et al.* (2014)'s groundbreaking reduction of evolution to coordination games using MW-like update rules. Omitted steps are given in §A.

### 3.1 Meir & Parkes (2015) and Chastain et al. (2014)'s Reduction

Recently, Meir & Parkes (2015) extended Chastain *et al.* (2014)'s observation that under $GA_{na}$, the marginal allele density of locus $l$ after $T$ generations

$$\sum_{\mathbf{c} \in \mathcal{C} | a_l = a} p_{\mathbf{c}}^{(T)}, \tag{9}$$

is *precisely* the mixed strategy of player $l$ after $T > 0$ rounds of a coordination game similar to the one described in §2.3 above, except with two-sided payoffs(/losses), and with the MW-like update rule —*aka* 'polynomial weights' or 'PW' (Meir & Parkes 2015, §2.3)—

$$X_l^{(T)}(a) \propto X_l^{(T-1)}(a) \mathbb{E}[(1+\eta)\Delta(\mathbf{c}) \mid a_l = a], \tag{10}$$

with the 'learning rate' $\eta = s$ set to the *selection strength* of the fitness landscape $s = \max f(\cdot) - \min f(\cdot)$, and the payoffs/losses normalized to $\Delta = (f - 1)/s \in [-1, 1]$:

> **Lemma 2**. *After $T$ generations, the marginal allele densities under $GA_{na}$ are identical to the mixed strategies after $T$ rounds of a two-sided coordination game with payoffs/losses $\Delta = (f - 1)/s$, and PW updates with 'learning rate' $\eta = s$.*
>
> **Proof**: Follows immediately from Corollary 1 in Meir & Parkes (2015), extended to the multiple genes case *via* Proposition A.3 in (*ibid*).

Furthermore, since the PW rule in (10) *also* ensures diminishing regret dynamics (*cf.* Meir & Parkes 2015, §5.1), this reduction immediately implies a regret bound on $f$ —albeit a less advantageous one than the one in Lemma 1, due to the inability to 'tune' $s$ in advance (*cf.* Corollary 3 in Meir & Parkes 2015):

> **Lemma 3**. After $T$ generations, *$GA_{na}$'s time-averaged fitness regret is at most*
>
> $$s^2 + O(1/T).$$

**Proof**: Follows directly from Lemma 2 above and Corollary 3 in Meir & Parkes (2015).

Put in another way, *GA$_{na}$'s marginal population dynamics can be seen as a PW coordination game between genes* where the strategies are the alleles, the shared payoffs/losses are the (centered, normalized) fitnesses of the chromosomes, and the 'learning rate' is the selection strength of the fitness landscape. As a result, *PW's diminishing average payoff regret also implies diminishing time-averaged fitness regret at rate $O(1/T)$, albeit ultimately bounded by $s^2$ (cf.* Meir & Parkes 2015, §5).

In the next two subsections, I extend this reduction to GA and MA, respectively.

### 3.2 Extension to GA

Crucially, the reduction above can be extended to the case of GA with little modification.

To see why, note that under single-point crossover, the population dynamics (before mutation) are almost identical to the "SR dynamics" in Equation A.8 in Meir & Parkes (2015) —except for the inner summation term, which should now range over all pairs of homologous regions recombinable under single-point crossover (namely $\{\emptyset, \{1, \ldots, m\}\}, \{\{1\}, \{2, \ldots, m\}\}, \{\{1, 2\}, \{3, \ldots, m\}\}$, and so forth) instead:

**Lemma 4.** *For $t > 0$, the population densities (before mutation) under GA are given by*

$$\tilde{p}_{\mathbf{c}}^{(t+1)} = p_r \sum_{\mathbf{d} \in \mathcal{C}} \frac{1}{|\mathcal{R}|} \sum_{J \in \mathcal{R}} \frac{f(\mathbf{c}_{J_1} \diamond \mathbf{d}_{J_2}) p_{\mathbf{c}_{J_1} \diamond \mathbf{d}_{J_2}}^{(t)} f(\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}) p_{\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}}^{(t)}}{\left(Z_{\bar{f}}^{(t)}\right)^2} + (1 - p_r) \frac{f(\mathbf{c}) p_{\mathbf{c}}^{(t)}}{Z_{\bar{f}}^{(t)}},$$

*where $\mathcal{R} = \{\{\emptyset, \{1, \ldots, m\}\}, \{\{1\}, \{2, \ldots, m\}\}, \{\{1, 2\}, \{3, \ldots, m\}\}, \ldots\}$ is the set of all pairs of recombinable homologous regions, and $Z_{\bar{f}}^{(t)} = \sum_{\mathbf{c} \in \mathcal{C}} p_{\mathbf{c}}^{(t)} f(\mathbf{c})$ is the expected fitness of chromosome $\mathbf{c}$ at generation $t$.*

**Proof**: Summing over all possible recombinations and normalizing gives the result.

From Lemma 4, it can be shown (*cf.* §A.1) that the marginal allele densities (after mutation) factor into the same form of Lemma A.2 in Meir & Parkes (2015) —albeit only with high probability, due to the distorting effect of mutation:

**Lemma 5.** *For $T > 0$, with high probability, GA's marginal allele densities equal GA$_{na}$'s.*

**Proof**: See §A.1.

Lemma 5, in turn, readily implies Lemmas 2 and 3 above also apply to the case of GA —albeit, again, only with high probability. The result is Theorem 1 (restated here as Theorem 3):

**Theorem 3.** *With high probability, GA's time-averaged fitness regret is at most*

$$s^2 + O(1/T).$$

**Proof**: Lemma 5 above implies Lemma 2 (and hence Lemma 3) applies to the case of GA (with high probability). The result follows.

In other words, *GA's population dynamics also reduce to a PW coordination game between genes*. As a result, *GA also follows $O(1/T)$ diminishing regret dynamics, bounded by $s^2$*.

### *3.3 Extension to MA*

In the case of MA, by a similar reasoning, note that its marginal population dynamics are precisely the update dynamics of the coordination game with exponentially-weighted (*aka* Hedge) update rules described above in §2.3. To see why, note that its population dynamics obey the following lemma (*cf.* Meir & Parkes 2015, Equation A.8):

**Lemma 6**. *For $t > 0$, the population densities (before mutation) under MA are given by*

$$\tilde{p}_{\mathbf{c}}^{(t+1)} = p_r \sum_{\mathbf{d} \in \mathcal{C}} \frac{1}{|\mathcal{R}|} \sum_{J \in \mathcal{R}} \frac{e^{-\eta(1-f(\mathbf{c}_{J_1} \diamond \mathbf{d}_{J_2}))} p_{\mathbf{c}_{J_1} \diamond \mathbf{d}_{J_2}}^{(t)} e^{-\eta(1-f(\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}))} p_{\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}}^{(t)}}{\left(Z_{\exp}^{(t)}\right)^2} + (1-p_r) \frac{e^{-\eta(1-f(\mathbf{c}))} p_{\mathbf{c}}^{(t)}}{Z_{\exp}^{(t)}}$$

*where $\mathcal{R} = \{\{\emptyset, \{1, \ldots, m\}\}, \{\{1\}, \{2, \ldots, m\}\}, \{\{1, 2\}, \{3, \ldots, m\}\}, \ldots\}$ is the set of all pairs of recombinable homologous regions, and $Z_{exp}^{(t)} = \sum_{\mathbf{c} \in \mathcal{C}} p_{\mathbf{c}}^{(t)} e^{-\eta(1-f(\mathbf{c}))}$.*

**Proof**: Summing over all possible recombinations and normalizing gives the result.

As a result, it can be shown (*cf.* §A.2) that the marginal allele densities (after mutation) obey the dynamics below (*cf.* Meir & Parkes 2015, Lemma A.2) —albeit, again, only with high probability, due to the distorting effect of mutation:

**Lemma 7**. *With high probability, MA's marginal allele densities equal*

$$\frac{1}{Z_{\exp}^{(T)}} \sum_{\mathbf{c} \in \mathcal{C}|a_l = a} p_{\mathbf{c}}^{(T)} e^{-\eta(1-f(\mathbf{c}))}.$$

**Proof**: See §A.2.

In turn, Lemma 7 readily implies that, with high probability, the marginal allele densities under MA are precisely the strategies obtained under the coordination game with Hedge updates as described in §2.3 (*cf.* Meir & Parkes 2015, Proposition A.3):

**Lemma 8**. *With high probability, after $T > 0$ generations, the marginal allele densities under MA are identical to the mixed strategies after $T$ rounds of the one-sided coordination game described in §2.3, with payoffs $\Delta = f$, and Hedge updates with 'learning rate' $\eta$.*

**Proof**: By Lemma 7 (w.h.p.),

$$
\begin{aligned}
\sum_{\mathbf{c} \in \mathcal{C}|a_l=a} p_{\mathbf{c}}^{(T+1)} &= \frac{1}{Z_{\exp}^{(T)}} \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} p_{\mathbf{c}}^{(T)} e^{-\eta(1-f(\mathbf{c}))} \\
&= \frac{1}{Z_{\exp}^{(T)}} \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} \left( \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} p_{\mathbf{c}}^{(T)} p_{\mathbf{c}|a_l=a}^{(T)} \right) e^{-\eta(1-f(\mathbf{c}))} \\
&= \frac{1}{Z_{\exp}^{(T)}} \left( \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} p_{\mathbf{c}}^{(T)} \right) \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} e^{-\eta(1-f(\mathbf{c}))} p_{\mathbf{c}|a_l=a}^{(T)} \\
&= \frac{1}{Z_{\exp}^{(T)}} \left( \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} p_{\mathbf{c}}^{(T)} \right) \mathbb{E}[e^{-\eta(1-f(\mathbf{c}))} \mid a_l = a],
\end{aligned}
$$

which equals the Hedge dynamics in (7) with $\sum_{\mathbf{c} \in \mathcal{C}|a_l=a} p_{\mathbf{c}}^{(T+1)} = X_l^{(T)}(a)$.

We can then apply Hedge's regret bound in Lemma 1 above to MA's time-averaged fitness regret, which gives Theorem 2 (restated here as Theorem 4):

**Theorem 4**. *Let $\eta = 4(\ln |\mathcal{C}|/(1-\tilde{f}_{min}))^{1/2}$ and $\tilde{f}_{min} \geq f(\cdot) \in [0,1]$. Then with high probability, after $T > 0$ generations, MA's time-averaged expected fitness regret is at most*

$$
O\left( \sqrt{(1 - \tilde{\Delta}_{\min})/T} \right) + O(1/T).
$$

**Proof**: Lemma 8 implies Lemma 1 also holds for MA (w.h.p.) with $\Delta = f$ and $n^m = |\mathcal{C}|$. The result follows.

In other words, *MA's population dynamics can be seen as a Hedge coordination game between genes* where the payoffs are the fitnesses, and the 'learning rate' is the *tunable* parameter $\eta$ —which, if set to its optimal rate $\eta \approx \Theta(\ln |\mathcal{C}|)$, yields *no regret* dynamics —albeit at a slower worst-case convergence rate of $O(1/\sqrt{T})$. Note, however, that convergence may be as fast as $O(1/T)$ for $\tilde{f}_{\min} \to 1$.

*3.4 Taking Stock*

The bounds in Theorems 1 and 2 show that, with high probability, MA's time-averaged expected fitness reaches the best achievable local minimum at a worst-case rate $O(1/\sqrt{T})$ —or as high as $O(1/T)$ for $\tilde{f}_{\min} \to 1$. GA, on the other hand, reaches a 'glass ceiling' bounded by the selection strength $s$ —albeit at the much faster rate $O(1/T)$. The performance gap between the two will be particularly salient for 'difficult' landscapes with high selection strength $s$ (Figure 1, *left*). However, in the *weak selection* limit $s \to 0$, the 'glass ceiling' becomes immaterial, and GA reaches the best achievable local minimum at a much faster rate than MA (Figure 1, *right*). I note that evolutionary models often assume weak selection —if only for analytical convenience (*cf.* Bürger 2011, §3.4).

Also note that, as pointed out earlier, the MW regret bounds in Lemmas 1 and 3 are tight up to constant
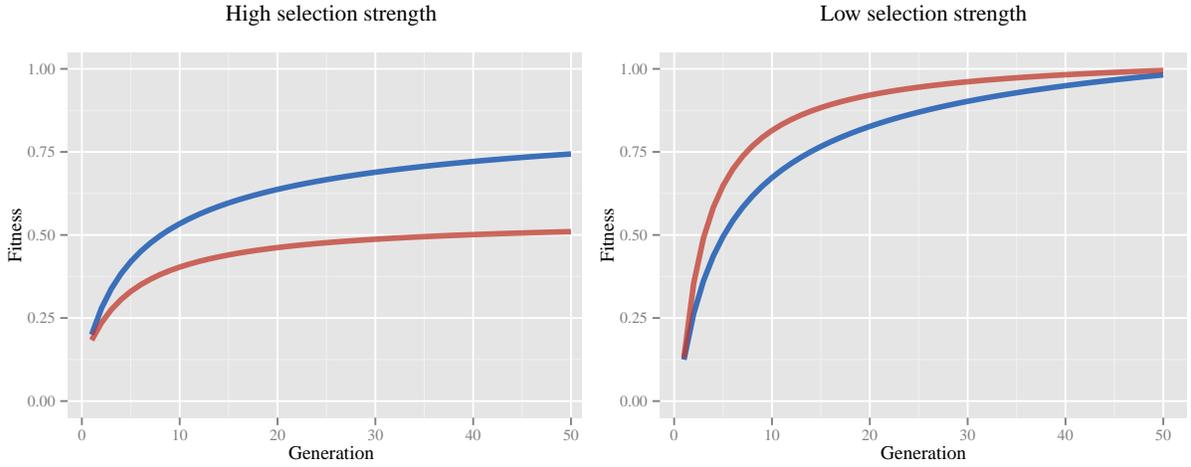
**Figure 1:** 'Mock' time-averaged expected fitness of MA (blue) *vs.* GA (red) under 'high' selection strength (*left*) and 'low' selection strength (*right*).

terms (*cf.* Arora, Hazan & Kale 2012, §4), so the 'mock' performance curves in Figure 1 ought to reflect well actual performance —provided of course the leading constants are small enough.

### 3.5 A heuristic interpretation of $\eta^*$

Interpretively, the bound in Theorem 2 is rather unintuitive, since it only holds when $\eta$ is set to the somewhat opaque optimal value of $\eta^* \approx \Theta(\ln|\mathcal{C}|)$. Nevertheless, I claim this order of magnitude admits a 'natural' interpretation: namely, the rate that (roughly) maximizes the entropy of the marginal allele densities subject to the information encoded in the fitness landscape —thus resulting in a *maximally non-committal* proxy for fitness that threads well the 'thermal equilibrium' between exploration and exploitation.

My argument is heuristic. Let $\check{X}_l = e^{-\lambda q(l)} / \sum_{l' \in \{1,\dots,m\}} e^{-\lambda q(l')}$ be the MaxEnt density for some quantity $q(l)$, and let $X_l^{(T)} = \sum_{\mathbf{c} \in \mathcal{C}} p_{\mathbf{c}}^{(T)}$ be the marginal allele densities after $T$ generations. Then by Lemma 2 in Schapire & Freund (1999), under MA with $\eta = \eta^*$, after $T > 0$ generations the relative entropy between the two is bounded by (w.h.p.)

$$\mathrm{RE}\left(\check{X}_l \,\middle\|\, X_l^{(T)}\right) \le \eta T \mathbb{E}[q(l)] - \frac{1 - \eta^2}{2\eta} \sum_{t=1}^{T} \mathbb{E}_{\mathbf{c} \sim P^{(t)}}[f(\mathbf{c})] + \ln|\mathcal{C}|.$$

Therefore, on every generation $T > 0$, with high probability, the marginal allele densities can be seen as a MaxEnt density that best reflects the quantity

$$\mathbb{E}[q(l)] \approx (1 - O(1/\sqrt{T})) \left( \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\mathbf{c} \sim P^{(t)}}[f(\mathbf{c}) \mid \mathbf{c}_l = l] \right) + O(1/\sqrt{T}) \ln|\mathcal{C}|.$$

In other words, with high probability, the marginal allele densities approximately converge —at rate $O(1/\sqrt{T})$— to the MaxEnt density that best reflects the expected time-averaged fitnesses for the allele.

This is not to say, however, that actual cultures *do* follow these dynamics. But it does show that they *should* from an information-theoretical point of view, insofar as they ensure the marginal allele densities approximately converge to an optimal encoding of the fitness landscape.

## 4 Experiments

Do these bounds hold in practice? As it turns out, the evidence is mixed.

### 4.1 Ms. Pac-Man Task

For my main experiment, I compared the performance of MA *vs.* GA on the task of evolving high-scoring controllers for *Ms. Pac-Man*: a challenging sequel to the original *Pac-Man* game with non-deterministic ghosts and complex scoring rules. For comparability purposes, I used the same parameters as in a similar experiment by Galván-López *et al.* (2010) except for: the selection method (which naturally changes), and the chromosome size (which was not reported therein).

The fitness landscape is defined by:

$$f(\mathbf{c}) = \frac{1}{\max\{\text{Score}(\cdot)\}} \text{Score}(\text{GE}_{\mathcal{G}}(\mathbf{c})), \tag{11}$$

where $\text{Score}(\cdot)$ denotes the (normalized) score of the given controller on a single *Ms. Pac-Man* game against 'legacy' ghosts on a benchmark Ms. Pac-Man simulator (*cf.* Lucas 2010; Rohlfshagen *et al.* 2013), and the chromosomes $\mathbf{c} \in \mathcal{C} = \bigotimes_{i=1}^{10}\{0, \ldots, 2^{16} - 1\}$ are 16-bit unsigned integer vectors of length $m = 10$. Controllers, in turn, are encoded using the 'grammatical evolution' (GE) scheme proposed in O'Neill & Ryan (2003): that is, as the result of a parameterized leftmost derivation (of length at most $m \times d = 10 \times 3$) on a given context-free grammar $\mathcal{G}$ such that, on the $i$-th rule application, the $(\mathbf{c}_{(i \bmod 10)} \bmod k)$-th alternative out of $k$ available is chosen —or a 'null' controller if the derivation length exceeds $m$ times $d = 3$ 'wrap-arounds'.

Grammar $\mathcal{G} = \{\mathcal{V}, \Sigma, \mathcal{R}, V_0\}$:    Chromosome $\mathbf{c} = (15, 92, 33, 81, 17, 46)$:

$$
\begin{aligned}
V_0 &\rightarrow B \,|\, BC \\
B &\rightarrow \texttt{if}(FEF)\{V_0\} \,|\, \texttt{if}(FEF)\{D\} \\
C &\rightarrow \texttt{else}\{D\} \,|\, \texttt{else}\{V_0\} \\
D &\rightarrow \texttt{go\_to(pill)} \,|\, \texttt{go\_to(ghost)} \\
E &\rightarrow \texttt{<} \,|\, \texttt{>} \,|\, \texttt{==} \\
F &\rightarrow \texttt{d\_ghost} \,|\, \texttt{d\_pill} \,|\, \texttt{dist}
\end{aligned}
$$

(15):    $B$
(92):    `if`$(FEF)\{D\}$
(33):    `if`$(\texttt{dist}EF)\{D\}$
(81):    `if(dist>`$F)\{D\}$
(17):    `if(dist>d_pill){`$D\}$
(46):    `if(dist>d_pill){go_to(ghost)}`

**Figure 2:** Toy example of GE encoding. Adapted from Galván-López *et al.* (2010).

To fix ideas, consider the toy example in Figure 2, adapted from Galván-López *et al.* (2010). Given a grammar $\mathcal{G} = \{\mathcal{V}, \Sigma, \mathcal{R}, V_0\}$ (*left*) and chromosome $\mathbf{c} = (15, 92, 33, 81, 17, 46)$ (*right*), we compute

$\text{GE}_\mathcal{G}(\mathbf{c})$ as follows:

- On the first rule application $V_0 \to B|BC$, we choose the $(15 \bmod 2)$-st or *first* alternative —which yields

$$\text{GE}_\mathcal{G}(\mathbf{c}) = B.$$

- On the second rule application $B \to \texttt{if(}FEF\texttt{)}\{V_0\} \mid \texttt{if(}FEF\texttt{)}\{D\}$, we choose the $(92 \bmod 2)$-st or *first* alternative —which yields

$$\text{GE}_\mathcal{G}(\mathbf{c}) = \texttt{if(}FEF\texttt{)}\{D\}.$$

- On the third rule application $F \to \texttt{d\_ghost} \mid \texttt{d\_pill} \mid \texttt{dist}$, we choose the $(34 \bmod 3)$-st or *third* alternative —which yields

$$\text{GE}_\mathcal{G}(\mathbf{c}) = \texttt{if(dist}EF\texttt{)}\{D\}.$$

- And so on until no non-terminals are left, 'cycling' through the chromosome up to $d = 3$ times if necessary. However, if after 3 'wrap-arounds' there are still non-terminals left, return `NULL`.

The actual grammar used is given in §B.1. Note the basis functions already encapsulate considerable high-level functionality, so the generalizability of these results to more general problem domains should not be overstated.

In addition to the fitness function $f(\cdot)$ and chromosome set $\mathbf{c} \in \mathcal{C}$ described above, I used the following parameters for both MA and GA —as per Galván-López *et al.* (2010, §4)—: initial random population of size $|\mathcal{P}_0| = 100$, number of generations $T = 50$, mutation probability $p_m = 0.1$, crossover probability $p_r = 0.7$, and $\tilde{f}_{\min} = 0$.
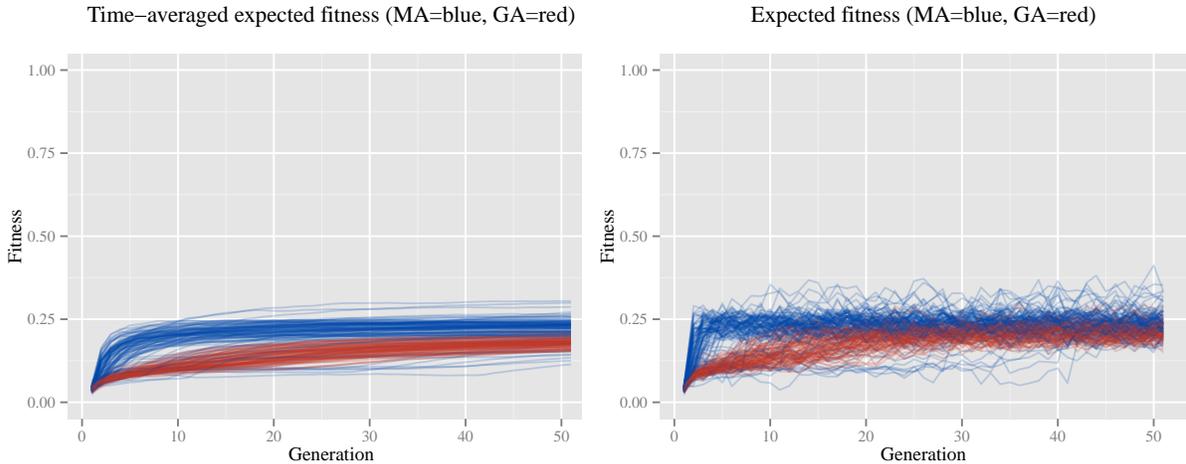


**Figure 3:** Empirical time-averaged (*left*) and 'instantaneous' (*right*) expected fitness of MA (blue) *vs.* GA (red) on the Ms. Pac-Man controller task ($N = 100$ replications).

The results for $N = 100$ Monte Carlo replications are given in Figure 3 (*left*). The evidence is mixed. Even though technically MA's time-averaged expected fitness is 'significantly' higher than GA's on every generation (one-sided KS test statistic $D_{N,N}^+ \geq 0.66$, significant at the $\alpha < 10^{-19}$ level), there

are important discrepancies with theory. Given the difficulty of the task, one would have expected a 'jagged' landscape (*i.e.* high selection strength $s$) and a low minimum achievable fitness $\tilde{f}_{\min}$, which should result in both: (a) a large performance gap between MA and GA, and (b) a slower convergence rate of $O(1/\sqrt{T})$ for MA (as opposed to $O(1/T)$ for GA). Instead, the (implied) gap or 'glass ceiling' in performance between MA and GA after $T$ generations concentrates barely above $0.05$; and the time-averaged expected fitnesses under GA do not *seem* to be following a $O(1/T)$ law, but a $O(1/\sqrt{T})$ law instead that seems, at any rate, *slower* than MA's[2]. These results strongly suggest that the leading constants in the bounds in Theorems 1 and 2 might be more significant than expected.

For comparison, the 'instantaneous' expected fitness curves are given in Figure 3 (*right*). Note, however, the bounds in Theorems 1 and 2 do *not* explicitly analyze this quantity —so I refrain from interpreting these results.

## 4.2 Sum-Of-Bits Task

For my second experiment, I compared the performance of MA *vs.* GA on the much simpler sum-of-bits task from CS63 last semester, with joint parameters $\mathbf{c} \in \mathcal{C} = \bigotimes_{i=1}^{15}\{0,1\}$, $f(\mathbf{c}) = f((a_1, \ldots, a_{15})) = \frac{1}{15}\sum_{i=1}^{15} a_i$, $|\mathcal{P}_0| = 50$, $T = 50$, $p_m = 0.01$, $p_r = 0.6$ and $\tilde{f}_{\min} = 0$.
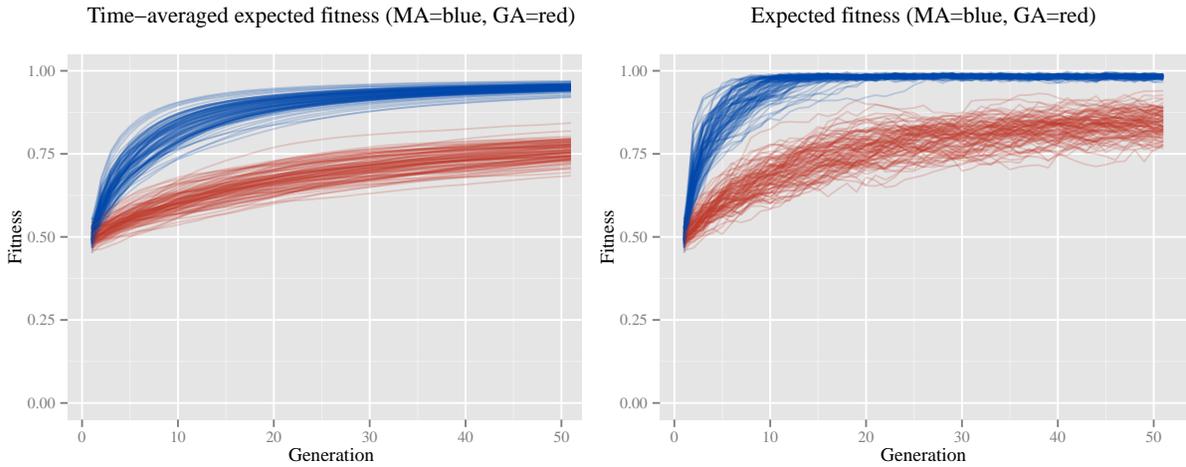


**Figure 4:** Empirical time-averaged (*left*) and 'instantaneous' (*right*) expected fitness of MA (blue) *vs.* GA (red) on the sum-of-bits task ($N = 100$ replications).

The results for 100 Monte Carlo replications are given in Figure 4 (*left*). Again, the results are mixed: even though MA's time-averaged expected fitness does 'significantly' outperform GA's on every generation (one-sided KS test statistic $D_{N,N}^+ \geq 0.86$, significant at the $\alpha < 10^{-33}$ level), and there is some

---

[2]More concretely, fitting a $\beta_1 - \beta_2 x^{-1/2}$ curve to each of the $N = 100$ time-averaged expected fitness curves under GA yields $R^2$ values in the (interquartile) range of $[0.94, 0.98]$, a mean of $0.96$, and a max of $0.99$; while fitting a $\beta_1 - \beta_2 x^{-1}$ curve instead yields $R^2$ values in the (interquartile) range of $[0.47, 0.58]$, a mean of $0.52$, and a maximum of $0.73$.

Interestingly, in the case of MA, fitting a $\beta_1 - \beta_2 x^{-1/2}$ curve to each of its $N = 100$ time-averaged expected fitness curves yields $R^2$ values in the (interquartile) range of $[0.96, 0.99]$, a mean of $0.98$, and a maximum of $0.99$; while fitting a $\beta_1 - \beta_2 x^{-1}$ curve $R^2$ values in the (interquartile) range of $[0.78, 0.94]$, a mean of $0.89$, and a maximum of $0.99$ —indicative perhaps of a non-trivial $O(1/T)$ term. Note, however, these observations are largely heuristic and qualitative.

evidence of a 'glass ceiling' or gap between MA and GA; the apparent convergence rates do not match the theoretical predictions of MA converging *slower* than GA. For comparison, the (uninterpreted) 'instantaneous' expected fitness curves are given in Figure 4 (*right*).

## 5 Discussion and Open Problems

Theorems 1 and 2 above suggest that in 'jagged' landscapes far from the weak selection regime, the accelerated spread of memes in cultural evolution *may* improve its convergence properties *vis-à-vis* genetic evolution. Analytically, these results follow readily from the recent groundbreaking reduction of Meir & Parkes (2015) and Chastain *et al.* (2014) of evolution's marginal population dynamics to the update dynamics of a coordination game between genes/memes with MW-like update rules. Interpretively, however, the reduction is hardly intuitive, the empirical evidence is mixed, and the *semantics* of the improvement mechanism are far from clear.

Can these bounds be improved? Certainly they could —and should— be restated in finite-sample terms —especially since the core analysis relies so heavily on the 'true' population densities $p_\mathbf{c}$. Lemma 5 in Freund & Schapire (1999) provides a useful starting point here, as it bounds Hedge's cumulative payoff deviations from their expected value by $O(1/\sqrt{T})$ (with high probability).

Another extension worth investigating further is translating the time-averaged regret bounds in Theorems 1 and 2 to the more natural (albeit much less studied) notion of *instantaneous* regret in (5). Note, however, this is generally *unfeasible* unless we are willing to either: (a) accept *some* kind of polynomially-discounted, time-averaged regret notion instead (*cf.* Cesa-Bianchi & Lugosi 2006, §2.11), or (b) make further assumptions about the fitness landscape (*cf.* Hazan & Seshadhri 2007). Having said that, the heuristic observation that MA also *seems* to (slightly) outperform GA in terms of 'instantaneous' regret on both the Ms. PacMan task and the sum-of-bits task (see also §B.2) should, at the very least, motivate further research attention on the connection between the two notions.

Also note that, for EC purposes, these bounds are largely impractical since they focus on the *expected*, *time-averaged* fitness of evolved populations, not its 'instantaneous' *maximum*. To get such a bound, techniques for studying the dynamics of second-order quantites under MW might be necessary —so the work in Cesa-Bianchi *et al.* (2007) might provide a useful starting point. Note, however, that these bounds rely on *adaptive*, time-varying learning rates $\eta_t$ —which may be rather difficult to 'motivate' interpretively.

Lastly, from a sociobiological perspective, it is an open question whether or not meme transmission in *actual* cultures follows an exponential law with rate $\eta^* \approx \Theta(\ln |\mathcal{C}|)$. Indeed, merely *quantifying* the value of $\eta$ in actual cultures seems like a formidable empirical task. Having said that, the normative interpretation proposed in §3.5 does suggest intriguing avenues of inquiry. For instance, is $\eta$ constant across the anthropological record? Does $\eta$ change across societies? If not, can we distinguish 'developed' from 'developing' societies by their differences in $\eta$? Or, *in extremis*, is societal collapse associated with *sub-optimal* values of $\eta$? To be able to address these issues quantitatively would surely be worth the empirical effort.

# Appendix

## A Proofs of Lemmas 5 and 7

The proofs below rely on the following crucial lemma:

**Lemma 9**. *With high probability, after $T$ generations, the $\ell_2$ distortion induced by mutation on the marginal allele densities of both GA and MA vanishes with size.*

**Proof**: To lighten the notation, I substitute $X_T$ for $\sum_{\mathbf{c} \in \mathcal{C} | a_l = a} \tilde{p}_{\mathbf{c}}^{(T)}$ —so that $\tilde{X}_T$ and $X_T$ denote the marginal allele densities of GA (or MA) before and after mutation, respectively.

Assume without loss of generality that $X_0 = \tilde{X}_0 = 0$ (almost surely). Note for $T > 0$, the martingale differences $\frac{1}{2}\|(X_T - \tilde{X}_T) - (X_{T-1} - \tilde{X}_{T-1})\|_2$ are bounded a.s. by

$$
\begin{aligned}
\tfrac{1}{2}\|(X_T - \tilde{X}_T) - (X_{T-1} - \tilde{X}_{T-1})\|_2 &= \tfrac{1}{2}\|(X_T - X_{T-1}) - (\tilde{X}_T - \tilde{X}_{T-1})\|_2 \\
&= \tfrac{1}{2}\|(X_T - X_{T-1}) - (\tilde{Q}X_T - \tilde{Q}X_{T-1})\|_2 \\
&= \tfrac{1}{2}\|(I - \tilde{Q})(X_T - X_{T-1})\|_2 \\
&\leq \tfrac{1}{2}\|(I - \tilde{Q})\|_2 \|(X_T - X_{T-1})\|_2 \\
&\leq (\tfrac{m}{m-1} - p_m) \cdot \min\{2, O(1/\sqrt{T})\} \\
&\leq (1 - p_m) \cdot \min\{1, O(1/\sqrt{T})\} \\
&\leq \min\{1, O(1/\sqrt{T})\} \\
&\leq 1,
\end{aligned}
$$

where $\tilde{Q} = \frac{1}{m}p_m J + (1 - \frac{m}{m-1}p_m)I$ is a $m \times m$ symmetric, doubly stochastic matrix with $\lambda_{\max}(\tilde{Q}) = p_m + (1 - \frac{m}{m-1})$ that encodes the effect of mutation.

Therefore, by extension of the Azura-Hoeffding inequality to vector-valued martingales (*cf.* Theorem 7.4 in Hayes 2005), the probability that the $\ell_2$ distortion induced by mutation is greater than $\varepsilon > 0$ decreases exponentially with $\varepsilon$:

$$
\Pr\left[\|X_T - \tilde{X}_T\|_2 \geq \varepsilon\right] < 2e^{-(2\varepsilon - O(1/\sqrt{T}))^2},
$$

which gives the result.

### A.1 Proof of Lemma 5

Follows immediately from Lemma 9.

### A.2 Proof of Lemma 7

The proof proceeds almost identically to that of Lemma A.2 in Meir & Parkes (2015). To lighten

the notation, I substitute $\alpha_{(\cdot)}^{(T)}$ for $e^{-\eta(1-f(\cdot))}p_{(\cdot)}^{(T)}$, and $J^l$ for $\{J_1 \cup \{l\}, J_2 \setminus \{l\}\}$.

By Lemmas 4 and 9, MA's marginal allele densities $\sum_{\mathbf{c} \in \mathcal{C}|a_l=a} p_{\mathbf{c}}^{(T+1)}$ are given by (w.h.p.)

$$\sum_{\mathbf{c} \in \mathcal{C}|a_l=a} \left( p_m \sum_{\mathbf{d} \in \mathcal{C}} \frac{1}{|\mathcal{R}|} \sum_{J \in \mathcal{R}} \frac{\alpha_{\mathbf{c}_{J_1} \diamond \mathbf{d}_{J_2}}^{(T)} \alpha_{\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}}^{(T)}}{\left( Z_{\exp}^{(T)} \right)^2} + (1-p_m) \frac{\alpha_{\mathbf{c}}^{(T)}}{Z_{\exp}^{(T)}} \right)$$

$$= p_m \left( \frac{1}{|\mathcal{R}|} \sum_{J \in \mathcal{R}} \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} \sum_{\mathbf{d}_{J_k} \in J_k} \frac{\alpha_{\mathbf{c}_{J_1} \diamond \mathbf{d}_{J_2}}^{(T)} \alpha_{\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}}^{(T)}}{\left( Z_{\exp}^{(T)} \right)^2} \right) + (1-p_m) \left( \frac{1}{Z_{\exp}^{(T)}} \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} \alpha_{\mathbf{c}}^{(T)} \right),$$

where the first factor $\frac{1}{|\mathcal{R}|} \sum_{J \in \mathcal{R}} \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} \sum_{\mathbf{d}_{J_k} \in J_k} \alpha_{\mathbf{c}_{J_1} \diamond \mathbf{d}_{J_2}}^{(T)} \alpha_{\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}}^{(T)} / \left( Z_{\exp}^{(T)} \right)^2$ can be written as

$$\frac{1}{\left( Z_{\exp}^{(T)} \right)^2 |\mathcal{R}|} \sum_{J \in \mathcal{R} \,|\, \{l\} \notin J} \left( \sum_{\mathbf{c}_{J_k}, \mathbf{d}_{J_k} \in J_k} \alpha_{\mathbf{c}_{J_1} \diamond \mathbf{d}_{J_2}}^{(T)} \alpha_{\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}}^{(T)} + \sum_{\mathbf{c}_{J_k^l}, \mathbf{d}_{J_k^l} \in J_k^l} \alpha_{\mathbf{c}_{J_1^l} \diamond \mathbf{d}_{J_2^l}}^{(T)} \alpha_{\mathbf{d}_{J_1^l} \diamond \mathbf{c}_{J_2^l}}^{(T)} \right)$$

$$= \frac{1}{\left( Z_{\exp}^{(T)} \right)^2 |\mathcal{R}|} \sum_{J \in \mathcal{R} \,|\, \{l\} \notin J} \left( \sum_{\mathbf{c}_{J_2} \in J_2} \sum_{\mathbf{d}_{J_1} \in J_1} \alpha_{\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}}^{(T)} \sum_{\mathbf{c}_{J_1} \in J_1} \sum_{\mathbf{d}_{J_2} \in J_2} \alpha_{\mathbf{c}_{J_1} \diamond \mathbf{d}_{J_2}}^{(T)} \right.$$

$$\left. + \sum_{\mathbf{c}_{J_1^l} \in J_1^l} \sum_{\mathbf{d}_{J_2^l} \in J_2^l} \alpha_{\mathbf{c}_{J_1^l} \diamond \mathbf{d}_{J_2^l}}^{(T)} \sum_{\mathbf{d}_{J_1^l} \in J_1^l} \sum_{\mathbf{c}_{J_2^l} \in J_2^l} \alpha_{\mathbf{d}_{J_1^l} \diamond \mathbf{c}_{J_2^l}}^{(T)} \right)$$

$$= \frac{1}{Z_{\exp}^{(T)} |\mathcal{R}|} \sum_{J \in \mathcal{R} \,|\, \{l\} \notin J} \left( \sum_{\mathbf{c}_{J_2} \in J_2} \sum_{\mathbf{d}_{J_1} \in J_1} \alpha_{\mathbf{d}_{J_1} \diamond \mathbf{c}_{J_2}}^{(T)} + \sum_{\mathbf{c}_{J_1^l} \in J_1^l} \sum_{\mathbf{d}_{J_2^l} \in J_2^l} \alpha_{\mathbf{c}_{J_1^l} \diamond \mathbf{d}_{J_2^l}}^{(T)} \right)$$

$$= \frac{1}{Z_{\exp}^{(T)} |\mathcal{R}|} \sum_{J \in \mathcal{R} \,|\, \{l\} \notin J} \left( \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} \alpha_{\mathbf{c}}^{(T)} + \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} \alpha_{\mathbf{c}}^{(T)} \right)$$

$$= \frac{1}{Z_{\exp}^{(T)}} \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} \alpha_{\mathbf{c}}^{(T)} \left( \frac{2}{|\mathcal{R}|} \sum_{J \in \mathcal{R} \,|\, \{l\} \notin J} 1 \right)$$

$$= \frac{1}{Z_{\exp}^{(T)}} \sum_{\mathbf{c} \in \mathcal{C}|a_l=a} \alpha_{\mathbf{c}}^{(T)}.$$

Thus we have (w.h.p.)

$$
\sum_{\mathbf{c}\in\mathcal{C}|a_l=a} p_{\mathbf{c}}^{(T+1)} \;=\; p_m\left(\frac{1}{Z_{\exp}^{(T)}}\sum_{\mathbf{c}\in\mathcal{C}|a_l=a}\alpha_{\mathbf{c}}^{(T)}\right) + (1-p_m)\left(\frac{1}{Z_{\exp}^{(T)}}\sum_{\mathbf{c}\in\mathcal{C}|a_l=a}\alpha_{\mathbf{c}}^{(T)}\right)
$$

$$
=\; \frac{1}{Z_{\exp}^{(T)}}\sum_{\mathbf{c}\in\mathcal{C}|a_l=a} e^{-\eta(1-f(\mathbf{c}))}p_{\mathbf{c}}^{(T)}.
$$

## B Further Experimental Details

### B.1 Grammar

The full grammar used $\mathcal{G} = \{\mathcal{V}, \Sigma, \mathcal{R}, \texttt{<prog>}\}$ is given in Figure 5 below. It is essentially identical to the one used in Galván-López *et al.* (2010, Figure 2) except for slight differences in the method calls. The result is a valid JavaScript function with signature `get_move(g)`, where `g` is a reference to the main game object (of type `pacman.game.Game`) of the Java simulator used in Galván-López *et al.* (2010). The interfacing is implemented *via* Java 8's *Nashorn* JavaScript interpreter, which allows seamless referencing of Java classes from within the interpreter's environment.

```
     <prog>  →  function get_move(g){
                    var next = Java.type('pacman.game.Constants').MOVE.NEUTRAL);
                    var ctl = Java.type('main.ge_pacman_controller');
                    var gt = <gt>, ws = <ws>, agd = <agd>;
                    if(ctl.edible_ghosts == 0){ <stat> } else{ <stat> }
                    return next; }
       <gt>  →  1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
                16 | 17 | 18 | 19 | 20
       <ws>  →  3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19
      <agd>  →  1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
     <stat>  →  <ifs> | <ifs> else { <act_or_stat> }
      <ifs>  →  if(<var><cmp><var>){ <act_or_stat> } |
                if(ctl.adbg(g)<le>gt) { <act_or_stat> } |
                if(ctl.igd(g)<le>ws) { next = (<avc_or_ppc>)(g); }
      <var>  →  gt | ws | agd | ctl.adbg | ctl.igd
      <cmp>  →  < | > | <= | >= | ==
       <le>  →  < | <=
<act_or_stat> →  next = (<act>)(g); | <stat>
      <act>  →  ctl.npd | ctl.ang | ctl.ngd | (ctl.np>0)?ctl.npd:ctl.ang
<avc_or_ppc> →  avc | ppc
```

**Figure 5:** Full grammar used in the main experiment. Adapted from Galván-López *et al.* (2010, Figure 2).

Note the resulting code calls upon static methods from the class `main.ge_pacman_controller`, which implements the high-level function basis $\{adbg, ang, avc, igd, npd, ngd, nig, nppd\}$ as described (where applicable) in Galván-López *et al.* (2010, *cf.* Table 1). Also note Table 1 in (*ibid*) is missing the functions `adbg` (which computes the average distance between ghosts), `avc` (which avoids the closest ghost) and `igd` (which computes the distance to the closest inedible ghost).

### B.2 Simulator Details

Experiments were run on version 6.2 of the benchmark Java simulator used first in the IEEE CIG 2009 Ms. Pac-Man competition (*cf.* Lucas 2010; Rohlfshagen *et al.* 2013). Note it differs slightly from the original Ms. Pac-Man in that it has no bonus 'fruit' pills and the speed of Ms. Pac-Man and the ghosts (when inedible) are identical.

For a given controller, the function Score($\cdot$) described in §4.1 returns the unnormalized game score on a single level ($\max\{\text{Score}(\cdot)\} = 15420$) of the controller against the built-in `controllers.Legacy` ghosts. Note the game is run in 'asynchronous' mode —that is, with no time limit for the controllers to make a move.

### B.3 Additional Figures

Figure 6 reports the best fitness of populations evolved under MA (*blue*) and GA (*red*) for the Ms. Pac-Man controller task (*left*) and the sum-of-bits task (*right*). Note, however, the bounds in Theorems 1 and 2 do *not* explicitly control these quantities, so these results should be considered exploratory and heuristic at best.
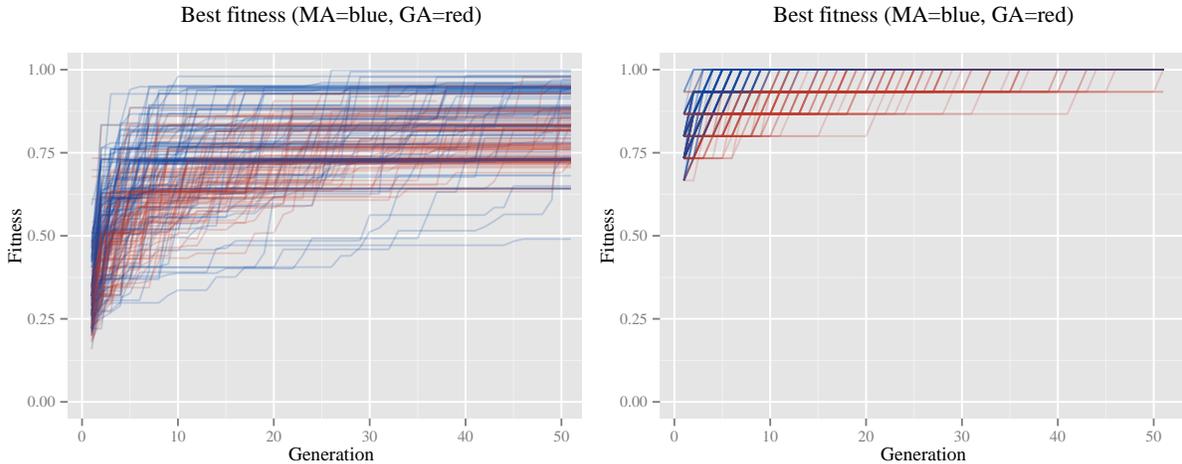


**Figure 6:** Best fitnesses of MA (blue) *vs.* GA (red) on the Ms. Pac-Man controller task (*left*) and the sum-of-bits task (*right*) ($N = 100$ replications).

# References

Arora, S., Hazan, E. and Kale, S. (2012) The multiplicative weights update method: a meta-algorithm and applications *Theory of Computing* 8:121-164.

Boyd, R. and Richerson, P.J. (1985) *Culture and the Evolutionary Process*. University of Chicago Press.

Boyd, R., Richerson, P.J. and Henrich, J. (2011) Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral Ecology and Sociobiology* 65(3):431-444.

Bürger, R. (2011) Some mathematical models in evolutionary genetics. *The Mathematics of Darwin's Legacy*. Springer.

Cavalli-Sforza, L.L. and Feldman, M.W. (1981) *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press.

Cesa-Bianchi, N. and Lugosi, G. (2006) *Prediction, Learning and Games*. Cambridge University Press.

Cesa-Bianchi, N., Mansour, Y. and Stoltz, G. (2007) Improved second-order bounds for prediction with expert advice. *Machine Learning* 66:321-352.

Chastain, E., Livnat, A., Papadimitriou, C. and Vazirani, U. (2014) Algorithms, games, and evolution. *Proceedings of the National Academy of Sciences* 111(29):10620-10623.

Dawkins, R. (1976) *The Selfish Gene*. Oxford University Press.

De la Maza, M. and Tidor, B. (1991) Boltzmann weighted selection improves performance of genetic algorithms. A.I. Memo 1345, CSAIL.

Derrida, B. (1981) Random-energy model: an exactly solvable model of disordered systems. *Physical Review B* 24(5):2613-2626.

Feldman, V. (2008) Evolution from learning algorithms. *STOC 2008*.

Freund, Y. and Schapire, R.E. (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55:119-139.

Freund, Y. and Schapire, R.E. (1999) Adaptive game playing using multiplicative weights. *Games and Economic Behavior* 29(1):79-103.

Galván-López, E., Swafford, J.M., O'Neill, M. and Brabazon, A. (2010) Evolving a Ms. PacMan controller using grammatical evolution. *EvoApplications 2010*.

Goldberg, D.E. (1990) A note on Boltzmann tournament selection for genetic algorithms and population-oriented simulated annealing. *Complex Systems* 4:445-460.

Hayes, T.P. (2005) A large-deviation inequality for vector-valued martingales. [https://cs.unm.edu/~hayes/papers/].

Hazan, E. and Seshadhri, C. (2007) Adaptive algorithms for online decision problems. *ECCC* 14.

Holland, J.H. (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

Kanade, V.N. (2012) *Computational Questions in Evolution*. Ph.D. Thesis. Harvard University.

Lucas, S. (2010) Ms. Pac-Man competition. [http://cswww.essex.ac.uk/staff/sml/pacman/PacManContest.html].

Meir, R. and Parkes, D. (2015) On sex, evolution, and the multiplicative weights update algorithm. arXiv: 1502.05056v1.

O'Neill, M. and Ryan, C. (2003) *Grammatical Evolution: Evolutionary Automatic Programming in an Arbitrary Language*. Kluwer.

Prügel-Bennett, A. and Shapiro, J.L. (1994) An analysis of genetic algorithms using statistical mechanics. *Physical Review Letters* 72:1305-1309.

Rohlfshagen, P., Robles, D. and Lucas, S. (2013) *Ms. Pac-Man vs. Ghosts League*. [https://web.archive.org/web/20140318083951/http://www.pacman-vs-ghosts.net/].

Stoltz, G. and Lugosi, G. (2005) Internal regret in on-line portfolio selection. *Machine Learning* 59(1-2):125.

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D., Roskies, R., Scott, J.R. and Wilkins-Diehr, N. (2014) XSEDE: accelerating scientific discovery. *Computing in Science & Engineering* 16(5):62-74.

Valiant, L. (2009) Evolvability. *Journal of the ACM* 56(1):3.