Supporting Appendix

These appendices aim to provide comprehensive background and interpretation for all of the key concepts and results discussed in this work. This work draws from several rich areas of theory, but necessary concepts from information theory, game theory, and learning theory are introduced throughout. It is our hope that any interested reader can achieve an understanding of these topics by working through this material without having to rely on other sources. Appendices A, B, and C review relevant concepts and results from the literature and establish the formalisms used in this work. ***Readers interested only in proofs of key results from this work can turn to Appendix D. Supplemental methods and results from our selection experiments can be found in Appendix E.***

Contents

A Modeling Natural Selection **2**

A.1 The replicator dynamics model of natural selection . . . 2

A.1.1 Discrete-time replicator dynamics (discrete generations) . . . 2

A.1.2 Continuous-time replicator dynamics (overlapping generations) . . . 3

A.2 Modeling variable environments . . . 4

B Natural Selection as an Information Acquisition Process **5**

B.1 Information Measures . . . 5

B.2 Natural selection acquires information about the environment . . . 7

B.2.1 Information gain . . . 7

B.2.2 Potential information . . . 9

B.2.3 Natural selection establishes mutual information between types and environmental histories . . . 10

C Natural Selection as a Learning Process **11**

C.1 Natural selection and Bayesian learning . . . 11

C.2 Natural selection as an online learning process . . . 11

C.2.1 The online learning problem . . . 11

C.2.2 The Population versus Environment game . . . 12

C.2.3 The Multiplicative Weights Updating (MWU) algorithm . . . 13

C.2.4 Natural selection as a fitness-based instance of MWU . . . 14

C.2.5 Two fitness-based notions of loss . . . 15

C.3 Stationary-target and general learning settings for evolving populations . . . 17

D The Efficiency of Natural Selection **19**

D.1 Mismatch load quantifies the cumulative fitness loss of selection . . . 19

D.1.1 Substitution load . . . 19

D.1.2 Mismatch load . . . 20

D.2 Regret quantifies relative growth performance . . . 20

D.2.1 Regret as relative mismatch load . . . 20

D.2.2 Regret as relative lineage fitness . . . 21

D.2.3 Target regret and Hindsight regret . . . 21

D.3 Regret and potential information . . . 22

D.3.1 Regret and changes in potential information . . . 22

D.3.2 Target regret bounds for stationary-target learning problems . . . 25

D.3.3 Hindsight regret bounds for general learning problems . . . 28

D.3.4 Comparison of regret bounds across learning rates . . . 30

D.4 Information gain and regret . . . 36

D.4.1 Information gain in a single update . . . 36

D.4.2 Information gain and target regret . . . 38

D.4.3 Information gain and hindsight regret . . . 40

D.5 Supplemental example learning problems . . . 42

E Selection Experiments **43**
E.1 Strain information . . . 43
E.2 Supplemental methods . . . 45
E.2.1 Transfer protocol . . . 45
E.2.2 Flow cytometry . . . 47
E.3 Supplemental results of selection experiments . . . 49

Appendix A: Modeling Natural Selection

A.1 The replicator dynamics model of natural selection

Replicator dynamics are a standard model of evolution that describe the action of natural selection in large populations without mutation, recombination, or other sources of new variation. Replicators are entities that generate copies of themselves and may represent genotypes, strategies, beliefs, etc. Populations may consist of multiple types of replicators, each with a characteristic fitness that gives its expected reproductive output. Replicator types with above-average fitness will increase their frequency in the population, and types with below-average fitness will decrease in frequency. Replicator dynamics are widely studied in population genetics, ecology, evolutionary game theory, learning theory, economics, and other contexts.

A.1.1 Discrete-time replicator dynamics (discrete generations)

Fitness and growth

First we consider a population of replicators with a set of n alternative types that reproduce in discrete generations. Let N_i^t be the number of individuals of the ith type in generation t. The expected number of descendants in generation $t+1$ per individual of type i in generation t is given by $W_i > 0$, which is referred to as the **Wrightian fitness**, or one-generation growth factor, of the ith type (In general, the fitness of a type may depend on the current environmental conditions, which we will develop further in Appendix A.2). The number of individuals of type i after a total of T generations is given by

$$N_i^T = N_i^0 (W_i)^T \,, \qquad \text{[A1]}$$

where N_i^0 is the initial number of individuals of type i. The expected number of descendants per initial individual after T generations is given by the ratio $N_i^T / N_i^0 = (W_i)^T$.

Suppose that we observe N_i^T individuals of type i after T generations. The average per-generation growth rate of this subpopulation over the interval is

$$\hat{r}_i := \frac{1}{T} \log\left(\frac{N_i^T}{N_i^0}\right).$$

When Wrightian fitness is constant across generations, $N_i^T = N_i^0 (W_i)^T$, and therefore

$$\hat{r}_i = \log W_i.$$

Thus, W_i is the *one-generation growth factor* of type i, while $\log W_i$ is its corresponding *per-generation growth rate*.

The **relative fitness** of the ith type, w_i, is considered with respect to the highest fitness of any type in the population, $W_* := \max_i W_i$:

$$w_i = \frac{W_i}{W_*} \,. \qquad \text{[A2]}$$

By this definition, the relative fitness of the optimal type is equal to 1 (i.e., $w_* = W_*/W_* = 1$), and the relative fitnesses of all other types take values in $(0, 1]$. As such, we may also choose to represent the relative fitness of the ith type as

$$w_i = 1 - k_i \,, \qquad \text{[A3]}$$

where $k_i \in [0, 1)$ is the **Wrightian selection coefficient** of the ith type, and $k_* = 0$ is the selection coefficient of the optimal type. This expression of relative fitness can be interpreted as saying that for every 1 offspring left by the optimal type there are $1 - k_i$ offspring left by the ith type.

Dynamics of type frequency change

The frequency of type i in the population in generation t is defined as

$$p_i^t = \frac{N_i^t}{\sum_z N_z^t} \,. \qquad \text{[A4]}$$

Then the change in the frequency of type i from one generation to the next is given by

$$p_i^{t+1} = \frac{N_i^t W_i}{\sum_z N_z^t W_z} = \frac{p_i^t W_i}{\sum_z p_z^t W_z} \,. \quad [A5]$$

Discrete-time replicator dynamics refer to this dynamical map

$$p_i^{t+1} = \frac{p_i^t W_i}{\langle W \rangle_{\boldsymbol{p}^t}} \,, \quad [A6]$$

where $\langle W \rangle_{\boldsymbol{p}^t} = \sum_i p_i^t W_i$ is the population mean fitness (equivalently the population's one-generation growth factor) at time t. This dynamic re-weights the frequencies of types proportional to their fitnesses such that types with above-average fitness increase in frequency and those with below-average fitness decrease in frequency.

A.1.2 Continuous-time replicator dynamics (overlapping generations)

Fitness and growth

While not the focus of the main text, we can also consider the evolution of populations with overlapping generations. Rather than adding individuals to the population at discrete unit intervals, individuals are added to the population at multiple instances per unit time. To capture this, we can express the expected reproductive output per unit time as

$$W_i = \left(1 + \frac{r_i}{\epsilon}\right)^{\epsilon} \,, \quad [A7]$$

where r_i gives the expected number of offspring above replacement (or below, if negative) that each i individual contributes to the population per unit time, and ϵ gives the number of instances that offspring are added per unit time. When $\epsilon = 1$, then this definition reduces to the discrete generation case, where $W_i = 1 + r_i$ defines the absolute Wrightian fitness of the ith type. On the other hand, fully continuous growth is represented by allowing new individuals to be added to the population at an infinite number of points per unit time, which results in the familiar exponential growth term

$$\lim_{\epsilon \to \infty} \left(1 + \frac{r_i}{\epsilon}\right)^{\epsilon} = e^{r_i} \,. \quad [A8]$$

Then the number of i individuals at time t in a continuously growing population is given by

$$N_i^t = N_i^0 e^{r_i t} \,, \quad [A9]$$

which is known as the Malthusian growth model. Here we see that the parameter r_i sets the exponential growth rate of the ith type, and we refer to this rate r_i as the **Malthusian fitness** of the ith type. The **Malthusian selection coefficient** s_i of the ith type is defined as the difference between the Malthusian fitness (i.e., growth rate) of the optimal type and that of the ith type

$$s_i = r_* - r_i \,. \quad [A10]$$

Dynamics of type frequency change

In this section, we consider the continuous-time analog of the discrete replicator dynamics presented above. By definition, the Malthusian fitness r_i is the instantaneous per-capita growth rate of type i (Equation A9). The abundance of type i therefore changes according to

$$\frac{dN_i^t}{dt} = r_i N_i^t. \quad [A11]$$

At time t, the instantaneous per-capita growth rate of the total population is the frequency-weighted mean of the growth rates of its constituent types. Consequently,

$$\frac{dN_{\text{tot}}^t}{dt} = \left(\sum_z p_z^t r_z\right) N_{\text{tot}}^t = \langle r \rangle_{\boldsymbol{p}^t} N_{\text{tot}}^t, \quad [A12]$$

where $\langle r \rangle_{\boldsymbol{p}^t} := \sum_z p_z^t r_z$ is the mean Malthusian fitness of the population.

The frequency of type i is its abundance divided by the total population abundance:

$$p_i^t = \frac{N_i^t}{N_{\text{tot}}^t}. \quad [A13]$$

Applying the quotient rule to this ratio and substituting the two growth equations above gives

$$\frac{dp_i^t}{dt} = \frac{d}{dt}\left(\frac{N_i^t}{N_{\text{tot}}^t}\right) \tag{A14}$$

$$= \frac{(r_i N_i^t)N_{\text{tot}}^t - N_i^t\left(\langle r\rangle_{\boldsymbol{p}^t} N_{\text{tot}}^t\right)}{\left(N_{\text{tot}}^t\right)^2} \tag{A15}$$

$$= \frac{N_i^t}{N_{\text{tot}}^t}\left(r_i - \langle r\rangle_{\boldsymbol{p}^t}\right) \tag{A16}$$

$$= p_i^t\left(r_i - \langle r\rangle_{\boldsymbol{p}^t}\right). \tag{A17}$$

This is the **continuous-time replicator equation**. A type increases in frequency when its Malthusian fitness exceeds the population mean fitness, and decreases in frequency when its Malthusian fitness falls below it.

A.2 Modeling variable environments

In general, the fitness of an individual depends on both its type and its environment, and we are ultimately interested in modeling the process of natural selection in contexts where conditions and fitnesses change over time. In this section we extend the standard replicator models to support both heterogeneous and time-varying environments, broadly defined.

Suppose the population occupies an environment that is comprised of a set of m distinct environmental conditions. We assume that every individual in the population experiences an independent environmental condition, as if each individual occupies a separate part (e.g., micro-environment) of the physical environment. The association of specific individuals to particular conditions is assumed to be random, and the probability that a given individual experiences the jth condition at time t is given by x_j^t. The condition distribution $\boldsymbol{x}^t$ defines the makeup of the environment at time t, and scenarios of environmental change can be modeled by specifying a particular sequence of environments $\{\boldsymbol{x}^0, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^{T-1}\}$.

We take a very expansive view of what constitutes an "environment." Essentially any set of contextual factors that have an impact on the fitnesses of types can be considered an environmental condition. For example, conditions may represent the abiotic conditions experienced, the types of other replicators encountered, or even the identity of other alleles carried by individuals. Similarly, the sequence of environments $\{\boldsymbol{x}^0, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^{T-1}\}$ may be generated according to any process, including as a function of time, as a function of type frequencies, as a stochastic process, as a "pre-programmed" sequence, etc. Throughout this work we consider various environmental contexts of interest as well as the fully general setting where no assumptions are made about the sequence of environments.

In the main text, we frame this model of discrete-time replicator dynamics in variable environments in terms of repeated play of a two-player, Population versus Environment game, which lends itself to learning theoretic analysis; see Appendix C.2.2 for more information about this formalization.

Discrete-time replicator dynamics in variable environments:

Let W_{ij} denote the Wrightian fitness, or one-generation growth factor, of type i in environmental condition j. The effective growth factor of type i under the condition distribution $\boldsymbol{x}^t$ is

$$\langle W_i\rangle_{\boldsymbol{x}^t} := \sum_j x_j^t W_{ij},$$

and the population mean fitness, equivalently its one-generation growth factor, is

$$\langle W\rangle_{\boldsymbol{p}^t,\boldsymbol{x}^t} := \sum_{i,j} p_i^t x_j^t W_{ij}.$$

Taking the logarithm of either growth factor gives the corresponding per-generation growth rate. Thus, $\log\langle W_i\rangle_{\boldsymbol{x}^t}$ is the per-generation growth rate of type i, while $\log\langle W\rangle_{\boldsymbol{p}^t,\boldsymbol{x}^t}$ is the per-generation growth rate of the total population. The discrete-time replicator dynamics in this variable-environment model are therefore given by

$$p_i^{t+1} = \frac{p_i^t\langle W_i\rangle_{\boldsymbol{x}^t}}{\langle W\rangle_{\boldsymbol{p}^t,\boldsymbol{x}^t}}. \tag{A18}$$

Continuous-time replicator dynamics in variable environments:

Let r_{ij} denote the Malthusian fitness, or instantaneous per-capita growth rate, of type i in environmental condition j. The effective Malthusian fitness of type i under the condition distribution $\boldsymbol{x}^t$ is

$$\langle r_i\rangle_{\boldsymbol{x}^t} := \sum_j x_j^t r_{ij},$$

and the mean Malthusian fitness, equivalently the instantaneous per-capita growth rate of the total population, is

$$\langle r \rangle_{\boldsymbol{p}^t, \boldsymbol{x}^t} := \sum_{i,j} p_i^t x_j^t r_{ij}.$$

The continuous-time replicator dynamics in this variable-environment model are therefore given by

$$\frac{dp_i^t}{dt} = p_i^t \left(\langle r_i \rangle_{\boldsymbol{x}^t} - \langle r \rangle_{\boldsymbol{p}^t, \boldsymbol{x}^t} \right) . \quad [A19]$$

Appendix B: Natural Selection as an Information Acquisition Process

B.1 Information Measures

Information theory uses probability theory and statistics to formalize the familiar notion that information is that which reduces uncertainty. The more an observation reduces one's uncertainty about the state of some system, the more information that observation provides. Information-theoretic measures allow us to rigorously quantify both uncertainty and information in a system.

Entropy quantifies uncertainty

Consider a random variable X that can take some value $i \in \mathcal{X}$ (e.g., the outcome of a die roll), and suppose you assign a probability p_i to each outcome value based on your belief about how likely it is to occur. If you are fairly certain that one of the outcomes will occur, you will ascribe high probability to that outcome and low probabilities to all other outcomes. If you are more uncertain about which outcome will occur, your probability distribution $\boldsymbol{p}$ will be more uniform. This notion of uncertainty is quantified by the **entropy**:

$$H(X) = -\sum_{i \in \mathcal{X}} p_i \log p_i = \mathbb{E}_{i \sim \boldsymbol{p}} \left[-\log p_i \right] = \mathbb{E}_{i \sim \boldsymbol{p}} \left[\log \frac{1}{p_i} \right] . \quad [B1]$$

Entropy is a property of the spread of the distribution $\boldsymbol{p}$ and does not depend on the values of the random variable X. As such, the entropy of $\boldsymbol{p}$ may also be denoted $H(\boldsymbol{p})$.

Entropy can be interpreted as the expected **surprisal** of an outcome, where the surprisal of the outcome i under the probability distribution $\boldsymbol{p}$ is defined as $\log(1/p_i)$. If an outcome that was ascribed a high probability of occurring does occur, this gives low surprisal ($\log(1/p_i) \to 0$ as $p_i \to 1$), but the occurrence of an outcome that was ascribed low probability gives high surprisal ($\log(1/p_i) \to \infty$ as $p_i \to 0$). The more uncertain you are about the outcome — the more uniform your probability distribution — the more surprising observed outcomes are on average, and the higher the entropy.

The **joint entropy** $H(X, Y)$ of two discrete random variables X and Y is the entropy of their joint probability distribution

$$H(X, Y) = -\sum_{i \in \mathcal{X}, j \in \mathcal{Y}} p(i, j) \log p(i, j) . \quad [B2]$$

The **conditional entropy** $H(X|Y)$ measures the expected amount of uncertainty about X that remains when the value of Y is known, averaged over possible values of Y

$$H(X|Y) = \sum_{j \in \mathcal{Y}} p(j) H(X|Y = j) = -\sum_{j \in \mathcal{Y}} p(j) \sum_{i \in \mathcal{X}} p(i|j) \log p(i|j). \quad [B3]$$

Reflecting the relationship between information and uncertainty, other information measures can be expressed as linear combinations of entropies.

Relative entropy (KL divergence) quantifies information gain

Consider a random variable X that is hypothesized to follow some probability distribution $\boldsymbol{q}$. Now suppose that you receive new evidence that leads you to update your hypothesis to some new probability distribution $\boldsymbol{p}$. How much information did this update provide? **Information gain** quantifies the reduction in expected surprisal obtained by replacing the prior distribution $\boldsymbol{q}$ with the posterior distribution $\boldsymbol{p}$:

$$D(\boldsymbol{p} \| \boldsymbol{q}) = H_{\boldsymbol{p}}(\boldsymbol{q}) - H_{\boldsymbol{p}}(\boldsymbol{p}) = \mathbb{E}_{i \sim \boldsymbol{p}} \left[\log \frac{1}{q_i} \right] - \mathbb{E}_{i \sim \boldsymbol{p}} \left[\log \frac{1}{p_i} \right] . \quad [B4]$$

Here both expectations of surprisal are calculated under the posterior distribution $\boldsymbol{p}$, the latest working hypothesis of the true event probabilities. The expected surprisal under the prior is the cross-entropy $H_{\boldsymbol{p}}(\boldsymbol{q})$, whereas the expected surprisal under the posterior is the entropy $H_{\boldsymbol{p}}(\boldsymbol{p})$. Information gain $D(\boldsymbol{p} \| \boldsymbol{q})$ is also referred to as the **relative entropy** or **Kullback-Leibler divergence** between distributions $\boldsymbol{p}$ and $\boldsymbol{q}$.

Information gain (i.e., relative entropy, KL divergence) is always nonnegative and is zero if and only if $\boldsymbol{p} = \boldsymbol{q}$. It is often useful to think of this measure as the "distance" between distributions, but the KL divergence is not a true distance measure as it is asymmetric ($D(\boldsymbol{p} \| \boldsymbol{q}) \neq D(\boldsymbol{q} \| \boldsymbol{p})$) and does not satisfy the triangle inequality.

Mutual information quantifies non-independence

Often we are interested in quantifying how informative an observation is about the state of a system. In other words, how much is our uncertainty about the state of a system X reduced by observing the state of another variable Y? Of course, the observation Y can only carry information about the system X if the states of X and Y are somehow dependent on each other. The more dependent X and Y are on each other, the more knowledge of Y specifies the state of X (and vice versa).

The **mutual information** $MI(X;Y)$ of random variables X and Y quantifies precisely this notion of information as the reduction in entropy (i.e., uncertainty) of one random variable given observation of another random variable

$$MI(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)\,. \tag{B5}$$

Mutual information can be equivalently expressed as the divergence from independence of two random variables. When X and Y are independent random variables, their joint distribution is equal to the product of their marginal distributions. Mutual information is equivalent to the relative entropy (KL divergence) between the actual joint distribution $\boldsymbol{p}_{X,Y}$ and the hypothetical product distribution $\boldsymbol{p}_X \boldsymbol{p}_Y$ that assumes independence:

$$MI(X;Y) = D(\boldsymbol{p}_{X,Y} \,\|\, \boldsymbol{p}_X \boldsymbol{p}_Y) = \sum_{i \in \mathcal{X}, j \in \mathcal{Y}} p_{X,Y}(i,j) \log \frac{p_{X,Y}(i,j)}{p_X(i) p_Y(j)}\,. \tag{B6}$$

Mutual information can be interpreted in terms of relative entropy as the amount of information we gain by updating our posterior to reflect some dependence between X and Y from a prior that assumed independence.

Bits as units of information

When the base of the logarithms is 2, entropy and corresponding information measures are measured in *bits*. For an interpretation of the bit as a unit of measure, consider the following example (adapted from Cover and Thomas (2006)). Let X be a random variable that can take some value $i \in \{a, b, c, d\}$ with probability distribution

$$p_i = \begin{cases} 1/2, & i = a, \\ 1/4, & i = b, \\ 1/8, & i = c, \\ 1/8, & i = d. \end{cases}$$

Suppose X takes a random value and we wish to determine this value with as few binary ("yes-no") questions as possible. The most efficient approach is to ask the question whose two outcomes are closest to equiprobable at each stage. Doing so either identifies the value of X or reduces the space of possible outcomes by about half. In this example, one should start by asking, "*Is* $X = a$?" If yes, X has been determined with 1 question; if no, follow up by asking "*Is* $X = b$?", and so on. X is a random variable, so the number of binary questions necessary to determine its value is also a random variable. The minimum expected number of binary questions to determine X in this example is

$$\frac{1}{2}(1 \text{ question}) + \frac{1}{4}(2 \text{ questions}) + \frac{1}{8}(3 \text{ questions}) + \frac{1}{8}(3 \text{ questions}) = 1.75 \text{ questions}\,, \tag{B7}$$

which is equal to the entropy of X

$$H(X) = -\sum_{i \in \mathcal{X}} p_i \log_2 p_i = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) = 1.75 \text{ bits}\,. \tag{B8}$$

The minimum expected number of binary questions required to determine X is equal to the entropy $H(X)$ for dyadic probability distributions (i.e., where all probabilities are powers of 2—such as the one in this example) and lies between $H(X)$ and $H(X) + 1$ in general.

Therefore, bits can be interpreted as the expected number of binary questions necessary to determine the value of a random variable. Bits of *entropy* $H(\boldsymbol{p})$ represent uncertainty in terms of the average number of questions that must be asked to arrive at certainty about a random variable X that takes values according to $\boldsymbol{p}$. Bits of *relative entropy* $D(\boldsymbol{p} \,\|\, \boldsymbol{q})$ represent a relative excess in uncertainty in terms of the expected number of extra questions that must be asked when choosing questions based on some distribution $\boldsymbol{q}$ instead of the true underlying distribution $\boldsymbol{p}$. Bits of *mutual information* $MI(X;Y)$ represent reduction in uncertainty in terms of how many questions about X are saved on average when given some other observation Y. In all cases, 1 bit refers to 1 binary question that must be asked on average to arrive at certainty. *In other words, 1 bit corresponds to 1 two-fold difference in uncertainty.*

B.2 Natural selection acquires information about the environment

B.2.1 Information gain

In the context of natural selection, the distribution of type frequencies $\boldsymbol{p}^t$ can be seen as the population's hypothesis about which types are best suited to the selective environments it has experienced. The relative reproductive success of each type provides evidence about the suitability of each type, and replicator dynamics updates the population's hypothesis such that highly fit types are given more weight, and poorly fit types are given less. This process encodes information about the observed fitnesses of types into the population's type composition. We can measure how much information is gained in these updates by measuring how much the population's type frequency distribution changes during this process. The relative entropy (Kullback–Leibler divergence) of the population's updated type distribution $\boldsymbol{p}^T$ from its initial distribution $\boldsymbol{p}^0$ quantifies the amount of **information gain** I^T for a period of selection of duration T

$$I^T = D(\boldsymbol{p}^T \,\|\, \boldsymbol{p}^0) = \sum_i p_i^T \log \frac{p_i^T}{p_i^0}\,. \tag{B9}$$

In Bayesian terms, $D(\boldsymbol{p}^T \,\|\, \boldsymbol{p}^0)$ is a measure of the information gained by revising the population's hypothesis from the prior distribution $\boldsymbol{p}^0$ to the posterior distribution $\boldsymbol{p}^T$.

Specific information

A closely related measure is the **specific information** introduced by Adami and Cerf (2000), which quantifies the entropy removed from a maximally unconstrained ensemble by a particular selective environment. When the initial population is uniform, $\boldsymbol{p}^0 = \boldsymbol{u}$ with $u_i = 1/n$, the information gain becomes

$$\begin{aligned} I^T &= D(\boldsymbol{p}^T \,\|\, \boldsymbol{u}) && \text{[B10]}\\ &= \sum_i p_i^T \log p_i^T + \log n && \text{[B11]}\\ &= \log n - H(\boldsymbol{p}^T) && \text{[B12]}\\ &= H(\boldsymbol{p}^0) - H(\boldsymbol{p}^T). && \text{[B13]} \end{aligned}$$

Thus, for a uniform initial distribution, I^T reduces to the specific information. Adami and Cerf (2000) further connect this population-level quantity to algorithmic information by relating specific information to the ensemble-averaged mutual Kolmogorov complexity between genetic sequences and their environment.

Specific information describes the total information encoded at time T relative to a maximally unconstrained baseline, whereas I^T measures the information acquired over the specified interval $[0, T]$ relative to the population's actual initial distribution, which may already be nonuniform. In the general case, I^T therefore need not equal $H(\boldsymbol{p}^0) - H(\boldsymbol{p}^T)$, but it retains a corresponding interpretation as the expected reduction in the surprisal of types observed at time T when they are described using the updated distribution $\boldsymbol{p}^T$ rather than the ancestral distribution $\boldsymbol{p}^0$:

$$I^T = H_{\boldsymbol{p}^T}(\boldsymbol{p}^0) - H_{\boldsymbol{p}^T}(\boldsymbol{p}^T) = -\sum_i p_i^T \log p_i^0 + \sum_i p_i^T \log p_i^T\,.$$

In this sense, relative entropy extends the intuition of specific information from a uniform baseline to arbitrary initial distributions and arbitrary selection intervals.

Information gain is associated with fitness increase

Information gain quantifies the extent to which selection reweights a population toward higher-fitness types. The same reweighting changes the population's per-generation rate, $\log\langle W\rangle_{\boldsymbol{p}^t}$. Here we relate these quantities by showing that information gain bounds the increase in the population's per-generation rate, minus an endpoint Jensen gap.

Proposition B1. *If the environment consists of a single unchanging condition ($m = 1$) and all types present in the initial population have fixed positive fitnesses W_i, then, for any $T \geq 1$, the total information gain I^T of the trajectory of type distributions $\boldsymbol{p}^0, \ldots, \boldsymbol{p}^T$ generated by discrete-time replicator dynamics is bounded from below by the increase in the population's per-generation rate, minus the endpoint Jensen gap J^T:*

$$I^T \geq \left(\log\langle W\rangle_{\boldsymbol{p}^T} - \log\langle W\rangle_{\boldsymbol{p}^0}\right) - J^T. \tag{B14}$$

Here $J^T := \log\langle W\rangle_{\boldsymbol{p}^T} - \langle \log W\rangle_{\boldsymbol{p}^T} \geq 0$ is the endpoint Jensen gap, where $\langle W\rangle_{\boldsymbol{p}^t} := \sum_i p_i^t W_i$ and $\langle \log W\rangle_{\boldsymbol{p}^t} := \sum_i p_i^t \log W_i$.

***Proof*:** After T generations of discrete-time, fixed-fitness replicator dynamics,

$$p_i^T = p_i^0 \prod_{t=0}^{T-1} \frac{W_i}{\langle W \rangle_{\boldsymbol{p}^t}}. \tag{B15}$$

Substituting this expression into the definition of information gain gives

$$\begin{aligned} I^T &= \sum_i p_i^T \log \frac{p_i^T}{p_i^0} \\ &= T\langle \log W \rangle_{\boldsymbol{p}^T} - \sum_{t=0}^{T-1} \log \langle W \rangle_{\boldsymbol{p}^t} \\ &= \sum_{t=0}^{T-1} \left[\langle \log W \rangle_{\boldsymbol{p}^T} - \log \langle W \rangle_{\boldsymbol{p}^t} \right]. \end{aligned} \tag{B16}$$

With type fitnesses held fixed, population mean fitness is nondecreasing under selection, as in the standard fixed-environment, haploid form of Fisher's fundamental theorem of natural selection (Fisher 1930, Adami 2024). The same update also makes the population mean of the type growth rates, $\langle \log W \rangle_{\boldsymbol{p}^t}$, nondecreasing, since

$$\langle \log W \rangle_{\boldsymbol{p}^{t+1}} - \langle \log W \rangle_{\boldsymbol{p}^t} = \frac{\mathrm{Cov}_{\boldsymbol{p}^t}(W, \log W)}{\langle W \rangle_{\boldsymbol{p}^t}} \geq 0, \tag{B17}$$

because $\log W$ is an increasing function of W. Consequently,

$$\langle \log W \rangle_{\boldsymbol{p}^T} \geq \langle \log W \rangle_{\boldsymbol{p}^{t+1}} \qquad \text{for every } t \in \{0, \ldots, T-1\}.$$

Moreover, the replicator update gives

$$\begin{aligned} \langle \log W \rangle_{\boldsymbol{p}^{t+1}} - \log \langle W \rangle_{\boldsymbol{p}^t} &= \sum_i p_i^{t+1} \log \frac{p_i^{t+1}}{p_i^t} \\ &= D\big(\boldsymbol{p}^{t+1} \big\| \boldsymbol{p}^t\big) \geq 0. \end{aligned} \tag{B18}$$

It follows that every summand in Equation B16 is nonnegative. Separating the $t = 0$ term therefore yields

$$\begin{aligned} I^T &= \langle \log W \rangle_{\boldsymbol{p}^T} - \log \langle W \rangle_{\boldsymbol{p}^0} + \sum_{t=1}^{T-1} \left[\langle \log W \rangle_{\boldsymbol{p}^T} - \log \langle W \rangle_{\boldsymbol{p}^t} \right] \\ &\geq \langle \log W \rangle_{\boldsymbol{p}^T} - \log \langle W \rangle_{\boldsymbol{p}^0}. \end{aligned} \tag{B19}$$

Finally, using

$$\langle \log W \rangle_{\boldsymbol{p}^T} = \log \langle W \rangle_{\boldsymbol{p}^T} - J^T$$

gives

$$I^T \geq \left(\log \langle W \rangle_{\boldsymbol{p}^T} - \log \langle W \rangle_{\boldsymbol{p}^0} \right) - J^T,$$

which completes the proof. □

Corollary B1.1. *If the environment comprises multiple conditions whose exogenous distribution remains fixed ($\boldsymbol{x}^t \equiv \boldsymbol{x}$ for all t) and all types present in the initial population have fixed positive effective fitnesses $\langle W_i \rangle_{\boldsymbol{x}}$, then, for any $T \geq 1$, the total information gain I^T of the trajectory of type distributions $\boldsymbol{p}^0, \ldots, \boldsymbol{p}^T$ generated by discrete-time replicator dynamics is bounded from below by the increase in the population's per-generation rate, minus the endpoint Jensen gap $J_{\boldsymbol{x}}^T$:*

$$I^T \geq \left(\log \langle W \rangle_{\boldsymbol{p}^T, \boldsymbol{x}} - \log \langle W \rangle_{\boldsymbol{p}^0, \boldsymbol{x}} \right) - J_{\boldsymbol{x}}^T. \tag{B20}$$

Here $J_{\boldsymbol{x}}^T := \log \langle W \rangle_{\boldsymbol{p}^T, \boldsymbol{x}} - \langle \log \langle W_i \rangle_{\boldsymbol{x}} \rangle_{\boldsymbol{p}^T} \geq 0$ is the endpoint Jensen gap, where $\langle W_i \rangle_{\boldsymbol{x}} := \sum_j x_j W_{ij}$ and $\langle W \rangle_{\boldsymbol{p}^T, \boldsymbol{x}} := \sum_i p_i^T \langle W_i \rangle_{\boldsymbol{x}} = \sum_{i,j} p_i^T x_j W_{ij}$.

Proof. When $\boldsymbol{x}^t \equiv \boldsymbol{x}$, the discrete-time replicator update depends on W_{ij} only through the fixed effective fitnesses. The dynamics are therefore identical to those in Proposition B1, with W_i replaced by $\langle W_i \rangle_{\boldsymbol{x}}$. Applying that proposition gives the stated result. □

Corollary B1.2. *Under the conditions of Proposition B1, the endpoint Jensen gap J^T is zero if and only if all types in the support of $\boldsymbol{p}^T$ have the same fitness. Whenever this condition holds, the total information gain I^T is bounded from below by the increase in the population's per-generation rate:*

$$I^T \geq \log\langle W\rangle_{\boldsymbol{p}^T} - \log\langle W\rangle_{\boldsymbol{p}^0}. \tag{B21}$$

In particular, $J^T \to 0$ for any trajectory converging to a distribution supported entirely on types of equal fitness, including a trajectory converging to fixation of a single type.

Under the conditions of Corollary B1.1, the analogous conclusions hold when all types in the support of $\boldsymbol{p}^T$ have the same effective fitness $\langle W_i\rangle_{\boldsymbol{x}}$. Whenever this condition holds,

$$I^T \geq \log\langle W\rangle_{\boldsymbol{p}^T,\boldsymbol{x}} - \log\langle W\rangle_{\boldsymbol{p}^0,\boldsymbol{x}}. \tag{B22}$$

Likewise, $J^T_{\boldsymbol{x}} \to 0$ along any trajectory converging to a distribution supported entirely on types of equal effective fitness.

Proof. The strict concavity of log gives the stated conditions under which the Jensen gaps vanish. The bounds then follow from Proposition B1 and Corollary B1.1, and the limiting statements follow by continuity. □

B.2.2 Potential information

Relative entropy (KL divergence) quantifies the difference between two probability distributions and can be interpreted as the information gained in updating from one distribution (e.g., a prior) to another (e.g., a posterior). We have defined the relative entropy $I^t := D(\boldsymbol{p}^t \,\|\, \boldsymbol{p}^0)$ as the information that a population has gained by natural selection up to time t (Appendix B.2.1). Here we consider the relative entropy of some fixed reference distribution $\overline{\boldsymbol{p}}$ from the population's current frequency distribution $\boldsymbol{p}^t$:

$$D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t) = \sum_i \overline{p}_i \log\frac{\overline{p}_i}{p_i^t}. \tag{B23}$$

This relative entropy represents the amount of information that would be gained if the population's type frequency distribution were to change from $\boldsymbol{p}^t$ to $\overline{\boldsymbol{p}}$. We refer to this quantity as the **potential information** of the system: the amount of information that is available to be gained in moving to $\overline{\boldsymbol{p}}$ from the population's current state.

We would like to determine the conditions under which a population undergoing natural selection can be said to be learning. One way to formalize this question is to ask when the potential information $D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t)$ decreases. Let $\overline{\boldsymbol{p}}$ denote a fixed reference composition. When this composition represents a biologically meaningful endpoint of selection (e.g., a locally attracting stationary state), the potential information can be interpreted as the information that remains to be acquired in moving from the population's current composition to that state. We assume that every type represented in $\overline{\boldsymbol{p}}$ is also represented in $\boldsymbol{p}^0$, so that the potential information is finite along the trajectory.

Under continuous-time replicator dynamics with fixed Malthusian fitnesses, the rate of change of the potential information is

$$\frac{d}{dt}D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t) = -\sum_i \overline{p}_i \frac{1}{p_i^t}\frac{dp_i^t}{dt} \tag{B24}$$

$$= -\sum_i \overline{p}_i \left(r_i - \langle r\rangle_{\boldsymbol{p}^t}\right) \qquad \text{using } \frac{d}{dt}p_i^t = p_i^t\left(r_i - \langle r\rangle_{\boldsymbol{p}^t}\right) \tag{B25}$$

$$= \langle r\rangle_{\boldsymbol{p}^t} - \langle r\rangle_{\overline{\boldsymbol{p}}}. \tag{B26}$$

Thus, potential information decreases whenever the fixed reference composition has a greater mean growth rate than the population's current composition:

$$\langle r\rangle_{\overline{\boldsymbol{p}}} > \langle r\rangle_{\boldsymbol{p}^t} \quad\Longrightarrow\quad \frac{d}{dt}D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t) < 0. \tag{B27}$$

When this condition holds, the population is moving closer to the reference composition in this information-theoretic sense.

More generally, for any trajectory that converges to an endpoint composition $\overline{\boldsymbol{p}}$,

$$D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t) \longrightarrow 0, \tag{B28}$$

$$I^t = D(\boldsymbol{p}^t\|\boldsymbol{p}^0) \longrightarrow D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^0). \tag{B29}$$

Thus, the initial potential information equals the limiting information gained as the population approaches the endpoint composition. This is the full information associated with the compositional change from $\boldsymbol{p}^0$ to $\overline{\boldsymbol{p}}$ in that selective context.

Potential information also has a more specific stability interpretation in the broader class of frequency-dependent replicator models. Relative to an evolutionarily stable state $\overline{\boldsymbol{p}}$, the divergence $D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t)$ is a strict local Lyapunov function for continuous-time replicator dynamics. For an interior reference state, the converse also holds: $\overline{\boldsymbol{p}}$ is an evolutionarily stable state if and only if $D(\overline{\boldsymbol{p}}\|\boldsymbol{p})$ is a strict local Lyapunov function (Harper 2009a). Evolutionarily stable states are therefore one important class of attracting stationary states, but attracting stationary states need not be evolutionarily stable states in systems with three or more types. The general endpoint result above is consequently framed in terms of convergence rather than evolutionary stability.

B.2.3 Natural selection establishes mutual information between types and environmental histories

The sense in which a population contains information *about* its environment can be understood through a simple thought experiment. Imagine drawing an environmental history from some set of possible histories and then randomly drawing a type from a population that experienced that history. If different histories tend to produce characteristically different type distributions, then observing the sampled type can change the probabilities that we ascribe to those histories. Averaged over the possible types that might be sampled, such observations reduce our uncertainty about which history occurred. By contrast, if types occur with the same probabilities following every history, then observing a sampled type tells us nothing about which history the population experienced. In this sense, a population's type composition contains information about its environmental history.

Natural selection can establish such an association between types and environmental histories. Environmental conditions determine the relative fitnesses of types, causing types associated with high relative fitness under a particular history to increase in frequency. When different histories favor types differently, the resulting populations evolve in distinct, environment-dependent ways. Thus, natural selection can accumulate information about environmental history in the association between histories and the types that are likely to be observed following them.

To formalize this idea, let E be a random environmental history drawn according to $P(E)$, and let G be the type of an individual sampled at time T from a population that experienced that history. Here a specific environmental history $e = (\boldsymbol{x}^0, \ldots, \boldsymbol{x}^{T-1})$ is a sequence of distributions over environmental conditions (not to be confused with an individual environmental condition j represented within $\boldsymbol{x}^t$). Assume that all populations begin from a common initial type distribution $\boldsymbol{p}^0$, independently of their subsequent histories, and that the exogenous distribution $P(E)$ remains fixed. For the population experiencing history e, let $\boldsymbol{p}_e^T$ denote its type distribution at time T. Then

$$P^T(E = e) = P(e), \tag{B30}$$

$$P^T(G = i \mid E = e) = p_{e,i}^T, \tag{B31}$$

$$P^T(G = i, E = e) = P(e) p_{e,i}^T, \tag{B32}$$

$$P^T(G = i) = \langle p_{e,i}^T \rangle_E := \sum_e P(e) p_{e,i}^T, \tag{B33}$$

where $\langle \cdot \rangle_E$ denotes an average over the distribution of environmental histories.

The mutual information between types and environmental histories quantifies precisely the reduction in uncertainty (entropy) arising from their statistical association:

$$MI^T(G; E) := H(E) - H(E \mid G) \tag{B34}$$

$$= D\big(P^T(G, E) \| P^T(G) P(E)\big) \tag{B35}$$

$$= \big\langle D\big(\boldsymbol{p}_e^T \| \langle \boldsymbol{p}_e^T \rangle_E\big) \big\rangle_E . \tag{B36}$$

Thus, $MI^T(G; E)$ measures how much observing a type at time T reduces uncertainty about the environmental history experienced by its population.

For a specific history e, the information gain

$$I_e^T := D(\boldsymbol{p}_e^T \| \boldsymbol{p}^0) \tag{B37}$$

measures how much the type distribution of that population has changed from its common initial state. The mutual information across histories can be decomposed in terms of these individual information gains (McGee 2021):

$$MI^T(G; E) = \big\langle D(\boldsymbol{p}_e^T \| \boldsymbol{p}^0) \big\rangle_E - D\big(\langle \boldsymbol{p}_e^T \rangle_E \| \boldsymbol{p}^0\big) \tag{B38}$$

$$= \langle I_e^T \rangle_E - D\big(\langle \boldsymbol{p}_e^T \rangle_E \| \boldsymbol{p}^0\big) . \tag{B39}$$

The first term is the average information gain across environmental histories. The second term subtracts the change in the type distribution that remains after pooling across histories (roughly, the change that is common across histories), because this pooled change provides no information about which particular history occurred. Thus, the mutual information refers to the portion of evolutionary change that is specific to particular environmental histories.

For example, selection might shift every population away from $\boldsymbol{p}^0$ in exactly the same way. Each population would then have positive information gain I_e^T, but the endpoint distributions $\boldsymbol{p}_e^T$ would be identical across histories and $MI^T(G; E)$ would be zero. Adaptive information accumulates only when populations experiencing alternative environmental histories evolve in systematically different, environment-dependent ways.

Because selection acts through environment-dependent differences in relative fitness, it can systematically establish this association. By contrast, processes such as drift, mutation, or migration whose probabilities are independent of environmental history do not by themselves encode which history occurred when averaged over their randomness, although they may modify an association already

established by selection. This work focuses on infinite, deterministic, haploid populations in which selection is the only process changing type frequencies.

Accordingly, $MI^T(G;E)$ measures the adaptive information present at time T as a result of environment-dependent evolutionary change. Note that under arbitrary fluctuating histories, this information need not increase monotonically at every intermediate time, because later selection can weaken associations established earlier.

Appendix C: Natural Selection as a Learning Process

C.1 Natural selection and Bayesian learning

Suppose that a learner wishes to infer the true state of a system, which may take any one of n possible states. The learner's uncertainty is represented using a probability distribution that assigns a weight $\Pr(h_i)$ to each alternative hypothesis. When new evidence $\mathcal{D}$ is obtained, alternatives that give relatively high likelihood $\Pr(\mathcal{D} \mid h_i)$ to the observed evidence are up-weighted in the posterior distribution. Bayesian learning refers to this process of iteratively updating beliefs in light of new evidence using Bayes' rule:

$$\Pr(h_i \mid \mathcal{D}) = \frac{\Pr(\mathcal{D} \mid h_i)\Pr(h_i)}{P(\mathcal{D})}. \qquad \text{[C1]}$$

In the context of natural selection, each type in a population can be thought of as an alternative strategy for survival and reproduction, and the distribution of type frequencies $\boldsymbol{p}$ can be viewed as the population's current weighting of those alternatives. Like Bayesian learning, natural selection updates these weights in response to information supplied by the environment. As selection changes the frequencies of types according to their relative fitnesses, the population shifts toward types that have produced organisms relatively well adapted to the conditions experienced.

Indeed, there is a formal connection between Bayesian learning and replicator dynamics. The two update rules share the same form: selection increases the frequencies of types with high relative fitness in the same way that Bayes' rule increases the weights of hypotheses that assign high relative likelihood to the observed evidence:

Bayesian learning:	Replicator dynamics:
$\Pr^{t+1}(h_i) = \frac{\Pr(\mathcal{D}^t \mid h_i)\Pr^t(h_i)}{\sum_k \Pr(\mathcal{D}^t \mid h_k)\Pr^t(h_k)}$	$p_i^{t+1} = \frac{W_i^t p_i^t}{\sum_k W_k^t p_k^t}.$ [C2]

Natural selection can thus be understood as a learning process that "infers" which types are best adapted to the environment. A number of studies have formalized selection and other evolutionary processes in terms of Bayesian computation and learning (Harper 2009b, Shalizi 2009, Campbell 2016, Watson and Szathmáry 2016, Czégel et al. 2022). In fact, Bayesian updating is a special case of discrete replicator dynamics in which the "fitness" assigned to each alternative is given by its likelihood, allowing many results for replicator dynamics to be applied to Bayesian analysis more generally (Shalizi 2009).

C.2 Natural selection as an online learning process

C.2.1 The online learning problem

In the field of computational learning theory, *online learning* refers to a process of iteratively updating a strategy for responding to problems given feedback about the quality of past responses. That is, a learner is presented with a sequence of problems to which they must provide a response (e.g., an answer to a question or a prediction about the state of a system), and the learner holds a strategy that is used to generate their response to each problem in turn. After providing a response to each problem, the learner receives feedback about the quality of their response via a *loss* function that measures the discrepancy between the learner's response and the optimal response for that problem. The learner's goal is to minimize the cumulative loss that they suffer in the long run, which they may achieve by updating their strategy after each round such that their responses are more accurate in subsequent rounds. The problem of adapting a strategy for sequential prediction has been studied in a number of fields, including machine learning, game theory, and information theory.

We first formalize the online learning problem in the general framework of repeated games against Nature. In the following section, we interpret this formalism for the context of an evolving population (Appendix C.2.2). The repeated games framework is closely related to the prediction with expert advice framework that has been widely studied in computer science (Freund and Schapire 1999, Cesa-Bianchi and Lugosi 2006).

Suppose a learner has a set of n alternative pure strategies. A *pure strategy* for the learner consists of selecting a single alternative i, while a *mixed strategy* $\boldsymbol{p}^t$ is a probability distribution that assigns weight p_i^t to each alternative. In each round t, the learner selects a mixed strategy $\boldsymbol{p}^t$ and the environment generates a loss vector:

$$\left(\ell_1^t, \ell_2^t, \ldots, \ell_n^t\right). \qquad \text{[C3]}$$

The element ℓ_i^t gives the loss assigned to pure strategy i in round t. The loss vector is revealed after the learner chooses their mixed strategy, and the learner uses this feedback when choosing their strategy in the next round. The sequence of loss vectors may be fixed or generated deterministically, stochastically, arbitrarily, or even adversarially.

Let $\ell^t(\boldsymbol{p})$ denote the one-round loss assigned to a mixed strategy $\boldsymbol{p}$ under a specified rule for aggregating the losses of its constituent pure strategies. We require that this rule agree with the pure-strategy losses (i.e., $\ell^t(\boldsymbol{\delta}_i) = \ell_i^t$, where $\boldsymbol{\delta}_i$ is the point mass at i), but more than one aggregation rule may satisfy this condition and provide a meaningful evaluation of a mixed strategy. We leave the rule unspecified for now and distinguish two fitness-based formulations relevant to this work in Appendix C.2.5.

After T rounds, the learner's cumulative loss is

$$L^T := \sum_{t=0}^{T-1} \ell^t(\boldsymbol{p}^t)\,. \tag{C4}$$

Even the best available strategy may receive non-zero loss in each round, resulting in a cumulative loss that continues to increase over time. For example, an adversarial environment can make the learner's cumulative loss arbitrarily large by repeatedly generating loss vectors that are unfavorable to the learner's choices of $\boldsymbol{p}^t$. For this reason, it is often more informative to evaluate an online learning algorithm in terms of **regret**, which measures the difference between the learner's cumulative loss and that of a competing fixed strategy $\boldsymbol{q}$ over the same sequence of loss vectors generated by the environment:

$$R_{\boldsymbol{q}}^T := L^T - L_{\boldsymbol{q}}^T = \sum_{t=0}^{T-1} \left(\ell^t(\boldsymbol{p}^t) - \ell^t(\boldsymbol{q})\right)\,. \tag{C5}$$

Regret is commonly evaluated with respect to an optimal fixed strategy in hindsight:

$$\boldsymbol{q}_{\text{opt}}^T \in \arg\min_{\boldsymbol{q}\in\Delta_n} L_{\boldsymbol{q}}^T\,. \tag{C6}$$

The identity of this optimal reference strategy may depend on the rule used to evaluate the loss of a mixed strategy.

Regret analysis evaluates the ability of a learning algorithm to perform nearly as well as a favorable fixed strategy despite not knowing the future sequence of losses. A learning process is said to be *no-regret* with respect to $\boldsymbol{q}$ if it updates the learner's mixed strategy such that it incurs no positive asymptotic per-round regret:

$$\limsup_{T\to\infty} \frac{1}{T} R_{\boldsymbol{q}}^T \leq 0\,. \tag{C7}$$

The definitions above apply generically across rules $\ell^t(\boldsymbol{p})$ for evaluating mixed strategies. We distinguish two fitness-based formulations in Appendix C.2.5 after establishing the equivalence between natural selection and MWU. Appendix D returns to cumulative loss and regret in the biological setting.

C.2.2 The Population versus Environment game

The online learning formalism described in Appendix C.2.1 can be interpreted as the basis for a model of natural selection in variable environments.* We consider repeated play of a two-player game between the population† (acting as the learner) and the environment. The set of pure strategies available to the population player (i.e., rows of the game matrix) represents the set of replicator types in the population, and the population's mixed strategy $\boldsymbol{p}^t$ represents the type frequency distribution at time t. The set of pure strategies available to the environment player (i.e., columns of the game matrix) represents the set of possible environmental conditions. The environment player's mixed strategy $\boldsymbol{x}^t$ gives the probability distribution of conditions experienced by individuals in round t.

We assume that every individual in the population experiences an independent environmental condition, as if each individual occupies a separate micro-environment. A play of the game in round t can therefore be interpreted as the population assigning a type to each individual with frequencies $\boldsymbol{p}^t$ and the environment assigning a condition to each micro-environment with probabilities $\boldsymbol{x}^t$. The association of particular individuals with particular conditions is random and outside the control of both players.

The outcome of each play for the population is specified by an n-by-m fitness (payoff) matrix $\mathbf{W}$. Each element $W_{ij} > 0$ gives the Wrightian fitness of an individual of type i that experiences condition j. The environment player's own payoffs or losses are left unspecified. The environment's play $\boldsymbol{x}^t$ induces a vector of condition-averaged type fitnesses, giving the one-generation growth factor of each type:

$$\mathbf{W}\boldsymbol{x}^t = \left(\langle W_1\rangle_{\boldsymbol{x}^t}, \langle W_2\rangle_{\boldsymbol{x}^t}, \ldots, \langle W_n\rangle_{\boldsymbol{x}^t}\right)^{\mathsf{T}}\,. \tag{C8}$$

The resulting mean fitness of a population with composition $\boldsymbol{p}^t$ is

$$\langle W\rangle_{\boldsymbol{p}^t,\boldsymbol{x}^t} = \left(\boldsymbol{p}^t\right)^{\mathsf{T}} \mathbf{W}\boldsymbol{x}^t = \sum_{i,j} p_i^t x_j^t W_{ij}\,, \tag{C9}$$

*The Population versus Environment game model is wholly consistent with the models of replicator dynamics in variable environments outlined in Appendix A.2.

†This differs from standard evolutionary game theory in which each individual within the population is a separate player.

which is the population's one-generation growth factor.

Natural selection uses the vector of type growth factors, $\mathbf{W}\boldsymbol{x}^t$, as feedback to update the population's mixed strategy:

$$p_i^{t+1} = \frac{p_i^t \langle W_i \rangle_{\boldsymbol{x}^t}}{\sum_z p_z^t \langle W_z \rangle_{\boldsymbol{x}^t}} . \quad \text{[C10]}$$

In the language of online learning, the environment's play thus generates a round-specific vector of fitness feedback from which a corresponding vector of type losses can be defined. A precise transformation from fitness to loss that makes replicator dynamics (Equation C10) an instance of a well-studied learning algorithm known as Multiplicative Weights Updating is developed in Appendix C.2.4.

We again take a very expansive view of what constitutes an "environment." Environmental conditions may represent abiotic conditions experienced by individuals, the types of other replicators encountered by individuals, or even the identities of other alleles carried by individuals. Similarly, the environment player may select the sequence $\{\boldsymbol{x}^0, \boldsymbol{x}^1, \ldots, \boldsymbol{x}^T\}$ according to any process, including as a function of time, as a function of the population player's strategy, as a stochastic process, as a pre-programmed sequence, etc. Throughout this work we consider various environmental contexts of interest as well as the fully general setting where no assumptions are made about the sequence of condition distributions.

C.2.3 The Multiplicative Weights Updating (MWU) algorithm

Online learning problems of this form are common, and a simple yet powerful learning algorithm known as **Multiplicative Weights Updating (MWU)** has been shown to be an effective solution to such problems in computer science, machine learning, economics, and other fields. Here we highlight the conceptual intuition underlying this algorithm by deriving MWU as an entropy-regularized loss-minimization procedure before stating its recursive update rule.

Derivation of MWU

This section derives the MWU update rule using the method of Lagrange multipliers. The logic that sets up the optimization offers some insights, but readers simply interested in the definition of MWU can skip to the next section.

MWU can be motivated by first considering a natural retrospective decision rule. Suppose a learner seeking to minimize cumulative loss chooses to adopt the strategy that would have performed best against the previously observed sequence of loss vectors. This is known as the **Follow the Leader (FTL)** algorithm:

$$\boldsymbol{p}^T \in \arg\min_{\boldsymbol{q}\in\Delta_n} \sum_i q_i L_i^T , \quad \text{[C11]}$$

where $L_i^T = \sum_{t=0}^{T-1} \ell_i^t$ is the cumulative expected loss of the pure strategy i. FTL places all weight on a pure strategy that has received the lowest cumulative loss so far. This update can be unstable, however, as all of the weight may switch from one pure strategy to another between rounds. Furthermore, a strategy that performed well against past loss vectors need not continue to do so against future loss vectors, so a learner using unregularized FTL can be easily misled by an unfavorable or adversarial sequence.

To add stability to the update, we can add a convex regularization function $\mathcal{R}(\boldsymbol{q})$ that penalizes abrupt concentration of the mixed strategy. The resulting update rule is known as the **Follow the Regularized Leader (FTRL)** algorithm:

$$\boldsymbol{p}^T = \arg\min_{\boldsymbol{q}\in\Delta_n} \left\{ \sum_i q_i L_i^T + \frac{1}{\eta}\mathcal{R}(\boldsymbol{q}) \right\} . \quad \text{[C12]}$$

Here $\eta \in (0, \infty)$ is a learning-rate parameter that modulates the relative emphasis placed on minimizing past loss versus constraining change through the regularization term. When η is large, the weights respond strongly to differences in cumulative loss and the update approaches unregularized FTL. When η is small, the regularization term has greater influence and the weights change more conservatively.

A natural regularizer is the relative entropy $D(\boldsymbol{q} \,\|\, \boldsymbol{p}^0)$, where $\boldsymbol{p}^0$ is a specified initial distribution. This regularizer penalizes departures from the initial distribution, and it reduces to negative entropy regularization when $\boldsymbol{p}^0$ is uniform. With this choice, the FTRL update becomes

$$\boldsymbol{p}^T = \arg\min_{\boldsymbol{q}\in\Delta_n} \left\{ \sum_i q_i L_i^T + \frac{1}{\eta} D(\boldsymbol{q} \| \boldsymbol{p}^0) \right\} . \quad \text{[C13]}$$

Assume for now that $p_i^0 > 0$ for all i. The objective in Equation C13 is strictly convex and therefore has a unique minimum. To find this minimum, we introduce a Lagrange multiplier λ for the constraint $\sum_i q_i = 1$:

$$\mathcal{L}(\boldsymbol{q}, \lambda) = \sum_i q_i L_i^T + \frac{1}{\eta} \sum_i q_i \log \frac{q_i}{p_i^0} + \lambda \left(\sum_i q_i - 1 \right) . \quad \text{[C14]}$$

Setting the derivative with respect to q_i equal to zero gives

$$\frac{\partial \mathcal{L}}{\partial q_i} = L_i^T + \frac{1}{\eta}\left(1 + \log \frac{q_i}{p_i^0}\right) + \lambda = 0\,. \tag{C15}$$

Solving for q_i, we obtain

$$\log \frac{q_i}{p_i^0} = -\eta L_i^T - \eta\lambda - 1 \tag{C16}$$

$$q_i = p_i^0 \exp\left(-\eta L_i^T - \eta\lambda - 1\right)\,. \tag{C17}$$

The factor $\exp(-\eta\lambda - 1)$ is common to every strategy, and imposing the constraint $\sum_i q_i = 1$ determines its value:

$$\exp(-\eta\lambda - 1) = \frac{1}{\sum_z p_z^0\, e^{-\eta L_z^T}}\,. \tag{C18}$$

Substituting this expression into the solution for q_i, we obtain the minimizing distribution:

$$p_i^T = \frac{p_i^0\, e^{-\eta L_i^T}}{\sum_z p_z^0\, e^{-\eta L_z^T}}\,. \tag{C19}$$

Thus, the FTRL solution assigns exponentially greater weight to strategies having lower cumulative loss while retaining the initial distribution $\boldsymbol{p}^0$ as a prior weighting over the alternatives.

Because $L_i^{t+1} = L_i^t + \ell_i^t$ we can put this solution into a recursive form that depends only on the losses from the most recent round:

$$p_i^{t+1} = \frac{p_i^0\, e^{-\eta(L_i^t + \ell_i^t)}}{\sum_z p_z^0\, e^{-\eta(L_z^t + \ell_z^t)}} = \frac{p_i^t\, e^{-\eta \ell_i^t}}{\sum_z p_z^t\, e^{-\eta \ell_z^t}}\,. \tag{C20}$$

This is the recursive update rule that implements FTRL with relative-entropy regularization.

If some type has $p_i^0 = 0$, the same update applies on the support of $\boldsymbol{p}^0$. Just as natural selection cannot generate an absent type without mutation, MWU preserves zero initial weight.

The MWU update rule

In the previous section, we derived an update rule that implements the Follow the Regularized Leader algorithm with relative entropy regularization. This update rule has been repeatedly derived in many contexts and goes by several names, including **Multiplicative Weights Updating** (see also exponential weights algorithm, Hedge, exponentiated gradient, entropic online mirror descent):

$$p_i^{t+1} = \frac{p_i^t\, e^{-\eta \ell_i^t}}{\sum_z p_z^t\, e^{-\eta \ell_z^t}}\,. \tag{C21}$$

MWU increases the weights of strategies having relatively low loss and decreases the weights of strategies having relatively high loss.

MWU is no-regret under conventional expected-loss evaluation (see Appendix C.2.5) under a range of conditions and achieves useful finite-time regret bounds. This means that a learner that uses MWU to update their strategy can learn an approximation of an optimal strategy and achieves a cumulative loss that is nearly as low as that of the optimal strategy. In some settings, such as where the learning rate can be tuned or where the loss function has certain properties, it can be shown that MWU is an optimal learning process in the sense that no other learning algorithm can achieve a tighter bound on regret (Freund and Schapire 1999, Cesa-Bianchi and Lugosi 2006). In this work, we are particularly interested in the instance of MWU that is equivalent to natural selection (see Appendix C.2.4). The application of conventional MWU regret bounds to natural selection and population growth is considered in Appendix D.3.

C.2.4 Natural selection as a fitness-based instance of MWU

In Appendix C.2.2 we established the online learning problem as the basis for a model of evolution, and in Appendix C.2.3 we derived the MWU update rule, which is known to be an effective learning process for problems of this kind. Recent work has shown that discrete-time replicator dynamics is equivalent to an instantiation of MWU (Chastain et al. 2014, Mehta et al. 2015, Meir and Parkes 2015, Chastain 2017). In this section, we review this equivalence and make explicit the fitness-derived loss vector that underlies it.

Let $W_{*j} := \max_i W_{ij}$ denote the maximum Wrightian fitness available in condition j. A *condition-wise oracle* that matches each environmental condition (microhabitat) with a locally optimal type achieves the maximum possible mean fitness

$$\langle W_* \rangle_{\boldsymbol{x}^t} = \sum_j x_j^t W_{*j}\,. \tag{C22}$$

Because the oracle uses a locally optimal type in each condition, its mean fitness is at least that of any fixed type:

$$\langle W_* \rangle_{\boldsymbol{x}^t} \geq \langle W_i \rangle_{\boldsymbol{x}^t} \qquad \text{for every } i. \tag{C23}$$

The condition-wise oracle is a natural benchmark because it has zero loss and ensures that every type's loss is nonnegative. This benchmark is generally unattainable by the actual population because types are assigned to conditions at random rather than matched condition by condition, unless the same type is optimal in every condition.

Define the *growth loss* of type i in round t as the shortfall in its per-generation growth rate relative to the condition-wise oracle:

$$\ell_i^t := \log \langle W_* \rangle_{\boldsymbol{x}^t} - \log \langle W_i \rangle_{\boldsymbol{x}^t} = -\log \left(\frac{\langle W_i \rangle_{\boldsymbol{x}^t}}{\langle W_* \rangle_{\boldsymbol{x}^t}} \right). \tag{C24}$$

Substituting the type growth losses into the MWU update and setting the learning rate to $\eta = 1$ gives

$$p_i^{t+1} = \frac{p_i^t \, e^{-\eta \ell_i^t}}{\sum_z p_z^t \, e^{-\eta \ell_z^t}} \tag{C25}$$

$$= \frac{p_i^t \left(\langle W_i \rangle_{\boldsymbol{x}^t} / \langle W_* \rangle_{\boldsymbol{x}^t} \right)}{\sum_z p_z^t \left(\langle W_z \rangle_{\boldsymbol{x}^t} / \langle W_* \rangle_{\boldsymbol{x}^t} \right)} \tag{C26}$$

$$= \frac{p_i^t \langle W_i \rangle_{\boldsymbol{x}^t}}{\langle W \rangle_{\boldsymbol{p}^t, \boldsymbol{x}^t}}. \tag{C27}$$

This is exactly the discrete-time replicator equation in a variable environment (Equation A18). Thus, natural selection is equivalent to MWU with unit learning rate applied in each generation to the vector of type growth losses.

C.2.5 Two fitness-based notions of loss

Up to this point, $\ell^t(\boldsymbol{p})$ has denoted the one-round loss of a mixed strategy without committing to a particular rule for combining the pure-strategy losses ℓ_i^t. We now distinguish two ways of combining type-level growth losses ℓ_i^t to define the loss of a mixed strategy.

External loss

The type losses that generate replicator dynamics exactly are $\ell_i^t := \log \langle W_* \rangle_{\boldsymbol{x}^t} - \log \langle W_i \rangle_{\boldsymbol{x}^t}$ (Equation C24). Standard online learning analysis assigns a mixed strategy the expected value of these losses across types:

$$\ell_{\text{ext}}^t(\boldsymbol{p}) := \sum_i p_i \ell_i^t = \log \langle W_* \rangle_{\boldsymbol{x}^t} - \langle \log \langle W_i \rangle_{\boldsymbol{x}^t} \rangle_{\boldsymbol{p}}. \tag{C28}$$

This formulation first averages Wrightian fitness across conditions for each type and then takes the logarithm of each type's effective fitness. It then averages the resulting type-level losses according to $\boldsymbol{p}$. This is the conventional expected loss used to define the learning-theoretic quantity of "external regret" (hence the subscript ext), and applying MWU to these underlying type losses exactly reproduces replicator dynamics. However, this rule averages the growth rates of the constituent type subpopulations rather than calculating the growth rate of the population as a whole.

Population growth loss

To measure the actual population-level growth shortfall attributable to a mixed composition, Wrightian fitness must be averaged over both types and environmental conditions before taking the logarithm. We therefore define the one-generation **growth loss** of a population with composition $\boldsymbol{p}$ as

$$\ell_{\text{pop}}^t(\boldsymbol{p}) := \log \langle W_* \rangle_{\boldsymbol{x}^t} - \log \langle W \rangle_{\boldsymbol{p}, \boldsymbol{x}^t}. \tag{C29}$$

For the realized composition $\boldsymbol{p}^t$, this is the one-generation shortfall in population growth attributable to its type composition, relative to the condition-wise oracle. For a counterfactual fixed composition $\boldsymbol{q}$, it gives the corresponding shortfall evaluated in the same environment $\boldsymbol{x}^t$.

This growth loss also has a precise learning-theoretic interpretation. Using the type growth losses defined above (Equation C24), the population growth loss can be written, with unit learning rate $\eta = 1$, as

$$\ell_{\text{pop}}^t(\boldsymbol{p}) = -\log \left(\sum_i p_i e^{-\ell_i^t} \right). \tag{C30}$$

In learning theory, the right-hand side is the *mix loss* associated with the distribution $\boldsymbol{p}$ and the type-loss vector. Exponentiating the negative type losses recovers the types' oracle-relative growth factors; mix loss averages those factors under $\boldsymbol{p}$ and then converts the result back into a loss by taking the negative logarithm. This measures the growth shortfall of the mixed population as a whole, rather than the arithmetic average of the separate growth losses of its constituent types. Thus, the population growth formulation arises naturally within MWU: external loss is the arithmetic expectation of the type growth losses, whereas population growth loss is their mix loss.

Relationship between external loss and growth loss:

External loss and population growth loss differ by a nonnegative Jensen gap across types:

$$\ell^t_{\mathrm{ext}}(\boldsymbol{p}) - \ell^t_{\mathrm{pop}}(\boldsymbol{p}) = J^t(\boldsymbol{p})\,, \tag{C31}$$

where

$$J^t(\boldsymbol{p}) := \log \langle W \rangle_{\boldsymbol{p},\boldsymbol{x}^t} - \left\langle \log \langle W_i \rangle_{\boldsymbol{x}^t} \right\rangle_{\boldsymbol{p}} \geq 0\,. \tag{C32}$$

Non-negativity follows from the concavity of the logarithm. Thus, external loss is always at least as large as the population growth loss of the same composition in the same generation.

The gap has a direct biological interpretation. The composition-dependent population growth represented by mean fitness depends on total reproductive output, which is pooled across types before its logarithm is taken. Types with higher effective fitness therefore contribute disproportionately to the offspring population relative to their pre-selection frequencies. External loss instead averages the type growth losses using those pre-selection frequencies. Consequently, $J^t(\boldsymbol{p})$ is the amount by which this separate-type accounting exceeds the growth loss of the mixed population as a whole.

External loss and population growth loss coincide exactly if and only if all types present in $\boldsymbol{p}$ have the same effective fitness. More generally, external loss and population growth loss agree to first order under weak selection. Let ε measure selection strength, so that effective fitness differences among types are $O(\varepsilon)$. Equivalently, the type growth losses can be written as

$$\ell^{t,\varepsilon}_i = c^t + \varepsilon a^t_i + O(\varepsilon^2)\,, \tag{C33}$$

where c^t is common to all types and therefore does not affect the Jensen gap. A Taylor expansion gives

$$J^{t,\varepsilon}(\boldsymbol{p}) = \frac{1}{2}\operatorname{Var}_{\boldsymbol{p}}\left(\ell^{t,\varepsilon}_i\right) + O(\varepsilon^3) = \frac{\varepsilon^2}{2}\operatorname{Var}_{\boldsymbol{p}}\left(a^t_i\right) + O(\varepsilon^3) = O(\varepsilon^2)\,. \tag{C34}$$

Thus, the per-generation difference between external loss and population growth loss has no first-order term in the strength of selection.

Accordingly, external loss and population growth loss agree to first order under weak selection. External regret also provides a conservative upper bound on population growth regret for any reference composition whose Jensen gap is zero, including any pure reference composition. For mixed reference compositions more generally, the Jensen gap affects both the exact biological interpretation and whether external-regret bounds transfer to population growth regret.

Cumulative loss and regret under the two formulations:

The two definitions of one-round loss lead to corresponding definitions of cumulative loss and regret. Under the external formulation, the cumulative losses of the learner and a fixed competing strategy $\boldsymbol{q}$ over the first T rounds are

$$L^T_{\mathrm{ext}} := \sum_{t=0}^{T-1} \ell^t_{\mathrm{ext}}(\boldsymbol{p}^t)\,, \qquad L^T_{\mathrm{ext},\boldsymbol{q}} := \sum_{t=0}^{T-1} \ell^t_{\mathrm{ext}}(\boldsymbol{q})\,. \tag{C35}$$

The learner's external regret relative to $\boldsymbol{q}$ is

$$R^T_{\mathrm{ext},\boldsymbol{q}} := L^T_{\mathrm{ext}} - L^T_{\mathrm{ext},\boldsymbol{q}} \tag{C36}$$

$$= \sum_{t=0}^{T-1} \left(\ell^t_{\mathrm{ext}}(\boldsymbol{p}^t) - \ell^t_{\mathrm{ext}}(\boldsymbol{q})\right)\,. \tag{C37}$$

In computational learning theory, external regret is commonly evaluated against an optimal fixed strategy chosen in hindsight:

$$\boldsymbol{q}^T_{\mathrm{ext}} \in \arg\min_{\boldsymbol{q}\in\Delta(\mathrm{supp}(\boldsymbol{p}^0))} L^T_{\mathrm{ext},\boldsymbol{q}}\,. \tag{C38}$$

Because cumulative external loss is linear in $\boldsymbol{q}$, an optimal composition can always be chosen to be a pure strategy. Nevertheless, it is convenient to allow $\boldsymbol{q}$ to denote any mixed strategy.

The corresponding cumulative population growth losses are

$$L^T_{\mathrm{pop}} := \sum_{t=0}^{T-1} \ell^t_{\mathrm{pop}}(\boldsymbol{p}^t)\,, \qquad L^T_{\mathrm{pop},\boldsymbol{q}} := \sum_{t=0}^{T-1} \ell^t_{\mathrm{pop}}(\boldsymbol{q})\,. \tag{C39}$$

The population growth regret with respect to $\boldsymbol{q}$ is

$$R^T_{\mathrm{pop},\boldsymbol{q}} := L^T_{\mathrm{pop}} - L^T_{\mathrm{pop},\boldsymbol{q}} \quad [C40]$$

$$= \sum_{t=0}^{T-1} \left(\ell^t_{\mathrm{pop}}(\boldsymbol{p}^t) - \ell^t_{\mathrm{pop}}(\boldsymbol{q})\right) \quad [C41]$$

$$= \sum_{t=0}^{T-1} \left(\log \langle W\rangle_{\boldsymbol{q},\boldsymbol{x}^t} - \log \langle W\rangle_{\boldsymbol{p}^t,\boldsymbol{x}^t}\right) . \quad [C42]$$

This directly measures the difference between the cumulative growth of a counterfactual lineage having fixed composition $\boldsymbol{q}$ and that of the evolving lineage. An optimal fixed reference composition is

$$\boldsymbol{q}^T_{\mathrm{pop}} \in \arg \min_{\boldsymbol{q}\in\Delta(\mathrm{supp}(\boldsymbol{p}^0))} L^T_{\mathrm{pop},\boldsymbol{q}} . \quad [C43]$$

Unlike the external loss objective, the growth loss objective is not linear in $\boldsymbol{q}$, so a growth-optimal reference composition need not be a pure strategy and need not coincide with an optimal external reference.

The two regret measures are related by:

$$R^T_{\mathrm{ext},\boldsymbol{q}} - R^T_{\mathrm{pop},\boldsymbol{q}} = \sum_{t=0}^{T-1} \left(J^t(\boldsymbol{p}^t) - J^t(\boldsymbol{q})\right) . \quad [C44]$$

Although the external loss is at least as large as the growth loss when the two are evaluated at the same composition, this does not imply a general ordering between their associated regret measures. Regret subtracts the loss of one composition from that of another, and the difference between their Jensen gaps may have either sign. As with their corresponding losses, however, the two regret measures agree to first order and converge in the weak-selection limit.

External loss provides the conventional connection to MWU external regret theory, whereas the population growth formulation yields the primary biological definitions of mismatch load and relative lineage fitness (Appendix D) because these quantities measure differences in actual population growth. As shown above, these two formulations are closely related and agree to first order. Because our focus is on the biological setting, we treat the population growth formulation as the default in the main text and Appendix D. Except where we compare the external and growth formulations explicitly, we therefore omit the subscript pop and use the following notation:

$$\ell^t(\boldsymbol{q}) \equiv \ell^t_{\mathrm{pop}}(\boldsymbol{q}) ; \qquad L^T \equiv L^T_{\mathrm{pop}} ; \quad [C45]$$

$$L^T_{\boldsymbol{q}} \equiv L^T_{\mathrm{pop},\boldsymbol{q}} ; \qquad R^T_{\boldsymbol{q}} \equiv R^T_{\mathrm{pop},\boldsymbol{q}} . \quad [C46]$$

Appendix D develops the biological cumulative loss quantities and their associated regret measures in greater detail.

C.3 Stationary-target and general learning settings for evolving populations

The performance of natural selection can be evaluated using different fixed reference compositions. We distinguish a stationary-target setting, in which selection approaches a single time-independent learning target, from the general setting, in which no stationary target is assumed and performance is evaluated against the optimal fixed composition in hindsight. These settings are nested rather than mutually exclusive: a population with a stationary learning target may also be evaluated against its hindsight-optimal fixed composition, and these two reference compositions need not coincide.

In each round t, the distribution of environmental conditions $\boldsymbol{x}^t$ determines the vector of type growth factors

$$\mathbf{W}\boldsymbol{x}^t = \left(\langle W_1\rangle_{\boldsymbol{x}^t}, \ldots, \langle W_n\rangle_{\boldsymbol{x}^t}\right)^{\mathsf{T}} \quad [C47]$$

that governs the discrete-time selection update. Changes in $\boldsymbol{x}^t$ may therefore change the selection dynamics even when the underlying type-condition fitness matrix $\mathbf{W}$ remains fixed. The distribution $\boldsymbol{x}^t$ may be generated exogenously or may depend on the population state, as in a frequency-dependent game with $\boldsymbol{x}^t = \boldsymbol{p}^t$.

For specified selection dynamics, including the rule determining the sequence of environments, let $\mathcal{B}(\overline{\boldsymbol{p}})$ denote the set of initial compositions whose resulting trajectories converge to a time-independent composition $\overline{\boldsymbol{p}}$ (its "basin of attraction").

Definition. *A fixed type distribution $\overline{\boldsymbol{p}}$ is a* **stationary learning target** *for a selection trajectory $\boldsymbol{p}^0, \boldsymbol{p}^1, \ldots, \boldsymbol{p}^\infty$ if*

$$\boldsymbol{p}^0 \in \mathcal{B}(\bar{\boldsymbol{p}}), \qquad \textit{equivalently} \qquad \lim_{t\to\infty} \boldsymbol{p}^t = \bar{\boldsymbol{p}},$$

and the potential information with respect to $\overline{\boldsymbol{p}}$ does not increase along the trajectory:

$$D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^{t+1}) \le D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t) \qquad \textit{for discrete-time dynamics}, \quad [C48]$$

$$\frac{d}{dt} D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t) \le 0 \qquad \textit{for continuous-time dynamics}. \quad [C49]$$

We also require $\mathrm{supp}(\overline{\boldsymbol{p}}) \subseteq \mathrm{supp}(\boldsymbol{p}^0)$*, since selection cannot introduce a type that is absent from the initial population.*

As shown in Appendix B.2.2, an accessible ESS in a symmetric frequency-dependent game provides an important example of a stationary learning target for continuous-time replicator dynamics: locally, its potential information is a strict Lyapunov function. The converse does not hold in general, because an attracting fixed point need not be an ESS. An attracting fixed point that is not an ESS may nevertheless be a stationary learning target if it satisfies the conditions above. Neither an ESS nor a stationary learning target is required to maximize population mean fitness across alternative population states.

The term "stationary" refers to the learning target, not necessarily to the environmental distribution. The stationary-target setting therefore encompasses a fixed exogenous environment, changing exogenous environments whose induced dynamics share a common target, and a fixed autonomous frequency-dependent game in which $\boldsymbol{x}^t$ changes with $\boldsymbol{p}^t$ while the underlying game remains unchanged. In the last case, the fixed points are properties of the autonomous game and do not change merely because the population moves through different compositions.

Appendix D: The Efficiency of Natural Selection

This appendix develops the load and regret measures used to evaluate the performance of natural selection. Unless otherwise indicated, loss, cumulative loss, and regret refer to the population growth formulation introduced in Appendix C.2.5; the subscript ext is used to distinguish external loss-based quantities.

D.1 Mismatch load quantifies the cumulative fitness loss of selection

The *genetic load* of a population refers to "the extent to which [the population] is impaired by the fact that not all individuals in the population are of the optimum type" (Crow 1958). A population that includes types poorly matched to the current environment will have lower mean fitness than a population with the fitness-maximizing composition, and genetic load quantifies the resulting shortfall in population mean fitness. A number of factors can contribute to the presence of suboptimal types in a population and thus to its overall genetic load. For example, *mutation load* refers to the depression in population fitness maintained by recurrent deleterious mutation, and *segregation load* refers to the depression in population fitness caused by the production of less-fit homozygotes through allele segregation in the context of heterozygote advantage (Crow 1958, Kimura 1960). Our focus is *substitution load* and its extension, *mismatch load*: the relative shortfall in population fitness that results from mismatch between types and environmental conditions.

D.1.1 Substitution load

Haldane introduced the basic concept in terms of the selective deaths associated with replacing one allele by a fitter allele (Haldane 1937, 1957). Kimura subsequently defined substitution load as the cumulative difference between the growth rate of the optimal type and the mean growth rate of the evolving population over the course of a substitution (Kimura 1960, 1961). Kimura's formulation makes clear that this quantity need not represent literal deaths: it is a shortfall in expected reproductive output (before any common density regulation), relative to an optimally composed reference population.

Consider a constant, homogeneous environment in which the optimal type has maximal Wrightian fitness $W_* := \max_i W_i$ and initial frequency $p_*^0 > 0$. Wrightian fitness W_i gives the expected per-capita reproductive output of type i over one generation, and $\log W_i$ gives its per-generation growth rate (Appendix A.1.1). Likewise, the mean fitness $\langle W \rangle_{\boldsymbol{p}^t}$ gives the expected growth factor of the population in generation t, so the per-generation growth rate of the population as a whole is $\log \langle W \rangle_{\boldsymbol{p}^t}$. The difference between the growth rate of the optimal type and that of the population is therefore

$$\log W_* - \log \langle W \rangle_{\boldsymbol{p}^t} \ . \tag{D1}$$

This difference gives the population's one-generation growth shortfall due to the presence of less-fit types, relative to a population consisting entirely of the optimal type.

Substitution load L_{sub}^T is defined as the sum of these per-generation growth shortfalls over T generations of the substitution process:

$$L_{\mathrm{sub}}^T := \sum_{t=0}^{T-1} \left(\log W_* - \log \langle W \rangle_{\boldsymbol{p}^t} \right). \tag{D2}$$

This definition is the discrete-generation analog of Kimura's (1961) continuous-time Malthusian fitness formulation.

Equivalently, substitution load is the log ratio between the cumulative growth factors of the optimal type and the evolving population:

$$L_{\mathrm{sub}}^T = \log \frac{(W_*)^T}{\prod_{t=0}^{T-1} \langle W \rangle_{\boldsymbol{p}^t}}. \tag{D3}$$

These growth factors represent expected reproductive output before any common density-regulation or experimental dilution step. They therefore need not equal changes in observed census size when a population is maintained near carrying capacity or repeatedly returned to a controlled density.

The replicator update for the optimal type gives

$$\frac{p_*^{t+1}}{p_*^t} = \frac{W_*}{\langle W \rangle_{\boldsymbol{p}^t}}, \tag{D4}$$

and thus

$$L_{\mathrm{sub}}^T = \sum_{t=0}^{T-1} \left(\log \left(\frac{p_*^{t+1}}{p_*^t} \langle W \rangle_{\boldsymbol{p}^t} \right) - \log \langle W \rangle_{\boldsymbol{p}^t} \right) = \sum_{t=0}^{T-1} \log \frac{p_*^{t+1}}{p_*^t} = \log \frac{p_*^T}{p_*^0}. \tag{D5}$$

Therefore, if selection carries the optimal type to fixation, so that $p_*^T \to 1$ as $T \to \infty$, then (Kimura 1961):

$$\lim_{T \to \infty} L_{\mathrm{sub}}^T = -\log p_*^0. \tag{D6}$$

D.1.2 Mismatch load

The classical definitions of substitution load introduced by Haldane (1957) and Kimura (1960, 1961) consider evolution in a constant, homogeneous environment. In the substitution model considered above, an initially present type is uniquely optimal, and selection drives it toward fixation. Substitution load measures the cumulative fitness shortfall associated with selection completing this substitution gradually.

To extend the same growth comparison to heterogeneous and changing environments, let $W_{ij} > 0$ be the fitness of type i in condition j, let $W_{*j} := \max_i W_{ij}$, and let $\boldsymbol{x}^t$ be the distribution of conditions in generation t. A *condition-wise oracle* that matches each condition (microhabitat) with its locally optimal type achieves the highest possible one-generation growth factor

$$\langle W_* \rangle_{\boldsymbol{x}^t} = \sum_j x_j^t W_{*j}, \tag{D7}$$

whereas a population with composition $\boldsymbol{p}^t$ has growth factor

$$\langle W \rangle_{\boldsymbol{p}^t, \boldsymbol{x}^t} = \sum_{i,j} p_i^t x_j^t W_{ij}. \tag{D8}$$

The one-generation growth loss of the population $\boldsymbol{p}^t$ is the corresponding shortfall in per-generation growth rate relative to the condition-wise oracle,

$$\ell^t(\boldsymbol{p}^t) := \log \langle W_* \rangle_{\boldsymbol{x}^t} - \log \langle W \rangle_{\boldsymbol{p}^t, \boldsymbol{x}^t}, \tag{D9}$$

and the **mismatch load** is defined as the cumulative growth loss accumulated by the evolving population:

$$L^T := \sum_{t=0}^{T-1} \ell^t(\boldsymbol{p}^t) = \sum_{t=0}^{T-1} \left(\log \langle W_* \rangle_{\boldsymbol{x}^t} - \log \langle W \rangle_{\boldsymbol{p}^t, \boldsymbol{x}^t} \right). \tag{D10}$$

When there is a single constant condition, $\ell^t(\boldsymbol{p}^t)$ equals the one-generation load in Equation D1, and mismatch load reduces to substitution load ($L^T = L^T_{\mathrm{sub}}$). In heterogeneous or variable environments, by contrast, there may be no fixed type composition that can reproduce the condition-wise oracle. Mismatch load may therefore continue to accrue even after selection reaches a stationary composition. This does not imply that the population continues to perform poorly relative to feasible fixed compositions, however; that relative comparison is measured by regret.

Mismatch load is therefore precisely the cumulative population growth loss defined in Appendix C.2.5. The expected type loss used in standard external regret analysis is closely related, but exceeds growth loss by a Jensen gap across types.

D.2 Regret quantifies relative growth performance

Mismatch load compares the evolving population with a condition-wise oracle that matches each condition with its locally optimal type—a benchmark generally unattainable by any fixed composition. Regret instead compares the evolving population with a feasible fixed composition evaluated across the same realized sequence of environments. Regret therefore subtracts the baseline load experienced by the chosen reference composition and isolates the growth shortfall associated with the population's trajectory.

D.2.1 Regret as relative mismatch load

Regret compares the evolving population to a fixed reference composition evaluated across the same realized sequence of environments. Let $\boldsymbol{q}$ denote such a composition supported on the types present in the initial population. Its cumulative growth loss across the same realized sequence of environmental distributions is

$$L^T_{\boldsymbol{q}} := \sum_{t=0}^{T-1} \ell^t(\boldsymbol{q}). \tag{D11}$$

When $\boldsymbol{x}^t$ depends on the evolving population state, this comparison holds the realized environmental sequence fixed, as in standard external regret. That is, $L^T_{\boldsymbol{q}}$ evaluates $\boldsymbol{q}$ against the environments encountered along the actual trajectory, rather than those that a population fixed at $\boldsymbol{q}$ would itself generate.

The **growth regret** of the evolving population relative to $\boldsymbol{q}$ is then

$$R^T_{\boldsymbol{q}} := L^T - L^T_{\boldsymbol{q}} \tag{D12}$$

$$= \sum_{t=0}^{T-1} \left(\ell^t(\boldsymbol{p}^t) - \ell^t(\boldsymbol{q}) \right) \tag{D13}$$

$$= \sum_{t=0}^{T-1} \left(\log \langle W \rangle_{\boldsymbol{q}, \boldsymbol{x}^t} - \log \langle W \rangle_{\boldsymbol{p}^t, \boldsymbol{x}^t} \right). \tag{D14}$$

Growth regret is the difference between the cumulative population growth losses—or mismatch loads—of the evolving population and the fixed reference composition. Because both losses are measured relative to the same condition-wise oracle, the oracle growth terms cancel. Growth regret therefore isolates the difference in cumulative growth between the evolving population and the reference composition, rather than their respective shortfalls relative to the generally unattainable oracle.

D.2.2 Regret as relative lineage fitness

This difference in cumulative growth has a direct lineage interpretation. Thinking of a lineage as the descendants of an initial population, compare the evolving lineage with a counterfactual reference lineage whose composition is held fixed at $\boldsymbol{q}$ in every generation. Both lineages are evaluated against the same realized sequence of environmental distributions. Let

$$\Gamma^T := \prod_{t=0}^{T-1} \langle W \rangle_{\boldsymbol{p}^t, \boldsymbol{x}^t} \qquad \text{and} \qquad \Gamma_{\boldsymbol{q}}^T := \prod_{t=0}^{T-1} \langle W \rangle_{\boldsymbol{q}, \boldsymbol{x}^t} \tag{D15}$$

denote the cumulative growth factors of the evolving lineage and the reference lineage, respectively. Because reproductive output compounds multiplicatively across generations,

$$R_{\boldsymbol{q}}^T = \log \Gamma_{\boldsymbol{q}}^T - \log \Gamma^T \tag{D16}$$

$$= \log \frac{\Gamma_{\boldsymbol{q}}^T}{\Gamma^T}. \tag{D17}$$

Thus, $\exp(R_{\boldsymbol{q}}^T) = \Gamma_{\boldsymbol{q}}^T / \Gamma^T$ is the fold difference between the cumulative growth factors of the reference and evolving lineages, while growth regret is the logarithm of that fold difference. If the two lineages begin with the same abundance, a growth regret of $R_{\boldsymbol{q}}^T = \rho$ nats means that the reference lineage produces e^ρ times as many expected descendants after T generations as the evolving lineage.

By analogy with the relative fitness of a type, the ratio $\Gamma^T / \Gamma_{\boldsymbol{q}}^T$ can be interpreted as the **relative lineage fitness** of the evolving population with respect to the fixed reference composition. Regret is its negative logarithm:

$$R_{\boldsymbol{q}}^T = -\log \frac{\Gamma^T}{\Gamma_{\boldsymbol{q}}^T}. \tag{D18}$$

For a fixed reference composition and environmental sequence, minimizing growth regret is therefore equivalent to maximizing the evolving population's relative lineage fitness.

D.2.3 Target regret and Hindsight regret

The choice of reference composition determines which aspect of the selection process regret evaluates. We consider two related comparisons. *Target regret* compares the population's cumulative growth while approaching a stationary learning target with the cumulative growth that would have been achieved by using that target composition throughout. *Hindsight regret* compares the population's trajectory with the best fixed composition that could have been chosen after observing the realized sequence of environmental distributions. Unless stated otherwise, we use 'target regret' and 'hindsight regret' as shorthand for growth regret evaluated against these respective reference compositions.

Target growth regret

In a stationary-target setting, let $\overline{\boldsymbol{p}}$ denote the stationary learning target defined in Appendix C.3. This composition is the fixed endpoint that the population gradually learns to adopt. A counterfactual reference lineage with composition $\overline{\boldsymbol{p}}$ from the outset may still experience persistent mismatch load relative to the condition-wise oracle, particularly in a heterogeneous environment. Target regret subtracts this baseline load and isolates the additional cumulative growth shortfall associated with the population's gradual approach to the target. The cumulative loss of the target composition and the population's target regret are denoted

$$\bar{L}^T := L_{\overline{\boldsymbol{p}}}^T, \qquad \bar{R}^T := R_{\overline{\boldsymbol{p}}}^T = L^T - \bar{L}^T. \tag{D19}$$

In terms of relative lineage fitness, $\bar{R}^T$ is the log ratio of the cumulative growth factor of a reference lineage having the target composition throughout to that of the evolving lineage. It therefore measures the inefficiency of having to learn the stationary target rather than having adopted it from the beginning.

Hindsight growth regret

Target regret requires a stationary learning target to serve as the reference. Such a stationary target may not exist when the accessible learning target changes over time. A more general comparison can be made using the realized environmental sequence. After observing the environmental sequence over T generations, we ask which single fixed composition would have experienced the least cumulative loss (or, equivalently, achieved the greatest cumulative growth) if used throughout that sequence.

Hindsight regret remains a comparison against a fixed composition that does not change over generations to follow a moving learning target or to match each condition with its locally optimal type. Rather, "hindsight" means that the one fixed composition used

for comparison is selected retrospectively, with knowledge of the entire realized sequence. Define this hindsight-optimal composition as

$$\widetilde{\boldsymbol{p}}^T \in \arg\min_{\boldsymbol{q}\in\Delta(\mathrm{supp}(\boldsymbol{p}^0))} L_{\boldsymbol{q}}^T. \tag{D20}$$

The hindsight-optimal composition $\widetilde{\boldsymbol{p}}^T$ depends on the environmental sequence observed through generation T and may therefore change as the evaluation horizon is extended.

The cumulative loss of this composition and the corresponding hindsight regret are

$$\widetilde{L}^T := L_{\widetilde{\boldsymbol{p}}^T}^T, \qquad \text{and} \qquad \widetilde{R}^T := R_{\widetilde{\boldsymbol{p}}^T}^T = L^T - \widetilde{L}^T. \tag{D21}$$

Thus hindsight regret compares the evolving lineage with the fixed reference lineage that would have achieved the greatest cumulative growth over the realized environmental sequence. This is a retrospective counterfactual benchmark: it does not imply that such a reference lineage actually existed or that the population could have identified the optimal composition prospectively.

Because growth loss depends on $\log\langle W\rangle_{\boldsymbol{q},\boldsymbol{x}^t}$, it is nonlinear in $\boldsymbol{q}$. A mixed composition may hedge across heterogeneous or changing conditions and achieve greater cumulative growth than any single type. Consequently, the hindsight-optimal composition may not be pure.

Target regret and hindsight regret answer different questions, but they are nested rather than mutually exclusive comparisons. Hindsight regret can also be evaluated in a stationary-target setting. Whenever both comparisons are defined for the same trajectory, the stationary target $\overline{\boldsymbol{p}}$ is one of the admissible fixed compositions considered by the hindsight optimization. It follows that

$$\widetilde{L}^T \le \bar{L}^T \qquad \text{and} \qquad \widetilde{R}^T \ge \bar{R}^T. \tag{D22}$$

The hindsight comparison is therefore at least as demanding as the target comparison: it selects the lowest-loss fixed benchmark available, rather than committing to the dynamically defined learning target. The excess hindsight regret does not reflect any change in the population's trajectory or cumulative loss; it results solely from comparison with a stronger reference composition. These regret measures are equal when the stationary learning target is also a hindsight-optimal fixed composition for the evaluated sequence, but they need not coincide otherwise.

D.3 Regret and potential information

Regret and potential information capture complementary aspects of the same learning process. Growth regret accumulates the difference in population growth between the evolving population and a fixed reference composition, whereas the potential information $D(\boldsymbol{q}\|\boldsymbol{p}^t)$ measures the discrepancy that remains between the population's current composition and that reference.

For replicator dynamics, changes in potential information are directly related to this relative growth performance. A positive instantaneous or one-generation contribution to growth regret means that the reference composition $\boldsymbol{q}$ grows more rapidly than the evolving population at that time. The identities below show exactly how this relative growth advantage is reflected in changes in potential information, including the additional Jensen correction introduced by finite discrete updates.

D.3.1 Regret and changes in potential information

Continuous time

For continuous-time dynamics, let $r_{*j} := \max_i r_{ij}$ denote the greatest Malthusian fitness available in condition j, so that $\langle r_*\rangle_{\boldsymbol{x}^t} := \sum_j x_j^t r_{*j}$ is the instantaneous growth rate of the condition-wise oracle. The instantaneous population growth loss of a composition $\boldsymbol{q}$ is

$$\ell^t(\boldsymbol{q}) := \langle r_*\rangle_{\boldsymbol{x}^t} - \langle r\rangle_{\boldsymbol{q},\boldsymbol{x}^t}. \tag{D23}$$

The cumulative losses of a trajectory $\boldsymbol{p}^t$ and a fixed reference composition $\boldsymbol{q}$ over the interval $[0,T]$ are therefore

$$L^T := \int_0^T \ell^t(\boldsymbol{p}^t)\,dt \qquad \text{and} \qquad L_{\boldsymbol{q}}^T := \int_0^T \ell^t(\boldsymbol{q})\,dt, \tag{D24}$$

and the corresponding continuous-time growth regret is

$$R_{\boldsymbol{q}}^T := L^T - L_{\boldsymbol{q}}^T = \int_0^T \big(\ell^t(\boldsymbol{p}^t) - \ell^t(\boldsymbol{q})\big)\,dt. \tag{D25}$$

The following proposition shows that the rate at which potential information decreases is exactly equal to the instantaneous growth regret.

Proposition D2. *For any trajectory $\boldsymbol{p}^t$ generated by continuous-time replicator dynamics and any fixed composition $\boldsymbol{q}$ satisfying* $\operatorname{supp}(\boldsymbol{q}) \subseteq \operatorname{supp}(\boldsymbol{p}^0)$*, the instantaneous growth regret relative to $\boldsymbol{q}$ is equal to the negative rate of change in potential information:*

$$\frac{d}{dt} D\big(\boldsymbol{q}\big\|\boldsymbol{p}^t\big) = -\left(\ell^t(\boldsymbol{p}^t) - \ell^t(\boldsymbol{q})\right). \tag{D26}$$

Consequently,

$$R_{\boldsymbol{q}}^T = D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^T\big). \tag{D27}$$

Proof. The potential information relative to $\boldsymbol{q}$ is $D(\boldsymbol{q}\|\boldsymbol{p}^t) = \sum_i q_i \log(q_i/p_i^t)$. Because $\boldsymbol{q}$ is fixed, differentiating with respect to time gives

$$\begin{aligned}
\frac{d}{dt} D\big(\boldsymbol{q}\big\|\boldsymbol{p}^t\big) &= -\sum_i q_i \frac{1}{p_i^t}\frac{dp_i^t}{dt} && \text{[D28]}\\
&= -\sum_i q_i \left(\langle r_i\rangle_{\boldsymbol{x}^t} - \langle r\rangle_{\boldsymbol{p}^t,\boldsymbol{x}^t}\right) && \text{[D29]}\\
&= \langle r\rangle_{\boldsymbol{p}^t,\boldsymbol{x}^t} - \langle r\rangle_{\boldsymbol{q},\boldsymbol{x}^t} && \text{[D30]}\\
&= -\left(\ell^t(\boldsymbol{p}^t) - \ell^t(\boldsymbol{q})\right). && \text{[D31]}
\end{aligned}$$

The second line uses the continuous-time replicator equation in a variable environment (Equation A19), and the last line follows because the oracle term $\langle r_*\rangle_{\boldsymbol{x}^t}$ cancels from the difference in losses. This proves the instantaneous identity.

Integrating both sides over $[0, T]$ and using Equation D25 gives

$$\begin{aligned}
R_{\boldsymbol{q}}^T &= -\int_0^T \frac{d}{dt} D\big(\boldsymbol{q}\big\|\boldsymbol{p}^t\big)\, dt && \text{[D32]}\\
&= -\left[D\big(\boldsymbol{q}\big\|\boldsymbol{p}^t\big)\right]_{t=0}^{t=T} && \text{[D33]}\\
&= D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^T\big), && \text{[D34]}
\end{aligned}$$

which is Equation D27. □

Thus, cumulative growth regret is exactly the net amount of potential information eliminated over the interval. The term $D(\boldsymbol{q}\|\boldsymbol{p}^0)$ is the initial potential relative to the reference composition, while $D(\boldsymbol{q}\|\boldsymbol{p}^T)$ is the potential remaining at time T. If the population converges to $\boldsymbol{q}$, the remaining potential vanishes and total regret converges to the initial potential. For an arbitrary reference composition, the same identity records either progress toward or movement away from it.

Discrete time

The corresponding discrete-time identity uses the type-level growth losses ℓ_i^t and the Jensen gap $J^t(\boldsymbol{q}) = \ell_{\text{ext}}^t(\boldsymbol{q}) - \ell^t(\boldsymbol{q})$ defined in Appendix C.2.5. This gap measures the difference between averaging the type-level growth losses under $\boldsymbol{q}$ and taking the logarithm of the mean fitness of the composition as a whole. It is zero for a pure reference composition or, more generally, whenever all types supported by $\boldsymbol{q}$ have the same effective fitness in that round. Although generally nonzero for a mixed reference composition, the gap is second order in the strength of selection and therefore vanishes per generation in the weak-selection limit developed below.

Proposition D3. *For any trajectory $\boldsymbol{p}^0, \ldots, \boldsymbol{p}^T$ generated by discrete-time replicator dynamics and any fixed composition $\boldsymbol{q}$ satisfying* $\operatorname{supp}(\boldsymbol{q}) \subseteq \operatorname{supp}(\boldsymbol{p}^0)$*, the one-round growth regret is equal to the reduction in potential information plus the Jensen gap of the reference composition:*

$$\ell^t(\boldsymbol{p}^t) - \ell^t(\boldsymbol{q}) = D\big(\boldsymbol{q}\big\|\boldsymbol{p}^t\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}\big) + J^t(\boldsymbol{q}). \tag{D35}$$

Consequently, cumulative growth regret is equal to the net reduction in potential information plus the reference composition's cumulative Jensen gap:

$$R_{\boldsymbol{q}}^T = D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^T\big) + \sum_{t=0}^{T-1} J^t(\boldsymbol{q}). \tag{D36}$$

Proof. Using the MWU form of the discrete replicator update from Equation C27, write

$$p_i^{t+1} = \frac{p_i^t e^{-\ell_i^t}}{Z^t}, \qquad \text{and} \qquad Z^t := \sum_k p_k^t e^{-\ell_k^t} = e^{-\ell^t(\boldsymbol{p}^t)}. \tag{D37}$$

The one-round change in potential information is therefore

$$D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t}\big) \tag{D38}$$

$$= \sum_i q_i \log \frac{p_i^t}{p_i^{t+1}} \tag{D39}$$

$$= \sum_i q_i \left(\ell_i^t + \log Z^t\right) \tag{D40}$$

$$= \ell_{\mathrm{ext}}^t(\boldsymbol{q}) - \ell^t(\boldsymbol{p}^t). \tag{D41}$$

By definition, $J^t(\boldsymbol{q}) = \ell_{\mathrm{ext}}^t(\boldsymbol{q}) - \ell^t(\boldsymbol{q})$. Substituting this relation and rearranging gives Equation D35.

Summing that one-round identity over $t = 0, \ldots, T-1$ gives

$$R_{\boldsymbol{q}}^T = \sum_{t=0}^{T-1} \left[D\big(\boldsymbol{q}\big\|\boldsymbol{p}^t\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}\big)\right] + \sum_{t=0}^{T-1} J^t(\boldsymbol{q}) \tag{D42}$$

$$= D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^T\big) + \sum_{t=0}^{T-1} J^t(\boldsymbol{q}), \tag{D43}$$

because every intermediate divergence appears once with a positive sign and once with a negative sign. This is Equation D36. □

The first two terms in Equation D36 have the same potential-information interpretation as in continuous time. The additional sum records the cumulative discrepancy between type-averaged loss and population growth loss for the fixed reference composition. Consequently, progress toward a mixed reference composition does not by itself account for all growth regret when the types represented in that reference have unequal effective fitness along the realized trajectory.

The preceding proposition concerns growth regret, which is the default regret measure in this paper. Because the replicator update is also an instance of MWU, it is useful to place beside it the exact identity for standard external regret. This comparison shows precisely how the Jensen gap separates the population growth comparison from the conventional learning-theoretic one.

Corollary D3.1. *Under the conditions of Proposition D3, the cumulative external regret relative to $\boldsymbol{q}$ is equal to the reduction in potential information plus the cumulative Jensen gap of the evolving population:*

$$R_{\mathrm{ext},\boldsymbol{q}}^T = D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^T\big) + \sum_{t=0}^{T-1} J^t(\boldsymbol{p}^t). \tag{D44}$$

In particular, for any pure reference composition $\boldsymbol{\delta}_i$,

$$R_{\boldsymbol{\delta}_i}^T = D\big(\boldsymbol{\delta}_i\big\|\boldsymbol{p}^0\big) - D\big(\boldsymbol{\delta}_i\big\|\boldsymbol{p}^T\big), \tag{D45}$$

$$R_{\boldsymbol{\delta}_i}^T = R_{\mathrm{ext},\boldsymbol{\delta}_i}^T - \sum_{t=0}^{T-1} J^t(\boldsymbol{p}^t) \le R_{\mathrm{ext},\boldsymbol{\delta}_i}^T. \tag{D46}$$

Proof. In one round, the difference between external regret and growth regret is

$$\big(\ell_{\mathrm{ext}}^t(\boldsymbol{p}^t) - \ell_{\mathrm{ext}}^t(\boldsymbol{q})\big) - \big(\ell^t(\boldsymbol{p}^t) - \ell^t(\boldsymbol{q})\big) \tag{D47}$$

$$= J^t(\boldsymbol{p}^t) - J^t(\boldsymbol{q}). \tag{D48}$$

Adding this difference to Equation D35 cancels $J^t(\boldsymbol{q})$ and gives

$$\ell_{\mathrm{ext}}^t(\boldsymbol{p}^t) - \ell_{\mathrm{ext}}^t(\boldsymbol{q}) = D\big(\boldsymbol{q}\big\|\boldsymbol{p}^t\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}\big) + J^t(\boldsymbol{p}^t). \tag{D49}$$

Summing over time telescopes the divergence terms and proves Equation D44.

For a pure reference composition $\boldsymbol{\delta}_i$, averaging and taking the log of mean fitness are the same operation, so $J^t(\boldsymbol{\delta}_i) = 0$ in every round. Substituting $\boldsymbol{q} = \boldsymbol{\delta}_i$ into Equation D36 gives the first equality in Equation D46. Summing Equation D48 with $J^t(\boldsymbol{\delta}_i) = 0$ gives

$$R_{\mathrm{ext},\boldsymbol{\delta}_i}^T - R_{\boldsymbol{\delta}_i}^T = \sum_{t=0}^{T-1} J^t(\boldsymbol{p}^t) \ge 0, \tag{D50}$$

which proves the second equality and the inequality. □

The Jensen gaps appear with different compositions in the two identities because the two regret measures aggregate losses differently. Growth regret uses the population growth loss of each composition and therefore contains the reference composition's gap $J^t(\boldsymbol{q})$. Standard external regret averages type-level losses and therefore contains the evolving population's gap $J^t(\boldsymbol{p}^t)$ instead. For a pure reference composition, the reference gap vanishes, and the nonnegative population gaps make external regret an upper bound on growth regret relative to that same type. This implication does not generally extend to a mixed growth-regret reference composition, whose own gaps need not vanish.

D.3.2 Target regret bounds for stationary-target learning problems

By definition, a stationary learning target $\overline{\boldsymbol{p}}$ is a time-independent attracting composition toward which the population converges and relative to which the potential information $D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t)$ is nonincreasing along the population trajectory (Appendix C.3). An ESS in a symmetric frequency-dependent game provides one important example, as shown in Appendix B.2.2, but the following results apply to any stationary learning target satisfying the stated conditions.

This setting turns the identities above into finite-time bounds and limiting results. The population begins with a finite amount of potential information relative to the target. As selection learns the target, that potential decreases and can never fall below zero. The central question is whether the corresponding cumulative growth shortfall is exactly limited by this initial potential or whether discrete generations introduce an additional correction.

Continuous time

Theorem 1. *Let $\overline{\boldsymbol{p}}$ be a stationary learning target for a population beginning at $\boldsymbol{p}^0$. Under continuous-time replicator dynamics, the population's target regret is equal to the potential information eliminated by time T, is bounded above by the initial potential information, and converges to that initial value:*

$$0 \leq \bar{R}^T = D\left(\overline{\boldsymbol{p}}\middle\|\boldsymbol{p}^0\right) - D\left(\overline{\boldsymbol{p}}\middle\|\boldsymbol{p}^T\right) \leq D\left(\overline{\boldsymbol{p}}\middle\|\boldsymbol{p}^0\right), \tag{D51}$$

with

$$\lim_{T\to\infty} \bar{R}^T = D\left(\overline{\boldsymbol{p}}\middle\|\boldsymbol{p}^0\right). \tag{D52}$$

Proof. Substituting $\boldsymbol{q} = \overline{\boldsymbol{p}}$ into Equation D27 gives the equality in Equation D51. The definition of a stationary learning target requires $D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^t)$ to be nonincreasing along the trajectory, so the difference between its initial and current values is nonnegative. The same difference cannot exceed its initial value because KL divergence is nonnegative. This proves the two inequalities.

Because a stationary-target trajectory satisfies $\boldsymbol{p}^T \to \overline{\boldsymbol{p}}$, continuity gives

$$\lim_{T\to\infty} D\left(\overline{\boldsymbol{p}}\middle\|\boldsymbol{p}^T\right) = D(\overline{\boldsymbol{p}}\|\overline{\boldsymbol{p}}) = 0. \tag{D53}$$

Taking the limit of the finite-time identity therefore gives Equation D52. □

The continuous-time result gives the initial potential information a direct growth interpretation. Target regret is the portion of this initial discrepancy that the population has eliminated, and the remaining divergence is the portion still to be learned. Once the target is reached, the cumulative growth shortfall associated with learning the target is $D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^0)$. The environmental trajectory and strength of selection determine how quickly this regret is accumulated, but the limiting total depends only on the initial and target compositions.

Discrete time

In discrete time, the reduction in potential information does not generally account for target regret by itself. Each generation may also contribute a Jensen gap associated with the stationary target. The following theorem states the resulting identity and its finite- and infinite-time consequences.

Theorem 2. *Let $\overline{\boldsymbol{p}}$ be a stationary learning target for a population beginning at $\boldsymbol{p}^0$. Under discrete-time replicator dynamics, the population's target regret is equal to the potential information eliminated by time T plus the cumulative Jensen gap of the target composition:*

$$\bar{R}^T = D\left(\overline{\boldsymbol{p}}\middle\|\boldsymbol{p}^0\right) - D\left(\overline{\boldsymbol{p}}\middle\|\boldsymbol{p}^T\right) + \sum_{t=0}^{T-1} J^t\left(\overline{\boldsymbol{p}}\right). \tag{D54}$$

It consequently satisfies the finite-time bound

$$0 \leq \bar{R}^T \leq D\left(\overline{\boldsymbol{p}}\middle\|\boldsymbol{p}^0\right) + \sum_{t=0}^{T-1} J^t\left(\overline{\boldsymbol{p}}\right). \tag{D55}$$

If the target's per-generation Jensen gaps have a finite cumulative sum, then target regret converges to the initial potential information plus this cumulative Jensen correction:

$$\lim_{T\to\infty} \bar{R}^T = D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^0\big) + \sum_{t=0}^{\infty} J^t\left(\overline{\boldsymbol{p}}\right). \tag{D56}$$

Proof. Substituting $\boldsymbol{q} = \overline{\boldsymbol{p}}$ into Equation D36 gives Equation D54. Because $\overline{\boldsymbol{p}}$ is a stationary learning target, potential information is nonincreasing along the trajectory, and thus

$$0 \le D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^0\big) - D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^T\big) \le D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^0\big). \tag{D57}$$

Every Jensen gap is also nonnegative. Adding their cumulative sum proves Equation D55.

Finally, $\boldsymbol{p}^T \to \overline{\boldsymbol{p}}$ implies $D(\overline{\boldsymbol{p}}\|\boldsymbol{p}^T) \to 0$. If $\sum_{t=0}^{\infty} J^t(\overline{\boldsymbol{p}})$ converges, taking the limit of Equation D54 gives Equation D56. □

The discrete-time result contains the same reduction in potential information as the continuous-time result, together with the cumulative Jensen gaps of the stationary target. This correction is absent when the target is pure, but it can persist for a mixed target whose supported types have unequal effective fitness while the population is still away from the target.

The following corollary identifies the exact zero-gap case and recovers the same potential-information accounting obtained in continuous time.

Corollary D0.1. *Under the conditions of Theorem 2, suppose that all types in the support of $\overline{\boldsymbol{p}}$ have equal effective fitness in every realized environmental distribution $\boldsymbol{x}^t$. Then $J^t(\overline{\boldsymbol{p}}) = 0$ for every t, and*

$$0 \le \bar{R}^T = D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^0\big) - D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^T\big) \le D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^0\big), \tag{D58}$$

with

$$\lim_{T\to\infty} \bar{R}^T = D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^0\big). \tag{D59}$$

This condition always holds for a pure stationary learning target.

Proof. If all types supported by $\overline{\boldsymbol{p}}$ have the same effective fitness in generation t, then their type-level losses are also equal. Let this common loss be c^t. Then

$$\ell^t_{\text{ext}}\left(\overline{\boldsymbol{p}}\right) = \sum_i \overline{p}_i c^t = c^t, \tag{D60}$$

$$\ell^t\left(\overline{\boldsymbol{p}}\right) = -\log \sum_i \overline{p}_i e^{-c^t} = -\log e^{-c^t} = c^t. \tag{D61}$$

Thus $J^t(\overline{\boldsymbol{p}}) = 0$ in every generation. Substituting this equality into Equation D54 gives the finite-time identity and bound. Because $\boldsymbol{p}^T \to \overline{\boldsymbol{p}}$, the terminal potential information converges to zero, giving Equation D59. A pure target satisfies the premise because its support contains only one type. □

A mixed stationary target need not satisfy the zero-gap condition during the approach to the target. For example, types in the support of a frequency-dependent equilibrium have equal effective fitness at equilibrium, but the same target composition may assign positive weight to types with unequal effective fitness when evaluated against an off-equilibrium environmental distribution. The accumulated Jensen correction in Equation D56 accounts for this difference.

Weak-selection limit

The zero-gap condition above is exact but restrictive. To connect the general discrete-time result with continuous-time replicator dynamics, we next consider the limit of weak selection. The key observation is an order-of-magnitude comparison: changes in type frequency occur to first order in the selection coefficients, whereas the Jensen gap arises from curvature and is second order.

To define the weak-selection limit, imagine a family of versions of the same evolutionary process, all beginning from the same type distribution and differing only in the strength of selection. Let $\varepsilon \in (0, 1]$ measure this strength. For each type i and environmental condition j, define

$$W^{\varepsilon}_{ij} := W_{*j}\left(1 - \varepsilon k_{ij}\right), \tag{D62}$$

where $W_{*j} := \max_h W_{hj}$ and $k_{ij} = 1 - W_{ij}/W_{*j}$ is the original Wrightian selection coefficient of type i in condition j. When $\varepsilon = 1$, this recovers the original fitness W_{ij}. Reducing ε moves the fitness of every type toward the greatest fitness attainable in that condition while preserving the relative scale of the original selective differences. As $\varepsilon \to 0$, all types approach the same fitness within each condition and selection becomes increasingly weak.

Weakening selection also slows changes in type frequency. When all selection coefficients are reduced by a factor of ε, selection produces changes of order ε per generation. Approximately $1/\varepsilon$ times as many generations are therefore required to produce a comparable amount of evolutionary change. By contrast, the Jensen gap produced in each generation depends on squared differences among selection coefficients and is of order ε^2. Accumulating gaps of order ε^2 over an evolutionary timescale of order $1/\varepsilon$ produces a total correction of order ε, which vanishes as selection becomes weak (as $\varepsilon \to 0$).

Let $\boldsymbol{x}^{t,\varepsilon}$ denote the environmental distribution experienced in generation t by the process with selection strength ε. This notation allows the environmental trajectory itself to depend on the evolving population, as in a frequency-dependent environment. For an externally specified environmental sequence, $\boldsymbol{x}^{t,\varepsilon} = \boldsymbol{x}^t$ for every ε.

After averaging fitness over the environmental distribution in generation t, define the effective relative fitness and effective Wrightian selection coefficient of type i by

$$w_i^{t,\varepsilon} := \frac{\sum_j x_j^{t,\varepsilon} W_{ij}^{\varepsilon}}{\max_h \sum_j x_j^{t,\varepsilon} W_{hj}^{\varepsilon}} = 1 - k_i^{t,\varepsilon}. \tag{D63}$$

Thus, $k_i^{t,\varepsilon}$ measures the disadvantage of type i relative to the type with the greatest effective fitness in that generation. The common normalization does not affect the replicator update, growth regret, or the Jensen gap. Equivalently, adding the same constant to every type loss changes neither regret differences nor Jensen gaps. We may therefore work with the normalized type-level losses

$$\ell_i^{t,\varepsilon} = -\log w_i^{t,\varepsilon} = -\log\left(1 - k_i^{t,\varepsilon}\right). \tag{D64}$$

For the weak-selection processes considered below, suppose that there is a constant K, independent of t and ε, such that

$$0 \le k_i^{t,\varepsilon} \le K\varepsilon \tag{D65}$$

for every type i and generation t. This condition states that all effective selective disadvantages become proportionally smaller as selection is weakened.

Because the Jensen gap depends on squared selective differences, Equation D65 implies that the gap in any one generation is at most of order ε^2. The population may nevertheless require infinitely many generations to converge exactly to its target. A small positive gap in every generation could therefore accumulate to a nonvanishing total if selective differences fade too slowly. We thus also require that the differences in effective fitness among the types in the target diminish sufficiently quickly as the population approaches that target:

$$\sum_{t=0}^{\infty} \mathrm{Var}_{\overline{\boldsymbol{p}}}\left(k_i^{t,\varepsilon}\right) = O(\varepsilon). \tag{D66}$$

The variance measures how different the effective selection coefficients of the target's constituent types are in each generation. If these coefficients differ by amounts of order ε, their variance is of order ε^2. Accumulating such variance over an effective evolutionary timescale of order $1/\varepsilon$ then gives the order ε on the right-hand side. The condition permits selective differences to decay gradually over an infinite sequence of generations, but rules out a persistent tail whose cumulative effect remains substantial as selection becomes weak.

Corollary D0.2. *Suppose that the progressively weakened evolutionary processes defined above all begin at $\boldsymbol{p}^0$, converge to the same stationary learning target $\overline{\boldsymbol{p}}$, and satisfy Equation D65 and Equation D66. Then the cumulative Jensen-gap correction vanishes as selection becomes weak:*

$$\sum_{t=0}^{\infty} J^{t,\varepsilon}\left(\overline{\boldsymbol{p}}\right) = O(\varepsilon) \longrightarrow 0, \tag{D67}$$

and the limiting target regret converges to the initial potential information:

$$\lim_{\varepsilon \to 0} \lim_{T \to \infty} \bar{R}^{T,\varepsilon} = D\left(\overline{\boldsymbol{p}} \middle\| \boldsymbol{p}^0\right). \tag{D68}$$

Proof. For a fixed composition $\boldsymbol{q}$, define

$$f(k) := -\log(1-k). \tag{D69}$$

Using Equation D64, the Jensen gap can be expressed directly in terms of the effective Wrightian selection coefficients:

$$J^{t,\varepsilon}(\boldsymbol{q}) = \ell_{\mathrm{ext}}^{t,\varepsilon}(\boldsymbol{q}) - \ell^{t,\varepsilon}(\boldsymbol{q}) \tag{D70}$$

$$= -\left\langle \log\left(1 - k^{t,\varepsilon}\right)\right\rangle_{\boldsymbol{q}} + \log\left(1 - \left\langle k^{t,\varepsilon}\right\rangle_{\boldsymbol{q}}\right) \tag{D71}$$

$$= \left\langle f\left(k^{t,\varepsilon}\right)\right\rangle_{\boldsymbol{q}} - f\left(\left\langle k^{t,\varepsilon}\right\rangle_{\boldsymbol{q}}\right). \tag{D72}$$

Because

$$f''(k) = \frac{1}{(1-k)^2}, \tag{D73}$$

Taylor's theorem around $\langle k^{t,\varepsilon} \rangle_{\boldsymbol{q}}$ gives

$$\frac{1}{2} \operatorname{Var}_{\boldsymbol{q}} \left(k_i^{t,\varepsilon} \right) \leq J^{t,\varepsilon}(\boldsymbol{q}) \leq \frac{\operatorname{Var}_{\boldsymbol{q}} \left(k_i^{t,\varepsilon} \right)}{2 \left(1 - \max_i k_i^{t,\varepsilon} \right)^2}. \tag{D74}$$

Thus, the Jensen gap is governed by the variance of the selection coefficients rather than by their mean.

Taking $\boldsymbol{q} = \overline{\boldsymbol{p}}$ and applying Equation D65 gives, for sufficiently small ε,

$$0 \leq \sum_{t=0}^{\infty} J^{t,\varepsilon} \left(\overline{\boldsymbol{p}} \right) \leq \frac{1}{2 \left(1 - K\varepsilon \right)^2} \sum_{t=0}^{\infty} \operatorname{Var}_{\overline{\boldsymbol{p}}} \left(k_i^{t,\varepsilon} \right) \tag{D75}$$

$$= O(\varepsilon). \tag{D76}$$

This proves Equation D67.

Finally, Equation D56 gives

$$\lim_{T \to \infty} \bar{R}^{T,\varepsilon} = D\left(\overline{\boldsymbol{p}} \middle\| \boldsymbol{p}^0 \right) + \sum_{t=0}^{\infty} J^{t,\varepsilon} \left(\overline{\boldsymbol{p}} \right). \tag{D77}$$

Letting $\varepsilon \to 0$ and applying Equation D67 proves Equation D68. □

Thus, weak selection recovers the continuous-time result even for a mixed target whose supported types do not have exactly equal effective fitness throughout the discrete trajectory. The correction vanishes not because every one-generation gap is exactly zero, but because weakening selection suppresses the gap more strongly than it slows evolution. For example, reducing every selection coefficient tenfold makes selective changes in type frequency approximately tenfold slower, so that approximately ten times as many generations are needed. The Jensen gap, however, depends on squared differences among selection coefficients and therefore becomes approximately one hundred times smaller in each generation. Accumulating a one-hundred-fold smaller gap over ten times as many generations produces a total correction that is still tenfold smaller. The cumulative-variance condition extends this reasoning to a population that approaches its target only asymptotically: it requires the remaining selective differences to fade quickly enough that the infinite sequence of progressively smaller gaps does not leave a substantial residual correction.

D.3.3 Hindsight regret bounds for general learning problems

A population need not approach a stationary target for its performance to be evaluated retrospectively. Hindsight regret instead asks how much cumulative growth was lost relative to the single fixed composition that would have performed best across the realized sequence of environmental distributions. Because that reference composition is selected only after the sequence is observed, it remains well defined when the accessible target changes or when the population does not converge.

The stationary-target results followed the disappearance of potential information relative to a reference distribution identified as the endpoint toward which selection converges. In the general setting, the reference composition $\widetilde{\boldsymbol{p}}^T$ is chosen retrospectively and depends on the full realized sequence over the evaluation interval. Once that sequence and horizon are fixed, however, $\widetilde{\boldsymbol{p}}^T$ is one fixed composition, so the identities above can be applied directly.

Continuous time

For a continuous-time trajectory over $[0, T]$, let

$$\widetilde{\boldsymbol{p}}^T \in \arg \min_{\boldsymbol{q} \in \Delta(\operatorname{supp}(\boldsymbol{p}^0))} L_{\boldsymbol{q}}^T \tag{D78}$$

denote a fixed composition minimizing the integrated population growth loss. Write $\widetilde{L}^T := L_{\widetilde{\boldsymbol{p}}^T}^T$ and $\widetilde{R}^T := L^T - \widetilde{L}^T$.

Corollary D0.3. *For any trajectory generated by continuous-time replicator dynamics, hindsight regret satisfies the exact identity*

$$\widetilde{R}^T = D\left(\widetilde{\boldsymbol{p}}^T \middle\| \boldsymbol{p}^0 \right) - D\left(\widetilde{\boldsymbol{p}}^T \middle\| \boldsymbol{p}^T \right), \tag{D79}$$

and is therefore bounded above by the initial potential information relative to the hindsight-optimal composition:

$$\widetilde{R}^T \leq D\left(\widetilde{\boldsymbol{p}}^T \middle\| \boldsymbol{p}^0 \right). \tag{D80}$$

Proof. For the fixed horizon T, substitute $\boldsymbol{q} = \widetilde{\boldsymbol{p}}^T$ into Equation D27. This gives Equation D79. Because KL divergence is nonnegative, the terminal divergence enters with a nonpositive sign and may be dropped to obtain Equation D80. □

The continuous-time bound contains no Jensen correction: hindsight regret is at most the potential information initially separating the population from the hindsight-optimal fixed composition. The nonnegativity of hindsight regret is established below by comparing the hindsight-optimal composition with the population's own terminal composition (Theorem 7).

Continuous-time population growth loss is linear in the fixed composition $\boldsymbol{q}$ (because $\langle r \rangle_{\boldsymbol{q},\boldsymbol{x}^t}$ is linear in $\boldsymbol{q}$), and thus an optimal continuous-time reference can always be chosen to be pure. This differs from the discrete-time growth-loss comparison below, in which the logarithm of mean Wrightian fitness makes the objective nonlinear and can favor a mixed composition that hedges across changing conditions.

Discrete time

For discrete-time dynamics, the same endpoint divergences appear, but the Jensen gaps must be evaluated at the hindsight-optimal growth-loss composition.

Corollary D0.4. *For any trajectory generated by discrete-time replicator dynamics, hindsight regret is equal to the initial potential information relative to the hindsight-optimal composition, minus the potential information remaining at time T, plus the reference composition's cumulative Jensen gap:*

$$\widetilde{R}^T = D\big(\widetilde{\boldsymbol{p}}^T \big\| \boldsymbol{p}^0\big) - D\big(\widetilde{\boldsymbol{p}}^T \big\| \boldsymbol{p}^T\big) + \sum_{t=0}^{T-1} J^t\left(\widetilde{\boldsymbol{p}}^T\right). \tag{D81}$$

Consequently,

$$\widetilde{R}^T \leq D\big(\widetilde{\boldsymbol{p}}^T \big\| \boldsymbol{p}^0\big) + \sum_{t=0}^{T-1} J^t\left(\widetilde{\boldsymbol{p}}^T\right). \tag{D82}$$

Proof. For the fixed horizon T, substitute $\boldsymbol{q} = \widetilde{\boldsymbol{p}}^T$ into Equation D36. By the definitions $R^T_{\widetilde{\boldsymbol{p}}^T} = \widetilde{R}^T$ and $L^T_{\widetilde{\boldsymbol{p}}^T} = \widetilde{L}^T$, this gives Equation D81. Dropping the nonpositive terminal divergence term gives Equation D82. □

The terminal divergence can only reduce regret. The main obstacle to obtaining a discrete-time bound from the initial potential alone is therefore the accumulated Jensen gap of the hindsight reference composition. Because $\widetilde{\boldsymbol{p}}^T$ may be mixed, these gaps cannot generally be discarded.

The gaps can nevertheless be controlled when the type losses have a bounded range. Bounded losses prevent the type-averaged loss of a mixed composition from becoming arbitrarily large relative to its population growth loss. The following proposition establishes the required one-round comparison.

Proposition D1. *Suppose that, in every round,*

$$0 \leq \ell_i^t \leq B \qquad \textit{for all } i. \tag{D83}$$

Then, for any composition $\boldsymbol{q}$, its Jensen gap is bounded by a fixed multiple of its population growth loss:

$$J^t(\boldsymbol{q}) \leq \left(\frac{B}{1-e^{-B}} - 1\right) \ell^t(\boldsymbol{q}). \tag{D84}$$

Proof. Let $b := (1-e^{-B})/B$. For $0 \leq u \leq B$, convexity of e^{-u} gives

$$e^{-u} \leq 1 - bu. \tag{D85}$$

Applying this inequality to every type loss and averaging over $\boldsymbol{q}$ gives

$$\sum_i q_i e^{-\ell_i^t} \leq \sum_i q_i \left(1 - b\ell_i^t\right) \tag{D86}$$

$$= 1 - b\ell^t_{\mathrm{ext}}(\boldsymbol{q}). \tag{D87}$$

Because $-\log u$ is decreasing in u, taking negative logarithms and then using $-\log(1-v) \geq v$ gives

$$\ell^t(\boldsymbol{q}) = -\log \sum_i q_i e^{-\ell_i^t} \geq -\log\left(1 - b\ell^t_{\mathrm{ext}}(\boldsymbol{q})\right) \tag{D88}$$

$$\geq b\ell^t_{\mathrm{ext}}(\boldsymbol{q}). \tag{D89}$$

Rearranging gives

$$\ell^t_{\mathrm{ext}}(\boldsymbol{q}) \leq \frac{B}{1-e^{-B}}\, \ell^t(\boldsymbol{q}). \tag{D90}$$

Subtracting $\ell^t(\boldsymbol{q})$ from both sides and using $J^t(\boldsymbol{q}) = \ell^t_{\mathrm{ext}}(\boldsymbol{q}) - \ell^t(\boldsymbol{q})$ proves Equation D84. □

Theorem 3. *Suppose the bounded-loss condition in Equation D83 holds. The discrete-time hindsight regret is then bounded by the initial potential information relative to the hindsight-optimal composition plus a multiple of the mismatch load that even this reference composition experiences:*

$$\widetilde{R}^T \le D\big(\widetilde{\boldsymbol{p}}^T \big\| \boldsymbol{p}^0\big) + \Big(\frac{B}{1-e^{-B}} - 1\Big)\,\widetilde{L}^T. \tag{D91}$$

Equivalently, the population's cumulative mismatch load is bounded by the same initial potential information plus a multiple of the hindsight-optimal fixed composition's mismatch load:

$$L^T \le D\big(\widetilde{\boldsymbol{p}}^T \big\| \boldsymbol{p}^0\big) + \frac{B}{1-e^{-B}}\widetilde{L}^T. \tag{D92}$$

Proof. Begin with the exact hindsight-regret identity Equation D81. Dropping the nonpositive terminal-divergence term gives

$$\widetilde{R}^T \le D\big(\widetilde{\boldsymbol{p}}^T \big\| \boldsymbol{p}^0\big) + \sum_{t=0}^{T-1} J^t\left(\widetilde{\boldsymbol{p}}^T\right). \tag{D93}$$

Applying Equation D84 in each round yields

$$\sum_{t=0}^{T-1} J^t\left(\widetilde{\boldsymbol{p}}^T\right) \le \Big(\frac{B}{1-e^{-B}} - 1\Big)\sum_{t=0}^{T-1} \ell^t\left(\widetilde{\boldsymbol{p}}^T\right) \tag{D94}$$

$$= \Big(\frac{B}{1-e^{-B}} - 1\Big)\,\widetilde{L}^T. \tag{D95}$$

Substitution proves Equation D91.

Finally, $\widetilde{R}^T = L^T - \widetilde{L}^T$, so adding $\widetilde{L}^T$ to both sides gives Equation D92. □

The resulting load bound has a familiar learning-theoretic form. The divergence $D(\widetilde{\boldsymbol{p}}^T\|\boldsymbol{p}^0)$ measures the initial difficulty of reallocating the population toward the hindsight-optimal composition, while $\widetilde{L}^T$ is the mismatch load that even this best admissible fixed composition experiences across the realized environmental sequence. Natural selection's cumulative mismatch load is bounded by the sum of the first term and a multiplicative inflation of the second. This decomposition distinguishes two sources of cumulative mismatch load: the cost of reallocating an initially mismatched population toward a successful fixed composition, and the load that no single fixed composition could avoid along the realized sequence. When the loss range is small, such as under the weak-selection scaling considered above, $B/(1-e^{-B}) = 1 + B/2 + O(B^2)$, so the multiplier approaches one and the discrete-time bound approaches the corresponding continuous-time form.

When the type losses satisfy $0 \le \ell_i^t \le 1$, the bounds simplify to

$$\widetilde{R}^T \le D\big(\widetilde{\boldsymbol{p}}^T \big\| \boldsymbol{p}^0\big) + \frac{1}{e-1}\widetilde{L}^T, \tag{D96}$$

$$L^T \le D\big(\widetilde{\boldsymbol{p}}^T \big\| \boldsymbol{p}^0\big) + \frac{e}{e-1}\widetilde{L}^T. \tag{D97}$$

The loss range is an explicit biological assumption here. Rescaling the losses would change the learning rate required for MWU to reproduce the same selection dynamics and is therefore not without consequence.

D.3.4 Comparison of regret bounds across learning rates

The preceding growth-regret bound concerns the learning rate $\eta = 1$, for which MWU applied to the type growth losses used here exactly reproduces natural selection (Chastain et al. 2014, Mehta et al. 2015, Meir and Parkes 2015, Chastain 2017) (Appendix C.2.4). It is interesting to ask how the bound changes if the same sequence of type losses is presented to multiplicative updates with different learning rates. This places the learning rate of selection within the same responsiveness–stability comparison commonly made for external regret, while continuing to evaluate performance using the paper's biologically relevant measure of relative population growth.

To compare other rates while holding the learning problem fixed, consider a common, exogenously specified sequence of type-loss vectors satisfying

$$0 \le \ell_i^t \le B \qquad \text{for all } i, t, \tag{D98}$$

where $B > 0$. For each fixed learning rate $\eta > 0$, consider the counterfactual MWU trajectory

$$p_i^{t+1}(\eta) = \frac{p_i^t(\eta)\, e^{-\eta \ell_i^t}}{\sum_k p_k^t(\eta)\, e^{-\eta \ell_k^t}}, \qquad \boldsymbol{p}^0(\eta) = \boldsymbol{p}^0. \tag{D99}$$

Substituting in the type losses we obtain

$$p_i^{t+1}(\eta) = \frac{p_i^t(\eta)\left(\langle W_i\rangle_{\boldsymbol{x}^t}\right)^\eta}{\sum_k p_k^t(\eta)\left(\langle W_k\rangle_{\boldsymbol{x}^t}\right)^\eta}. \tag{D100}$$

The learning rate controls how strongly frequencies respond to the same type-loss differences. Rates $\eta < 1$ attenuate the effective-fitness ratios that determine compositional change, whereas rates $\eta > 1$ amplify them. Only $\eta = 1$ uses the original effective fitnesses without modification and therefore exactly reproduces natural selection. The other rates represent algorithmic counterfactuals that respond more weakly or strongly than selection to the same fitness differences.

For each value of η, the update produces a different counterfactual sequence of compositions $\boldsymbol{p}^0, \boldsymbol{p}^1(\eta), \ldots, \boldsymbol{p}^T(\eta)$. We evaluate the growth of each of these compositions under the original effective fitnesses. Thus, the population growth loss of any composition $\boldsymbol{q}$ remains

$$\ell^t(\boldsymbol{q}) \equiv \ell^t_{\mathrm{pop}}(\boldsymbol{q}) := -\log\left(\sum_i q_i e^{-\ell_i^t}\right) = \log\langle W_*\rangle_{\boldsymbol{x}^t} - \log\langle W\rangle_{\boldsymbol{q},\boldsymbol{x}^t}\,. \tag{D101}$$

Changing η changes the compositions produced by the update and therefore the growth losses those compositions experience. It does not change how the growth loss of a given composition is calculated. When $\eta \neq 1$, the resulting trajectory is an algorithmic counterfactual rather than the trajectory that natural selection would generate from the original fitnesses.

Then we write

$$L^T(\eta) := \sum_{t=0}^{T-1} \ell^t(\boldsymbol{p}^t(\eta)), \tag{D102}$$

$$\begin{aligned} R^T_{\boldsymbol{q}}(\eta) &:= L^T(\eta) - L^T_{\boldsymbol{q}} \\ &= \sum_{t=0}^{T-1}\left(\ell^t(\boldsymbol{p}^t(\eta)) - \ell^t(\boldsymbol{q})\right). \end{aligned} \tag{D103}$$

The reference loss $L^T_{\boldsymbol{q}}$ has no dependence on η because the same fixed composition is evaluated against the same type-loss sequence in every comparison.

To connect the η-dependent update with the fixed definition of population growth loss, define the auxiliary normalizer loss

$$l^t_\eta(\boldsymbol{q}) := -\log\left(\sum_i q_i e^{-\eta\ell_i^t}\right). \tag{D104}$$

This is an accounting quantity associated with the update at learning rate η. It coincides with population growth loss when $\eta = 1$, but for other values of η it should not be interpreted as the growth loss generated by the original, unscaled fitnesses.

The following comparison is the key step. It describes how much of the actual population growth loss must be represented in the normalizing factor of an update with learning rate η.

Proposition D1. *Under Equation D98, the auxiliary normalizer loss satisfies*

$$l^t_\eta(\boldsymbol{q}) \geq a_{\eta,B}\,\ell^t(\boldsymbol{q}), \tag{D105}$$

where

$$a_{\eta,B} := \begin{cases} \eta, & 0 < \eta \leq 1, \\ \dfrac{1-e^{-\eta B}}{1-e^{-B}}, & \eta \geq 1. \end{cases} \tag{D106}$$

The two expressions agree at $\eta = 1$, where $a_{1,B} = 1$ and $l^t_1(\boldsymbol{q}) = \ell^t(\boldsymbol{q})$.

Proof. Because $0 \leq \ell_i^t \leq B$, we have $e^{-\ell_i^t} \in [e^{-B}, 1]$. When $0 < \eta \leq 1$, the function v^η is concave for $v > 0$. Jensen's inequality therefore gives

$$\sum_i q_i e^{-\eta\ell_i^t} = \sum_i q_i \left(e^{-\ell_i^t}\right)^\eta \tag{D107}$$

$$\leq \left(\sum_i q_i e^{-\ell_i^t}\right)^\eta. \tag{D108}$$

Taking negative logarithms yields

$$l_\eta^t(\boldsymbol{q}) = -\log\left(\sum_i q_i e^{-\eta\ell_i^t}\right) \tag{D109}$$

$$\geq -\log\left[\left(\sum_i q_i e^{-\ell_i^t}\right)^\eta\right] \tag{D110}$$

$$= -\eta\log\left(\sum_i q_i e^{-\ell_i^t}\right) \tag{D111}$$

$$= \eta\ell^t(\boldsymbol{q}). \tag{D112}$$

When $\eta \geq 1$, the function v^η is convex. On the interval $[e^{-B}, 1]$, its graph lies below the chord joining its endpoint values. Consequently,

$$\left(e^{-\ell_i^t}\right)^\eta \leq 1 - c_{\eta,B}\left(1 - e^{-\ell_i^t}\right), \qquad \text{where } c_{\eta,B} := \frac{1-e^{-\eta B}}{1-e^{-B}}. \tag{D113}$$

Because $\left(e^{-\ell_i^t}\right)^\eta = e^{-\eta\ell_i^t}$, averaging this inequality over $\boldsymbol{q}$ gives

$$\sum_i q_i e^{-\eta\ell_i^t} \leq 1 - c_{\eta,B}\left(1 - \sum_i q_i e^{-\ell_i^t}\right). \tag{D114}$$

Because $c_{\eta,B} \geq 1$, convexity of the function $v^{c_{\eta,B}}$ and its tangent at $v = 1$ imply

$$\left(\sum_i q_i e^{-\ell_i^t}\right)^{c_{\eta,B}} \geq 1 - c_{\eta,B}\left(1 - \sum_i q_i e^{-\ell_i^t}\right). \tag{D115}$$

Combining these inequalities gives

$$\sum_i q_i e^{-\eta\ell_i^t} \leq \left(\sum_i q_i e^{-\ell_i^t}\right)^{c_{\eta,B}}. \tag{D116}$$

Taking negative logarithms now yields

$$l_\eta^t(\boldsymbol{q}) = -\log\left(\sum_i q_i e^{-\eta\ell_i^t}\right) \tag{D117}$$

$$\geq -\log\left[\left(\sum_i q_i e^{-\ell_i^t}\right)^{c_{\eta,B}}\right] \tag{D118}$$

$$= -c_{\eta,B}\log\left(\sum_i q_i e^{-\ell_i^t}\right) \tag{D119}$$

$$= c_{\eta,B}\ell^t(\boldsymbol{q}), \tag{D120}$$

which proves the second branch of Equation D105. □

The change at $\eta = 1$ reflects a genuine difference between weakening and strengthening a finite multiplicative update. For $\eta \leq 1$, concavity directly compares the weakened update with a proportional fraction of the actual growth loss. For $\eta \geq 1$, the bounded loss range is needed to control how strongly the amplified update can concentrate on the currently favored types.

Theorem 4. *Suppose that Equation D98 holds, and let* $\boldsymbol{q}$ *be any fixed composition satisfying* $\operatorname{supp}(\boldsymbol{q}) \subseteq \operatorname{supp}(\boldsymbol{p}^0)$. *The growth regret of the trajectory generated with learning rate* η *satisfies*

$$R_{\boldsymbol{q}}^T(\eta) \leq \begin{cases} \left(\dfrac{B}{1-e^{-B}} - 1\right) L_{\boldsymbol{q}}^T + \dfrac{1}{\eta} D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big), & 0 < \eta \leq 1, \\ \left(\dfrac{\eta B}{1-e^{-\eta B}} - 1\right) L_{\boldsymbol{q}}^T + \dfrac{1-e^{-B}}{1-e^{-\eta B}} D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big), & \eta \geq 1. \end{cases} \tag{D121}$$

At $\eta = 1$, *both branches reduce to*

$$R_{\boldsymbol{q}}^T(1) \leq D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big) + \left(\frac{B}{1-e^{-B}} - 1\right) L_{\boldsymbol{q}}^T, \tag{D122}$$

recovering the growth-regret bound for natural selection.

Proof. The change in potential information over one update is

$$D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}(\eta)\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t}(\eta)\big) \tag{D123}$$

$$= \sum_i q_i \log \frac{p_i^t(\eta)}{p_i^{t+1}(\eta)} \tag{D124}$$

$$= \eta \sum_i q_i \ell_i^t + \log \sum_i p_i^t(\eta) e^{-\eta \ell_i^t} \tag{D125}$$

$$= \eta \ell_{\text{ext}}^t(\boldsymbol{q}) - l_\eta^t\left(\boldsymbol{p}^t(\eta)\right). \tag{D126}$$

Equivalently,

$$l_\eta^t\left(\boldsymbol{p}^t(\eta)\right) = \eta \ell_{\text{ext}}^t(\boldsymbol{q}) + D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t}(\eta)\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}(\eta)\big). \tag{D127}$$

First suppose that $0 < \eta \leq 1$. By Equation D105,

$$\eta \ell^t(\boldsymbol{p}^t(\eta)) \leq \eta \ell_{\text{ext}}^t(\boldsymbol{q}) + D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t}(\eta)\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}(\eta)\big). \tag{D128}$$

Using

$$\ell_{\text{ext}}^t(\boldsymbol{q}) = \ell^t(\boldsymbol{q}) + J^t(\boldsymbol{q}) \tag{D129}$$

and rearranging gives

$$\ell^t(\boldsymbol{p}^t(\eta)) - \ell^t(\boldsymbol{q}) \leq J^t(\boldsymbol{q}) + \frac{1}{\eta}\left[D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t}(\eta)\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}(\eta)\big)\right]. \tag{D130}$$

Summing over $t = 0, \ldots, T-1$ telescopes the potential-information terms and gives

$$R_{\boldsymbol{q}}^T(\eta) \leq \sum_{t=0}^{T-1} J^t(\boldsymbol{q}) + \frac{1}{\eta}\left[D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{0}\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{T}(\eta)\big)\right]. \tag{D131}$$

The bounded-loss comparison established above gives

$$J^t(\boldsymbol{q}) \leq \left(\frac{B}{1-e^{-B}} - 1\right)\ell^t(\boldsymbol{q}). \tag{D132}$$

Applying this inequality in each generation therefore yields

$$R_{\boldsymbol{q}}^T(\eta) \leq \left(\frac{B}{1-e^{-B}} - 1\right) L_{\boldsymbol{q}}^T + \frac{1}{\eta}\left[D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{0}\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{T}(\eta)\big)\right] \tag{D133}$$

$$\leq \left(\frac{B}{1-e^{-B}} - 1\right) L_{\boldsymbol{q}}^T + \frac{1}{\eta} D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{0}\big), \tag{D134}$$

where the final inequality drops the nonpositive terminal-divergence term. This proves the first branch of Equation D121.

Now suppose that $\eta \geq 1$, and let

$$c_{\eta,B} := \frac{1-e^{-\eta B}}{1-e^{-B}}. \tag{D135}$$

Equations Equation D105 and Equation D127 give

$$c_{\eta,B}\,\ell^t(\boldsymbol{p}^t(\eta)) \leq \eta \ell_{\text{ext}}^t(\boldsymbol{q}) + D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t}(\eta)\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}(\eta)\big). \tag{D136}$$

The same bounded-loss argument used above implies

$$\ell_{\text{ext}}^t(\boldsymbol{q}) \leq \frac{B}{1-e^{-B}}\,\ell^t(\boldsymbol{q}). \tag{D137}$$

Substituting this bound, dividing by $c_{\eta,B}$, and subtracting $\ell^t(\boldsymbol{q})$ from both sides gives

$$\ell^t(\boldsymbol{p}^t(\eta)) - \ell^t(\boldsymbol{q}) \leq \left(\frac{\eta B}{1-e^{-\eta B}} - 1\right)\ell^t(\boldsymbol{q}) \tag{D138}$$

$$+ \frac{1-e^{-B}}{1-e^{-\eta B}}\left[D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t}(\eta)\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^{t+1}(\eta)\big)\right]. \tag{D139}$$

Summing over $t = 0, \dots, T-1$ again telescopes the potential-information terms:

$$R_{\boldsymbol{q}}^T(\eta) \le \left(\frac{\eta B}{1-e^{-\eta B}} - 1\right) L_{\boldsymbol{q}}^T + \frac{1-e^{-B}}{1-e^{-\eta B}} \left[D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big) - D\big(\boldsymbol{q}\big\|\boldsymbol{p}^T(\eta)\big)\right] \tag{D140}$$

$$\le \left(\frac{\eta B}{1-e^{-\eta B}} - 1\right) L_{\boldsymbol{q}}^T + \frac{1-e^{-B}}{1-e^{-\eta B}} D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big)\,. \tag{D141}$$

The final inequality again drops the nonpositive terminal-divergence term. This proves the second branch of Equation D121. Substituting $\eta = 1$ into either branch gives Equation D122. □

The two parts of the bound describe two different consequences of changing the responsiveness of the update. When $\eta < 1$, the population responds more conservatively to the observed fitness differences. This does not reduce the reference composition's cumulative growth loss, but it enlarges the initial-potential contribution from $D(\boldsymbol{q}\|\boldsymbol{p}^0)$ to $D(\boldsymbol{q}\|\boldsymbol{p}^0)/\eta$. The bound therefore records the effect of moving more slowly away from an initially mismatched composition.

When $\eta > 1$, stronger updates reduce the coefficient on the initial potential information but increase the coefficient on the reference compositions' cumulative growth loss. A population can move rapidly toward types that have performed well, so less of its regret is attributed to its initial distance from the reference composition, but an aggressive update can react strongly to a loss difference that is transient or that even the best fixed composition cannot consistently avoid. The learning-rate comparison therefore balances the effect of reallocating an initially mismatched population against the cost of responding too strongly to unavoidable or changing losses.

Applying the theorem to the hindsight-optimal composition gives the form most directly analogous to the hindsight external-regret comparison.

Corollary D0.1. *Let $\widetilde{\boldsymbol{p}}^T$ be a hindsight-optimal fixed composition for population growth loss, and define*

$$\widetilde{R}^T(\eta) := L^T(\eta) - \widetilde{L}^T. \tag{D142}$$

Under Equation D98,

$$\widetilde{R}^T(\eta) \le \begin{cases} \left(\dfrac{B}{1-e^{-B}} - 1\right) \widetilde{L}^T + \dfrac{1}{\eta} D\big(\widetilde{\boldsymbol{p}}^T\big\|\boldsymbol{p}^0\big)\,, & 0 < \eta \le 1, \\ \left(\dfrac{\eta B}{1-e^{-\eta B}} - 1\right) \widetilde{L}^T + \dfrac{1-e^{-B}}{1-e^{-\eta B}} D\big(\widetilde{\boldsymbol{p}}^T\big\|\boldsymbol{p}^0\big)\,, & \eta \ge 1. \end{cases} \tag{D143}$$

Proof. For the fixed horizon and common type-loss sequence, substitute $\boldsymbol{q} = \widetilde{\boldsymbol{p}}^T$ into Equation D121. Then use $L_{\widetilde{\boldsymbol{p}}^T}^T = \widetilde{L}^T$ and $R_{\widetilde{\boldsymbol{p}}^T}^T(\eta) = \widetilde{R}^T(\eta)$. □

For type losses in the range $[0, 1]$, this becomes

$$\widetilde{R}^T(\eta) \le \begin{cases} \dfrac{1}{e-1} \widetilde{L}^T + \dfrac{1}{\eta} D\big(\widetilde{\boldsymbol{p}}^T\big\|\boldsymbol{p}^0\big)\,, & 0 < \eta \le 1, \\ \left(\dfrac{\eta}{1-e^{-\eta}} - 1\right) \widetilde{L}^T + \dfrac{1-e^{-1}}{1-e^{-\eta}} D\big(\widetilde{\boldsymbol{p}}^T\big\|\boldsymbol{p}^0\big)\,, & \eta \ge 1. \end{cases} \tag{D144}$$

At $\eta = 1$, the two expressions meet at

$$\widetilde{R}^T(\eta = 1) \le D\big(\widetilde{\boldsymbol{p}}^T\big\|\boldsymbol{p}^0\big) + \frac{1}{e-1} \widetilde{L}^T, \tag{D145}$$

as required by the preceding discrete-time growth-regret theorem.

Learning rate minimizing the hindsight growth-regret bound

This learning-rate dependent bound also allows us to ask which learning rate gives the smallest upper bound on hindsight regret. For losses in the range $0 \le \ell_i^t \le B$, let $\mathfrak{R}_T(\eta)$ denote the right-hand side of Equation D143. Because the bound has different expressions below and above $\eta = 1$, we consider the two ranges separately.

For conservative learning rates $0 < \eta \le 1$,

$$\frac{d\mathfrak{R}_T(\eta)}{d\eta} = -\frac{D\big(\widetilde{\boldsymbol{p}}^T\big\|\boldsymbol{p}^0\big)}{\eta^2} \le 0. \tag{D146}$$

The bound therefore decreases, or remains unchanged, as η increases toward one. Among conservative rates in this range, $\eta = 1$ consequently gives the smallest bound. Intuitively, making the update more conservative slows movement away from the initial composition and increases the contribution of the initial potential information, and there is no compensating reduction in the cumulative loss term of this bound.

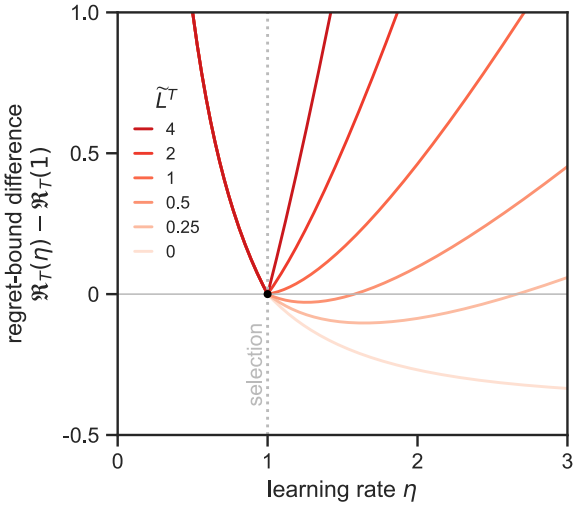


Fig. S1. The effect of learning rate on the worst-case regret bound for Multiplicative Weights Updating. For each fixed learning rate η, the figure shows the difference $\mathfrak{R}_T(\eta) - \mathfrak{R}_T(1)$ between the worst-case regret bound at that learning rate and the corresponding bound for natural selection at $\eta = 1$ (Equation D143). Each curve corresponds to a different cumulative mismatch load $\widetilde{L}^T$ of the hindsight-optimal composition. All curves assume $D(\widetilde{\boldsymbol{p}}^T \| \boldsymbol{p}^0) = 1$ and $\ell_{\max} = 1$. The dotted vertical line marks natural selection at $\eta = 1$, and the horizontal gray line marks equal regret bounds. Regret bound difference values above 0 indicate the learning rate η gives a higher regret bound than that for selection, whereas values below 0 indicate a lower bound. For $\eta < 1$, the curves overlap and the bound is always higher than that for selection. For $\eta > 1$, faster updating can lower the bound when $\widetilde{L}^T$ is small, but this advantage diminishes as $\widetilde{L}^T$ increases. Selection minimizes the bound when $\widetilde{L}^T / D(\widetilde{\boldsymbol{p}}^T \| \boldsymbol{p}^0) \geq (1 - e^{-1})/(e - 2) \approx 0.88$; among the cases shown, this holds for $\widetilde{L}^T = 1$, 2, and 4.

For aggressive learning rates $\eta \geq 1$, differentiation gives

$$\frac{d\mathfrak{R}_T(\eta)}{d\eta} = \frac{Be^{\eta B}}{\left(e^{\eta B} - 1\right)^2} \left[\widetilde{L}^T \left(e^{\eta B} - \eta B - 1 \right) - \left(1 - e^{-B} \right) D\left(\widetilde{\boldsymbol{p}}^T \middle\| \boldsymbol{p}^0 \right) \right]. \tag{D147}$$

The factor outside the brackets is positive, so the sign of the derivative is determined by the two terms inside the brackets. These terms express the tradeoff produced by increasing the learning rate. Increasing the learning rate reduces the coefficient on the initial divergence but increases the coefficient on $\widetilde{L}^T$. The bound therefore compares the benefit of reallocating more rapidly from an initially mismatched composition with the larger finite-update contribution controlled by the cumulative load of the fixed reference composition.

When $\widetilde{L}^T > 0$, an interior minimizing rate above one must satisfy

$$e^{\eta B} - \eta B - 1 = \left(1 - e^{-B} \right) \frac{D\left(\widetilde{\boldsymbol{p}}^T \middle\| \boldsymbol{p}^0 \right)}{\widetilde{L}^T}. \tag{D148}$$

The left-hand side is strictly increasing in $\eta > 0$, since its derivative is $B(e^{\eta B} - 1) > 0$. Thus, if Equation D148 has a solution above one, that solution is the unique minimizer in the aggressive range.

The right derivative of the bound at $\eta = 1$ is nonnegative when

$$\widetilde{L}^T \geq \frac{1 - e^{-B}}{e^B - B - 1} D\left(\widetilde{\boldsymbol{p}}^T \middle\| \boldsymbol{p}^0 \right). \tag{D149}$$

For instance, if $B = 1$, Equation D149 reduces to

$$\widetilde{L}^T \geq \frac{1 - e^{-1}}{e - 2} D\left(\widetilde{\boldsymbol{p}}^T \middle\| \boldsymbol{p}^0 \right) \approx 0.88\, D\left(\widetilde{\boldsymbol{p}}^T \middle\| \boldsymbol{p}^0 \right). \tag{D150}$$

When the cumulative loss of the hindsight-optimal composition is greater than the threshold given by Equation D149, the bound decreases as η approaches one from below and increases, or remains momentarily unchanged, as η rises above one. The unit learning rate of natural selection therefore minimizes the regret bound over all $\eta > 0$ in this regime where the hindsight-optimal composition has at least this much cumulative loss. Otherwise, if $\widetilde{L}^T$ is less than this threshold, the regret bound continues to decrease above $\eta = 1$. As η increases further, however, the term involving $\widetilde{L}^T$ eventually dominates, and the regret bound begins to increase.

These regimes have a direct interpretation. When the hindsight-optimal composition begins far from the initial population but experiences little cumulative growth loss, the bound can favor a rate above one because rapid reallocation reduces the contribution from the initial discrepancy. As $\widetilde{L}^T$ becomes large relative to this initial divergence, the increasing reference-load contribution outweighs that benefit, and the unit rate of selection minimizes the bound. Because $\widetilde{L}^T$ measures mismatch experienced by the best fixed composition, it may reflect ordinary mismatch between accessible types and environmental conditions, environmental variation across the sequence, or both.

The limiting case $\widetilde{L}^T = 0$ is different. If $D(\widetilde{\boldsymbol{p}}^T \| \boldsymbol{p}^0) > 0$, the bound decreases for every finite learning rate and approaches $(1 - e^{-B}) D(\widetilde{\boldsymbol{p}}^T \boldsymbol{p}^0)$ as $\eta \to \infty$, so no finite rate minimizes it. If both $\widetilde{L}^T = 0$ and $D(\widetilde{\boldsymbol{p}}^T \| \boldsymbol{p}^0) = 0$, the bound is zero for every learning rate.

These conclusions concern the worst-case upper bound, not necessarily the realized regret of a particular trajectory, and they do not establish a universally optimal biological learning rate. The comparison also requires the same loss sequence to be used for every value of η. If the environment depends on the evolving population, changing η generally changes both the population trajectory and the environmental sequence it generates. The resulting trajectories would then represent different learning problems rather than a controlled comparison of responsiveness on the same problem.

The result is specifically a discrete-generation comparison because its learning-rate dependence arises from finite multiplicative updates and their mix-loss curvature. In continuous time, multiplying a time-independent replicator vector field by a constant can be absorbed into a rescaling of time. Under weak selection with correspondingly scaled time, the finite-generation curvature corrections approach their continuous-time forms.

Relationship to standard external-regret bounds

The preceding results evaluate the counterfactual learning-rate trajectories using growth loss because growth regret directly measures differences in cumulative population growth. The MWU literature instead studies external regret, which is based on the frequency-weighted average of the type growth losses. Retaining this conventional comparison is useful for locating the present results relative to standard learning theory and to previous treatments of replicator dynamics as multiplicative updating (Kivinen and Warmuth 1995, Freund and Schapire 1999, Cesa-Bianchi and Lugosi 2006, Worden 1995, Chastain et al. 2014, Chastain 2017).

For type losses satisfying $0 \leq \ell_i^t \leq B$, the standard argument gives, for any admissible fixed composition $\boldsymbol{q}$ (e.g. Freund and Schapire 1999, Cesa-Bianchi and Lugosi 2006):

$$R^T_{\mathrm{ext},\boldsymbol{q}}(\eta) \leq \frac{B}{1-e^{-\eta B}} D\big(\boldsymbol{q}\big\|\boldsymbol{p}^0\big) + \left(\frac{\eta B}{1-e^{-\eta B}} - 1\right) L^T_{\mathrm{ext},\boldsymbol{q}}. \tag{D151}$$

This has the same broad interpretation as the growth-regret bound above. A smaller learning rate places a larger cost on moving away from the initial composition, whereas a larger learning rate can place a larger cost on responding aggressively to losses that the reference composition itself cannot avoid.

The external and growth formulations nevertheless use different loss functions and can select different hindsight-optimal compositions. Cumulative external loss is linear in the reference composition, so an external-optimal composition can always be chosen to be a pure type. For a pure reference composition $\boldsymbol{\delta}_i$, its Jensen gap is zero in every generation, and the nonnegative Jensen gaps of the evolving population imply $R^T_{\boldsymbol{\delta}_i}(\eta) \leq R^T_{\mathrm{ext},\boldsymbol{\delta}_i}(\eta)$. As a result, the standard external-regret bound also bounds growth regret relative to that same pure type.

This does not generally extend to hindsight growth regret, however. The composition $\widetilde{\boldsymbol{p}}^T$ that maximizes cumulative population growth may be mixed, and its Jensen gaps need not vanish. The direct growth-regret bound derived above is therefore the more relevant result for the paper's biological comparison.

Because the two bounds evaluate different losses and potentially different reference compositions, they also need not favor the same learning rate. In particular, the standard external-regret bound may be minimized by a rate below, equal to, or above one, whereas the direct growth-regret bound derived above is not improved by reducing the rate below one. These conclusions are not contradictory; they answer different performance questions.

D.4 Information gain and regret

Previous sections used potential information as an accounting variable for tracking performance relative to a fixed composition. We now turn to the information actually acquired along the evolutionary trajectory, $I^T = D(\boldsymbol{p}^T\|\boldsymbol{p}^0)$, and ask how it relates to improvements in population growth and to cumulative growth shortfalls. Although information gain and potential information are both KL divergences, they compare different distributions in different directions: information gain compares the current population with its initial state, whereas potential information compares the current population with a reference composition.

The results below connect these two informational views by showing how much information selection gains while changing the population's composition and how that information is bounded by differences in cumulative lineage growth. First we examine the information gained in a single discrete-time update, then the limiting relationship between information gain and regret when a stationary target is learned, and finally the general bounds that remain available when no stationary target exists.

D.4.1 Information gain in a single update

A selective update reweights the population toward types with greater effective fitness in the current environment. To isolate the population growth improvement produced by that reweighting, we can evaluate both the old and updated compositions against the same environmental distribution $\boldsymbol{x}^t$. This asks how much fitter the updated composition would be if it were used immediately under the conditions that generated the update. The next proposition shows that this look-ahead improvement involves two components: the KL information gained by the frequency update and a nonnegative Jensen correction.

Proposition D1. *Let $\boldsymbol{p}^{t+1}$ be obtained from $\boldsymbol{p}^t$ by one discrete-time selective update in environmental distribution $\boldsymbol{x}^t$. Then the information gained in that update is equal to the reduction in population growth loss obtained by replacing $\boldsymbol{p}^t$ with $\boldsymbol{p}^{t+1}$ in the same environment, minus*

the Jensen gap of the updated composition:

$$D\big(\boldsymbol{p}^{t+1}\big\|\boldsymbol{p}^{t}\big)=\ell^{t}(\boldsymbol{p}^{t})-\ell^{t}(\boldsymbol{p}^{t+1})-J^{t}(\boldsymbol{p}^{t+1}). \tag{D152}$$

Consequently, the one-update information gain is bounded above by this reduction in population growth loss:

$$D\big(\boldsymbol{p}^{t+1}\big\|\boldsymbol{p}^{t}\big)\le\ell^{t}(\boldsymbol{p}^{t})-\ell^{t}(\boldsymbol{p}^{t+1}). \tag{D153}$$

Proof. From Equation D37,

$$\frac{p_i^{t+1}}{p_i^{t}}=\frac{e^{-\ell_i^{t}}}{Z^{t}},\qquad\text{where }\log Z^{t}=-\ell^{t}(\boldsymbol{p}^{t}). \tag{D154}$$

Therefore,

$$\begin{aligned}D\big(\boldsymbol{p}^{t+1}\big\|\boldsymbol{p}^{t}\big)&=\sum_i p_i^{t+1}\log\frac{p_i^{t+1}}{p_i^{t}} &\text{[D155]}\\&=\sum_i p_i^{t+1}\left(-\ell_i^{t}-\log Z^{t}\right) &\text{[D156]}\\&=\ell^{t}(\boldsymbol{p}^{t})-\ell^{t}_{\mathrm{ext}}(\boldsymbol{p}^{t+1}). &\text{[D157]}\end{aligned}$$

Using $\ell^{t}_{\mathrm{ext}}(\boldsymbol{p}^{t+1})=\ell^{t}(\boldsymbol{p}^{t+1})+J^{t}(\boldsymbol{p}^{t+1})$ gives Equation D152. Finally, $J^{t}(\boldsymbol{p}^{t+1})\ge 0$ by Jensen's inequality, so removing it from the right-hand side gives Equation D153. □

The identity decomposes this one-step improvement in population growth loss into the information gained by the update and the nonnegative Jensen gap of the updated composition. Selection therefore cannot acquire more information in one discrete update than the improvement in per-generation population growth rate that the update produces when evaluated against the same environmental distribution. The right-hand side is a one-step look-ahead comparison, not the difference between the losses actually realized in rounds t and $t+1$ when the environment may change.

There is no directly analogous first-order identity for an infinitesimal continuous-time update. Over an interval of length dt, the change in composition is $O(dt)$, but the divergence between the neighboring compositions $D(\boldsymbol{p}^{t+dt}\|\boldsymbol{p}^{t})$ is $O(dt^{2})$. The cumulative continuous-time relationships therefore arise from changes in potential information over a finite interval, as developed below, rather than by integrating an instantaneous version of Equation D152.

Connection to Hledík et al.'s finite-population bound.

Hledík et al. measure how selection biases parental sampling relative to neutral reproduction (Hledík et al. 2022). Conditional on the current composition $\boldsymbol{p}^{t}$, neutrality samples type i with probability p_i^{t}, whereas selection samples it with probability $p_i^{t}\widehat{w}_i^{t}=p_i^{t+1}$, where

$$\widehat{w}_i^{t}:=\frac{\langle W_i\rangle_{\boldsymbol{x}^{t}}}{\langle W\rangle_{\boldsymbol{p}^{t},\boldsymbol{x}^{t}}} \tag{D158}$$

is relative fitness. Their cost per parental sample is therefore exactly our single-update information gain, expressed in bits:

$$C_{\mathrm{H}}^{t}:=\sum_i p_i^{t+1}\log_2\widehat{w}_i^{t}=\frac{D\left(\boldsymbol{p}^{t+1}\big\|\boldsymbol{p}^{t}\right)}{\ln 2}. \tag{D159}$$

Their relative fitness variance,

$$V^{t}:=\sum_i p_i^{t}\left(\widehat{w}_i^{t}-1\right)^{2}, \tag{D160}$$

also has a direct growth interpretation. Evaluated in the same environment, the selective update increases mean fitness by the factor

$$\frac{\langle W\rangle_{\boldsymbol{p}^{t+1},\boldsymbol{x}^{t}}}{\langle W\rangle_{\boldsymbol{p}^{t},\boldsymbol{x}^{t}}}=1+V^{t}. \tag{D161}$$

Combining this identity with Equation D152 gives

$$\begin{aligned}C_{\mathrm{H}}^{t}&=\frac{\log\left(1+V^{t}\right)-J^{t}\left(\boldsymbol{p}^{t+1}\right)}{\ln 2} &\text{[D162]}\\&\le\log_2\left(1+V^{t}\right)\le\frac{V^{t}}{\ln 2}. &\text{[D163]}\end{aligned}$$

The final inequality recovers Hledík et al.'s variance bound, while the intermediate term provides a tighter bound whenever fitness varies. Under weak selection, writing $\widehat{w}_i^t = 1 + \delta_i^t$ gives

$$C_{\mathrm{H}}^t = \frac{V^t}{2\ln 2} + O\left(\left\|\boldsymbol{\delta}^t\right\|_\infty^3\right), \tag{D164}$$

recovering their weak-selection relationship. To second order, the information gained and the Jensen correction each account for half of the growth improvement produced by the update. Hledík et al.'s finite-population result averages this per-sample quantity over stochastic population states and multiplies it by the number kN of parents sampled under selection in each generation.

D.4.2 Information gain and target regret

Over a longer trajectory, $I^T = D(\boldsymbol{p}^T \| \boldsymbol{p}^0)$ measures the net information gained between the initial and current compositions. It is not defined as the sum of the one-update divergences $D(\boldsymbol{p}^{t+1} \| \boldsymbol{p}^t)$. The one-step result therefore cannot simply be summed to obtain total information gain. Instead, in a stationary-target setting we use convergence of the population itself: once $\boldsymbol{p}^T$ approaches $\overline{\boldsymbol{p}}$, the endpoint divergence defining information gain approaches the initial potential information relative to that target. Combining this endpoint argument with the target-regret results from Appendix D.3.2 yields separate continuous- and discrete-time conclusions.

Continuous time

Theorem 5. *Let $\overline{\boldsymbol{p}}$ be a stationary learning target for a population beginning at $\boldsymbol{p}^0$. Under continuous-time replicator dynamics, both total information gain and target regret converge to the initial potential information:*

$$\lim_{T\to\infty} I^T = \lim_{T\to\infty} \bar{R}^T = D\left(\overline{\boldsymbol{p}} \middle\| \boldsymbol{p}^0\right). \tag{D165}$$

Proof. Because $\boldsymbol{p}^T \to \overline{\boldsymbol{p}}$ and $\mathrm{supp}(\overline{\boldsymbol{p}}) \subseteq \mathrm{supp}(\boldsymbol{p}^0)$, continuity of KL divergence in its first argument gives

$$\lim_{T\to\infty} I^T = \lim_{T\to\infty} D\left(\boldsymbol{p}^T \middle\| \boldsymbol{p}^0\right) = D\left(\overline{\boldsymbol{p}} \middle\| \boldsymbol{p}^0\right). \tag{D166}$$

Independently, Equation D52 gives

$$\lim_{T\to\infty} \bar{R}^T = D\left(\overline{\boldsymbol{p}} \middle\| \boldsymbol{p}^0\right). \tag{D167}$$

Equating these two limits proves Equation D165. □

At convergence, information gain and continuous-time target regret therefore reduce to the same initial-to-target divergence. Information gain records the endpoint's departure from the initial composition, while the potential identity records the cumulative disappearance of the discrepancy between the population and its target. Their equality is an endpoint result. This does not imply that information gain and target regret are equal at every intermediate time.

Discrete time

Finite discrete updates retain the same endpoint information gain but can add a cumulative Jensen correction to target regret. The next theorem states both the general relation and the conditions under which the continuous-time equality is recovered.

Theorem 6. *Let $\overline{\boldsymbol{p}}$ be a stationary learning target for a population beginning at $\boldsymbol{p}^0$. Under discrete-time replicator dynamics, total information gain converges to the initial potential information:*

$$\lim_{T\to\infty} I^T = D\left(\overline{\boldsymbol{p}} \middle\| \boldsymbol{p}^0\right). \tag{D168}$$

If the target's Jensen gaps have a finite cumulative sum, then

$$\lim_{T\to\infty} \bar{R}^T = \lim_{T\to\infty} I^T + \sum_{t=0}^{\infty} J^t\left(\overline{\boldsymbol{p}}\right). \tag{D169}$$

In particular, if the target has zero Jensen gap along the trajectory, including whenever the stationary learning target is pure, then

$$\lim_{T\to\infty} I^T = \lim_{T\to\infty} \bar{R}^T = D\left(\overline{\boldsymbol{p}} \middle\| \boldsymbol{p}^0\right). \tag{D170}$$

For general stationary targets, this equality is recovered in the weak-selection limit under the conditions of Corollary D0.2:

$$\lim_{\varepsilon\to 0}\lim_{T\to\infty} I^{T,\varepsilon} = \lim_{\varepsilon\to 0}\lim_{T\to\infty} \bar{R}^{T,\varepsilon} = D\left(\overline{\boldsymbol{p}} \middle\| \boldsymbol{p}^0\right). \tag{D171}$$

Proof. The convergence $\boldsymbol{p}^T \to \overline{\boldsymbol{p}}$ and continuity of $D(\boldsymbol{p}^T \| \boldsymbol{p}^0)$ give Equation D168 by the same endpoint argument used in the continuous-time proof.

When the cumulative target gap is finite, Equation D56 gives

$$\lim_{T\to\infty} \bar{R}^T = D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^0\big) + \sum_{t=0}^{\infty} J^t\left(\overline{\boldsymbol{p}}\right). \tag{D172}$$

Substituting Equation D168 proves Equation D169. The zero-gap equality then follows from Equation D59.

Under the weak-selection assumptions, Equation D68 shows that the cumulative Jensen correction vanishes. For every ε, convergence to the same target gives

$$\lim_{T\to\infty} I^{T,\varepsilon} = D\big(\overline{\boldsymbol{p}}\big\|\boldsymbol{p}^0\big). \tag{D173}$$

Taking $\varepsilon \to 0$ therefore proves Equation D171. □

For a general mixed discrete-time target, target regret can therefore exceed the endpoint information gain by the cumulative target Jensen gap. The excess reflects the finite size of the discrete updates and the distinction between averaging type growth rates and taking the growth rate of the mixed population as a whole. It disappears exactly in the zero-gap case and asymptotically under the stated weak-selection conditions.

Information gain and substitution load

The general results above recover the classical relationship between information gain and substitution load when a single initially present type is optimal. The continuous- and discrete-time forms are stated separately below.

Corollary D0.1. *Consider a constant, homogeneous environment in which an initially present type has maximal Malthusian fitness $r_* := \max_i r_i$. Under continuous-time replicator dynamics, the substitution load through time T is given by*

$$L^T_{\mathrm{sub}} := \int_0^T \left(r_* - \langle r\rangle_{\boldsymbol{p}^t}\right) dt\,, \tag{D174}$$

and information gain is bounded above by substitution load:

$$I^T \le L^T_{\mathrm{sub}}. \tag{D175}$$

Moreover, if selection carries the optimal type to fixation, then information gain and substitution load converge to the same value:

$$\lim_{T\to\infty} I^T = \lim_{T\to\infty} L^T_{\mathrm{sub}} = -\log p^0_*. \tag{D176}$$

Proof. The pure composition $\boldsymbol{\delta}_*$ has zero continuous-time population growth loss in every instant. It is therefore a hindsight-optimal fixed reference composition, so $\widetilde{L}^T = 0$ and $\widetilde{R}^T = L^T = L^T_{\mathrm{sub}}$. Substituting this equality into Equation D182 gives Equation D175.

The continuous-time replicator equation for the optimal type gives

$$\frac{d}{dt}\log p^t_* = r_* - \langle r\rangle_{\boldsymbol{p}^t}. \tag{D177}$$

Integrating over $[0,T]$ yields $L^T_{\mathrm{sub}} = \log(p^T_*/p^0_*)$. If $p^T_* \to 1$, this identity gives $L^T_{\mathrm{sub}} \to -\log p^0_*$. The same convergence gives

$$I^T = D\big(\boldsymbol{p}^T\big\|\boldsymbol{p}^0\big) \longrightarrow D\big(\boldsymbol{\delta}_*\big\|\boldsymbol{p}^0\big) = -\log p^0_*, \tag{D178}$$

which proves Equation D176. □

Corollary D0.2. *If the environment consists of a single unchanging condition and an initially present type $*$ has maximal Wrightian fitness, then under discrete-time replicator dynamics information gain is bounded above by substitution load:*

$$I^T \le L^T_{\mathrm{sub}}. \tag{D179}$$

If selection carries the optimal type to fixation, information gain and substitution load converge to the same value:

$$\lim_{T\to\infty} I^T = \lim_{T\to\infty} L^T_{\mathrm{sub}} = -\log p^0_*. \tag{D180}$$

Proof. In a constant, homogeneous environment, the fixed pure composition $\boldsymbol{\delta}_*$ has zero population growth loss in every round. It is therefore a hindsight-optimal reference composition with $\widetilde{L}^T = 0$ and $\widetilde{R}^T = L^T = L^T_{\mathrm{sub}}$. Substituting this equality into Equation D187 gives Equation D179.

If the optimal type fixes, then $\boldsymbol{p}^T \to \boldsymbol{\delta}_*$, and thus

$$\lim_{T\to\infty} I^T = D\big(\boldsymbol{\delta}_*\big\|\boldsymbol{p}^0\big) = -\log p^0_*. \tag{D181}$$

The substitution-load identity Equation D6 gives the same limit for L^T_{sub}, proving **??**. □

This special case shows how the general hierarchy collapses when a feasible fixed composition attains the condition-wise oracle. During substitution, load can exceed the information acquired so far; once the optimal type fixes, both quantities record the same change from its initial frequency p^0_* to frequency one.

D.4.3 Information gain and hindsight regret

Even when there is no stationary endpoint, the terminal composition $\boldsymbol{p}^T$ itself provides a useful bridge between information gain and regret. We can ask how much better a counterfactual lineage would have grown if it had used this final composition from the beginning rather than reaching it gradually through selection. The hindsight-optimal fixed composition can only provide a stronger comparison, because its cumulative loss is less than or equal to that of $\boldsymbol{p}^T$.

Continuous time

Theorem 7. *For every trajectory $\boldsymbol{p}^t$, $t \in [0, T]$, generated by continuous-time replicator dynamics, total information gain is bounded above by hindsight regret, which is itself bounded above by cumulative mismatch load:*

$$I^T \leq \widetilde{R}^T \leq L^T. \tag{D182}$$

Proof. Continuous-time replicator dynamics preserve the initial support, so the terminal composition $\boldsymbol{p}^T$ is an admissible fixed reference composition. Set $\boldsymbol{q} = \boldsymbol{p}^T$ in Equation D27:

$$R^T_{\boldsymbol{p}^T} = D\big(\boldsymbol{p}^T \big\| \boldsymbol{p}^0\big) - D\big(\boldsymbol{p}^T \big\| \boldsymbol{p}^T\big) \tag{D183}$$

$$= I^T. \tag{D184}$$

By the definition of the hindsight-optimal composition, $\widetilde{L}^T \leq L^T_{\boldsymbol{p}^T}$. Subtracting each side from the same population loss L^T reverses the ordering:

$$\widetilde{R}^T = L^T - \widetilde{L}^T \geq L^T - L^T_{\boldsymbol{p}^T} = R^T_{\boldsymbol{p}^T} = I^T. \tag{D185}$$

This proves the first inequality.

The condition-wise oracle has at least as great an instantaneous growth rate as any fixed composition, so $\ell^t(\widetilde{\boldsymbol{p}}^T) \geq 0$ and thus $\widetilde{L}^T \geq 0$. Therefore,

$$\widetilde{R}^T = L^T - \widetilde{L}^T \leq L^T, \tag{D186}$$

which proves the second inequality. □

In continuous time, regret relative to the fixed terminal composition equals the information gained exactly. The hindsight-optimal composition performs at least as well as that terminal composition, so hindsight regret is at least as large as information gain. Mismatch load is larger still because it also includes the loss that the hindsight-optimal fixed composition itself experiences relative to the generally unattainable condition-wise oracle.

Discrete time

Theorem 8. *For every trajectory $\boldsymbol{p}^0, \ldots, \boldsymbol{p}^T$ generated by discrete-time replicator dynamics, total information gain is bounded above by hindsight regret, which is itself bounded above by cumulative mismatch load:*

$$I^T \leq \widetilde{R}^T \leq L^T. \tag{D187}$$

Proof. Discrete-time replicator dynamics preserve the initial support, so $\boldsymbol{p}^T$ is an admissible fixed reference composition. Set $\boldsymbol{q} = \boldsymbol{p}^T$ in Equation D36:

$$R^T_{\boldsymbol{p}^T} = D\big(\boldsymbol{p}^T \big\| \boldsymbol{p}^0\big) - D\big(\boldsymbol{p}^T \big\| \boldsymbol{p}^T\big) + \sum_{t=0}^{T-1} J^t(\boldsymbol{p}^T) \tag{D188}$$

$$= I^T + \sum_{t=0}^{T-1} J^t(\boldsymbol{p}^T) \tag{D189}$$

$$\geq I^T. \tag{D190}$$

The second line uses $D(\boldsymbol{p}^T \| \boldsymbol{p}^T) = 0$ and the definition $I^T = D(\boldsymbol{p}^T \| \boldsymbol{p}^0)$; the inequality uses nonnegativity of every Jensen gap.

By the definition of the hindsight-optimal composition, $\widetilde{L}^T \leq L^T_{\boldsymbol{p}^T}$. It follows that

$$\widetilde{R}^T = L^T - \widetilde{L}^T \geq L^T - L^T_{\boldsymbol{p}^T} = R^T_{\boldsymbol{p}^T} \geq I^T, \tag{D191}$$

which proves the first inequality.

Finally, population growth loss is nonnegative, so $\widetilde{L}^T \geq 0$. Therefore,

$$\widetilde{R}^T = L^T - \widetilde{L}^T \leq L^T, \tag{D192}$$

which proves the second inequality. □

The discrete-time argument differs only in the comparison with the terminal composition. Regret relative to $\boldsymbol{p}^T$ equals information gain plus the cumulative Jensen gaps obtained by evaluating that fixed terminal composition across the earlier environmental sequence. This makes the first inequality at least as strong as in continuous time.

The two inequalities have distinct biological interpretations. The first states that the information encoded in the population's compositional change cannot exceed its cumulative growth shortfall relative to the best fixed composition in hindsight. The second states that hindsight regret cannot exceed total mismatch load, because regret subtracts the mismatch load experienced by the fixed reference composition itself. Mismatch load thus includes both the shortfall that no admissible fixed composition could avoid and the additional shortfall associated with the population's evolutionary trajectory. Hindsight regret isolates the latter, and information gain is bounded by it.

D.5 Supplemental example learning problems

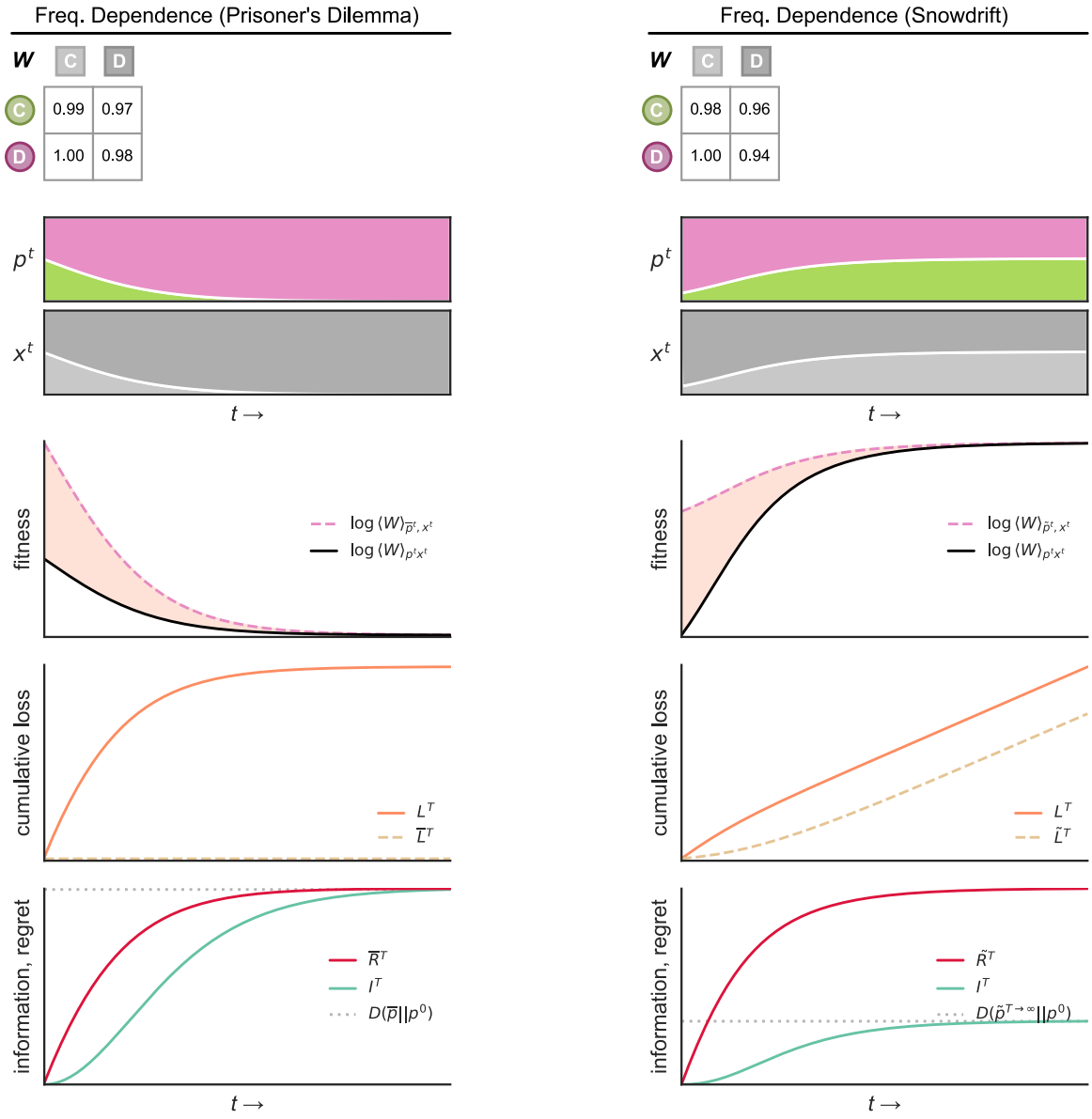


Fig. S2. Load, regret, and information gain under frequency-dependent selection. Replicator-dynamics simulations are shown for two additional contexts (columns). In each, the matrix $\mathbf{W}$ gives the fitness W_{ij} of each of two types in each of two environmental conditions. Colored Muller plots show the population composition $\boldsymbol{p}^t$ over time, and stacked grayscale plots show the environmental distribution $\boldsymbol{x}^t$. The middle panels show the cumulative mismatch load of the evolving population (L^T, orange) and the stationary target (ESS) composition ($\bar{L}^T$, dashed gold). The bottom panels show information gain (teal) and target regret relative to the target ($\bar{R}^T$, red). In both games, the environmental conditions are encounters with the types *Cooperate* and *Defect*, with $\boldsymbol{x}^t = \boldsymbol{p}^t$ in each generation. **(a) Prisoner's Dilemma (left).** Although all-cooperation maximizes population mean fitness, defectors have greater fitness at every composition. Replicator dynamics therefore drive the population toward all-defection, reducing mean fitness along the realized trajectory. Against that same realized sequence of environments, however, all-defection is the hindsight-optimal fixed composition. It is also a stationary learning target (ESS), and selection consequently achieves vanishing average target regret. **(b) Snowdrift (right).** Replicator dynamics carry the population toward a polymorphic stationary learning target (ESS) at which cooperators and defectors have equal effective fitness. Because defectors predominate along the realized trajectory before equilibrium, cooperation has greater effective fitness during this transient, making all-cooperation the hindsight-optimal fixed composition for the realized environmental sequence. Thus, the stationary target and the hindsight-optimal composition differ.

Appendix E: Selection Experiments

E.1 Strain information

Escherichia coli B (REL606) strains were used in selection experiments and related assays. A "wild type" strain (WT) and three strains with unique mutations in the *rpoB* gene (M1, M2, M3) were obtained with permission from the $-80°$C strain archive from Lindsey et al. (2013). Mutations to the *rpoB* gene gave each mutant strain a distinct exponential growth rate that was reduced from that of the WT strain (Supplementary Figure S3).

Each strain was transformed with an engineered marker plasmid. The pBR322 plasmid, which carries a *bla* gene conferring ampicillin resistance (Amp^R) and a *tetA* gene conferring tetracycline resistance (Tet^R), served as a vector backbone. Using Gibson assembly, the *bla* gene and corresponding promoter region were removed and replaced with an insert carrying a fluorescent protein gene under a strong constitutive *proC* promoter. As the strain with the optimal growth rate, the WT strain was transformed with a plasmid engineered to carry the green fluorescent protein (GFP) gene *mGFPmut2*. The *rpoB* mutant strains were each transformed with plasmids engineered to carry the red fluorescent protein (RFP) gene *mScarlet-I*.

Strains and culture samples were stored by mixing 1 mL of culture with 160 μL of 80% glycerol and freezing at $-80°$C. Strains were revived from freezer scrapings and incubated overnight prior to all assays and experiments.

	Chromosomal Genotype			Plasmid			
Strain ID	*E. coli* strain	Mutations	Selective Markers	Plasmid ID	Vector Backbone	FP Marker Gene Insert	Selective Markers
WT	B (REL606)	*See Table S2*	-	pRMbcGFP	pBR322	pEB1-mGFPmut2	Tet^R
M1	B (REL606)	*See Table S2*	Rif^R	pRMbcRFP	pBR322	pEB2-mScarlet-I	Tet^R
M2	B (REL606)	*See Table S2*	Rif^R	pRMbcRFP	pBR322	pEB2-mScarlet-I	Tet^R
M3	B (REL606)	*See Table S2*	Rif^R	pRMbcRFP	pBR322	pEB2-mScarlet-I	Tet^R

Table S1 Basic characteristics of bacterial strains used in selection experiments.

WT	M1	M2	M3	Position	Gene	Mutation	Annotation	Description
×	×	×	×	3,925,216	*gppA* ←	C→T	L272L (CTG→CTA)	guanosine pentaphosphatase
×				458,185	*acrR* →	IS1 (+) +9 bp	Coding (358 366/648 nt)	DNA binding transcriptional repressor
	×			4,162,387	*rpoB* →	T→G	L511R (CTG→CGG)	DNA-directed RNA polymerase subunit β
		×		4,161,298	*rpoB* →	A→T	Q148L (CAG→CTG)	DNA-directed RNA polymerase subunit β
		×		4,162,382	*rpoB* →	C→A	S509R (AGC→AGA)	DNA-directed RNA polymerase subunit β
			×	4,161,292	*rpoB* →	T→A	V146D (GTT→GAT)	DNA-directed RNA polymerase subunit β
			×	4,162,540	*rpoB* →	A→C	E562A (GAA→GCA)	DNA-directed RNA polymerase subunit β

Table S2 Mutations present in each bacterial strain relative to the *Escherichia coli* B (REL606) reference genome (presence of a mutation in a given strain is indicated by an X in that strain's column).

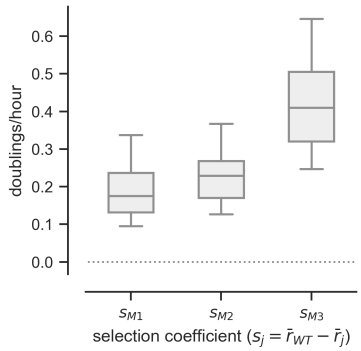


Fig. S3. Selection coefficients. The selection coefficient is a measure of the fitness of a type (i.e., strain) relative to another. In Malthusian fitness terms, the selection coefficient of a type j is defined as the difference between the exponential growth rate of the optimal type and that of the jth strain; that is, $s_j = r_* - r_j$ (Appendix A.1.2). In our experimental system, the WT strain had the fastest growth rate. Therefore the selection coefficient of a strain j was defined as $s_j = r_{\text{WT}} - r_j$. A positive selection coefficient $s_j > 0$ indicates that the growth rate of the jth strain is less than that of the WT strain. By definition, the selection coefficient of the WT strain is zero ($s_{\text{WT}} = 0$). Here we show the selection coefficients of the mutant strains (M1, M2, M3) relative to the WT strain as measured by an exponential growth assay. Individual strain cultures were sampled and transferred to fresh media to maintain exponential growth every 2 hours for a total of 8 hours. Cell enumeration was performed on culture samples using flow cytometry, and strain growth rates were calculated from changes in culture density over the course of the assay. Selection coefficients were then calculated according to the definition given above. These selection coefficients reflect the relative growth rates of strains after being transformed with the corresponding marker plasmids.

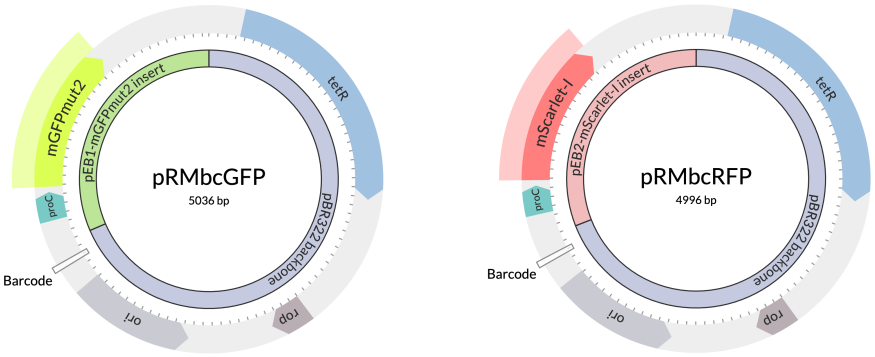


Fig. S4. Engineered plasmid maps. Each strain was transformed with an engineered marker plasmid. The pBR322 plasmid, which carries a *bla* gene conferring ampicillin resistance (Amp^R) and a *tetA* gene conferring tetracycline resistance (Tet^R), served as a vector backbone. Using Gibson assembly, the *bla* gene and corresponding promoter region was removed and replaced with an insert carrying a fluorescent protein gene under a constitutive *proC* promoter. **Left:** The pRMbcGFP plasmid carries a region encoding the green fluorescent protein gene *mGFPmut2* under the *proC* promoter, which was taken from the pEB1-mGFPmut2 plasmid (Addgene plasmid #103980). **Right:** The pRMbcRFP plasmid carries a region encoding the red fluorescent protein gene *mScarlet-I* under the *proC* promoter, which was taken from the pEB2-mScarlet-I plasmid (Addgene plasmid #104007).

E.2 Supplemental methods

Selection competition experiments were conducted between three pairs of strains: WT vs. M1, WT vs. M2, and WT vs. M3. Each pair of strains participated in three competitions with different initial strain frequency compositions: WT at approximately 12.5%, 25%, and 50% of the initial population, respectively.

Strains were cultured in standard Luria-Bertani (LB) broth with 15μg/mL tetracycline for plasmid retention (hereafter referred to as "media"). Cultures were incubated in 25 mL of media in 125 mL capacity flasks at 37°C with shaking. A single large batch of media was prepared and used for all cultures across all stages of all experiments, and all competitions were conducted simultaneously using the same incubator in order to mitigate potential day or block effects.

Strains were revived and incubated overnight to saturation density (~1×10^9 CFU/mL). Prior to the start of competitions, individual strain cultures were diluted 1000-fold into 25 mL of fresh media (~1×10^6 CFU/mL) and incubated for 1 hour, which was sufficient time for cells to begin resuming exponential growth. At the end of this "reanimation" period, the density of each strain culture was spot-checked using flow cytometry. Selection experiments were initiated by combining the strains to be competed in the designated ratios at a total density of ~1×10^6 CFU/mL in 25 mL of fresh media. Because the strains had entered exponential growth phase prior to initiating competitions, some growth occurred during the initiation process, which caused minor imprecisions in the initial strain frequencies.

Every 4 hours, a 1 mL sample was taken from each competition culture to measure strain densities and frequencies using flow cytometry (see Appendix E.2.2 for more information). At the same time, a sample of the culture was transferred to 25 mL of fresh media. An adaptive protocol was used to determine the transfer volume for each competition culture at each transfer in order to ensure that the culture would remain in exponential growth while maintaining a measurable density for flow cytometry measurements in the next growth interval (see Appendix E.2.1 for more information). Competitions proceeded in this fashion for 36 hours.

E.2.1 Transfer protocol

Selection experiments involved incubating mixed batch cultures and tracking the densities and frequencies of the competing strains for a total of 36 hours, which was the time necessary to observe the optimal WT strain approaching fixation in all competitions. These strains enter a stationary phase as the culture approaches a saturation density of approximately 1×10^9 CFU/mL, so regular transfers to fresh media were necessary to maintain active growth and selection over this duration. Basic transfer protocols often involve repeatedly growing cultures to saturation and transferring a sample of the saturated culture to fresh media at low density (e.g., ~1×10^5 CFU/mL). However, we were interested in measuring the outcomes of selection due to differential exponential growth rates, and thus we wished to maintain cultures in exponential phase to avoid the potential confounding effects of cells entering and exiting lag phase. In addition, highly accurate density estimation using flow cytometry required a culture density of at least ~1×10^7 CFU/mL at the time of sampling. Therefore, in order to ensure that the culture would remain in exponential growth while maintaining a measurable density for the flow cytometer we required a transfer protocol that would reliably maintain culture densities between ~$1 \times 10^7 - 1 \times 10^8$ CFU/mL at the end of each growth interval. The following transfer protocol was developed to dynamically calculate the transfer volume for each competition culture in order to achieve this (note that all competitions were conducted simultaneously).

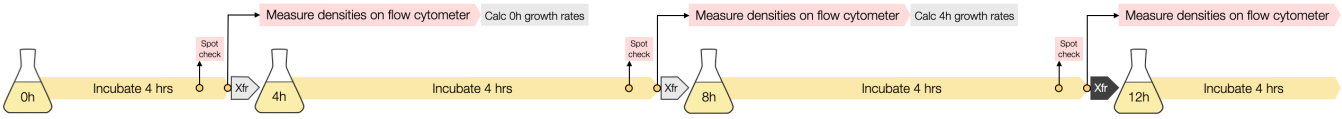


Fig. S5. Illustration of transfer protocol.

To make things concrete, we outline the transfer protocol from the perspective of an experimenter determining transfer volumes at the 12 hour mark of the overall experiment (black-shaded block arrow in Supplementary Figure S5), although the same process was used at all transfer points throughout the 36 hour experiment. Flasks are labeled by the time point at which they were inoculated (e.g., the '0h' culture flask was inoculated at the beginning of the experiment, the '4h' culture flask was inoculated via transfer at the 4 hour mark, and so on).

- **4 hours prior to transfer:** Data collection samples are taken from each competition, and competition cultures are transferred to fresh media.
    1. Samples are taken from the previous set of culture flasks (e.g., the '4h' cultures) for density estimation on the flow cytometer.
    2. The current set of flasks (e.g., the '8h' cultures) are inoculated via transfer and their incubation period begins.
    3. Cytometry samples (e.g., of the '4h' cultures) are prepared and measured over the following 1-2 hours.
    4. The newly obtained cytometry measurements provide endpoint density estimates for the previous set of culture flasks (e.g., the '4h' cultures). The initial density of these same cultures is estimated from the endpoint density of the prior set of cultures (e.g., obtained from cytometry measurements of the '0h' cultures) and the transfer volumes used at that time. The growth

rates of each of the previous cultures (e.g., the '4h' cultures) are estimated using the initial and endpoint densities for this interval (e.g., the 4h-8h interval).

- **30 minutes prior to transfer:** A "spot check" is performed on a subset of the current flasks in order to obtain a rough estimate of the densities and growth rates of the currently incubating cultures.
  1. Samples are quickly taken from the competition cultures with the lowest growth rate and the highest growth rate observed in the latest set of cytometry-based growth rate estimates (e.g., from the '4h' cultures). These flasks are immediately returned to the incubator to finish the rest of the incubation period.
  2. These samples are enumerated on the flow cytometer and used to estimate the endpoint densities and growth rates of the presumed fastest and slowest growing competition cultures in the current incubation period. This provides a presumed minimum and maximum growth rate for all competitions in the current growth interval.
  3. The endpoint densities for all competition cultures in the current growth interval are projected based on the initial densities of these cultures (using the previous cultures' endpoint densities and the transfer volumes involved in their inoculations) and the average of the growth rates estimated from the spot check measurement described above.
- **10 minutes prior to transfer:** The transfer volume to be used for each competition culture is calculated as follows (a spreadsheet was programmed to perform these calculations automatically using flow cytometry data collected previously in the experiment):
  - For each competition culture, working in order of highest projected endpoint density (from #3 above) to lowest:
    1. If this is the first competition considered, let the "trial" transfer volume for this competition culture be a default volume of 250 μL (i.e., a 100-fold dilution into 25 mL of media); else, let the "trial" transfer volume be the actual transfer volume of the previously considered competition.
    2. Calculate the projected initial density of the post-transfer culture based on the projected final density of the current culture and the "trial" transfer volume.
    3. Calculate the projected endpoint density of this competition culture in the next incubation period based on the projected initial density and the *minimum* growth rate that has been observed for any culture so far. Likewise, calculate the projected endpoint density of this competition culture in the next incubation period using the *maximum* growth rate that has been observed for any culture so far. This gives the range of plausible endpoint densities the post-transfer culture may realize if the "trial" transfer volume is used.
    4. If the range of plausible endpoint densities using the "trial" transfer volume falls within the acceptable range of endpoint densities (i.e., $1 \times 10^7 - 1 \times 10^8$ CFU/mL), then use the "trial" transfer volume to transfer this competition culture; else, do the following:
       * Backcalculate the initial density that would lead this culture to just reach the *minimum* acceptable endpoint density (i.e., 1×10^7 CFU/mL) if the culture were to grow at the *minimum* growth rate that has been observed for any culture so far. This is the smallest post-transfer initial density that is ensured to reach the minimum acceptable endpoint density in a worst-case, slow growth scenario (based on the range of growth rates observed to this point). (This heuristic is used because we deemed it more important to guarantee that cultures reach accurately measurable densities than to guarantee that cultures do not approach saturation.)
       * Calculate the transfer volume that would give this initial density based on the projected endpoint density for this culture in the current incubation period.
       * Use this "backcalculated" transfer volume to transfer this competition culture.
- **At time of transfer:** Data collection samples are taken from each competition, and competition cultures are transferred to fresh media using the calculated transfer volumes. The cycle repeats for the following incubation period and transfer.

This transfer protocol successfully maintained endpoint densities for all competition cultures within the range $8.3 \times 10^6 - 2.7 \times 10^8$ CFU/mL throughout the entire 36 hour experiment, with only a few culture endpoints falling outside of the tight target range of $1 \times 10^7 - 1 \times 10^8$ CFU/mL. All culture samples had densities that were accurately measurable on the flow cytometer, and no competition culture showed signs of cells exiting exponential phase to any appreciable degree.

E.2.2 Flow cytometry

Flow cytometry was used to enumerate cells in culture samples and estimate culture densities. Fixed fluorescence gates were established to separately enumerate cells expressing GFP and RFP markers. Culture samples were immediately centrifuged at 13,000 rpm for 5 minutes to remove growth media supernatant. Pelleted cells were then resuspended in 1 mL of flow buffer (1X PBST with EDTA added at a 500-fold dilution). Samples were then transferred to a round-bottom 96-well plate where a 10-fold dilution series was performed. The accuracy of cell enumeration was maximized when the sample density was in the range of ~$1 \times 10^4 - 1 \times 10^6$ CFU/mL, and samples were measured at 10-fold, 100-fold, and 1000-fold dilutions to ensure each sample had measurements in this range. Three replicate flow cytometer enumerations were performed for each sample at each dilution.

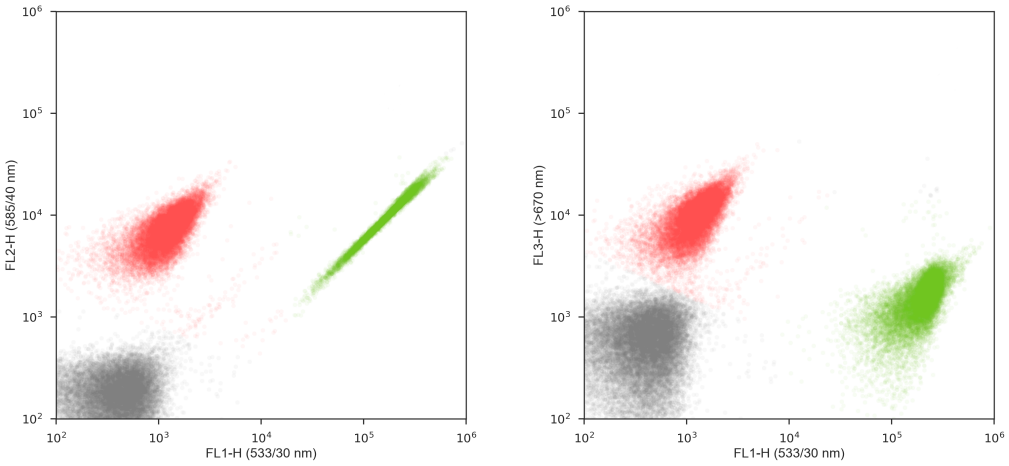


Fig. S6. Example flow cytometry data. Culture densities were estimated with cell enumeration using a BD Accuri C6 Flow Cytometer. These plots present example data collected from a single sample taken from a mixed culture of GFP-marked WT cells and RFP-marked M1 cells. The x- and y-axes refer to fluorescence intensities in the wavelength bands indicated in parentheses. Each point represents the fluorescence intensities of a particle drawn from the culture sample. Three distinct clusters of points are immediately apparent, and fixed fluorescence gates were established to automatically differentiate GFP-marked cells, RFP-marked cells, and other particles. Cells that fall within the RFP gate are colored red above, cells that fall within the GFP gate are colored green, and other particles that fall outside of both FP gate regions are colored gray. Such data provide counts of GFP-marked cells, RFP-marked cells, and total cells, which can be used to calculate the strain densities and frequencies in the culture from which the cytometry sample was taken.

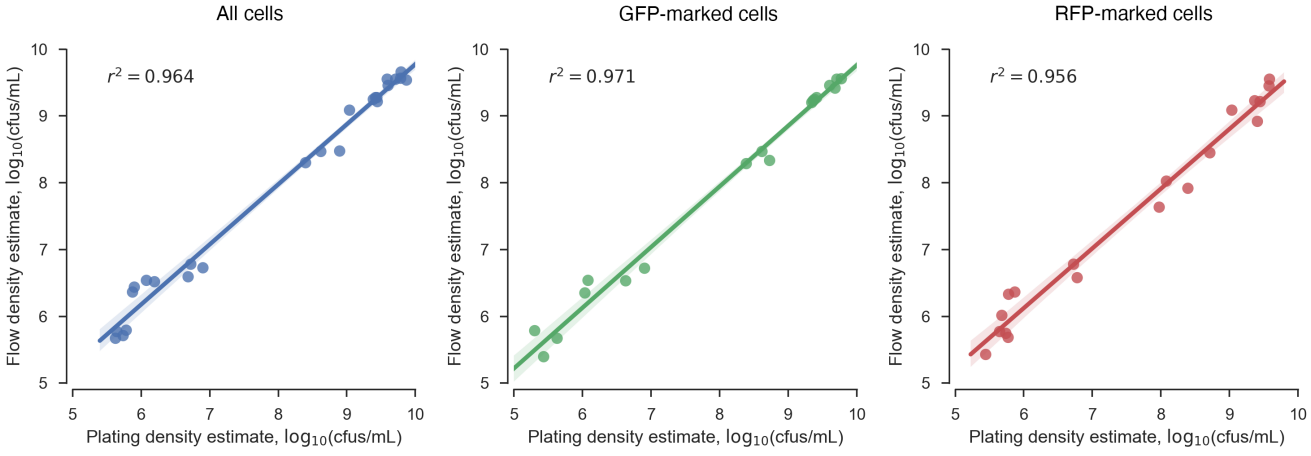


Fig. S7. Correspondence of cell enumeration using flow cytometry vs. agar plating. Mixed cultures of GFP-marked cells and RFP-marked cells were grown from low density (~1×10^6 CFU/mL) to saturation (~2×10^9 CFU/mL). Samples of these cultures were taken throughout the growth process with densities spanning several orders of magnitude. Culture densities were estimated both with cell enumeration using a BD Accuri C6 Flow Cytometer and with traditional agar plating and colony counting (colonies of GFP-marked cells appear green, and colonies of RFP-marked cells appear pink). These plots show how density estimates from flow cytometry compare to those from plating. Each point depicts the respective density estimates for the same culture sample. The leftmost plot shows the correspondence of density estimates for all cells, the middle plot shows the correspondence of density estimates for GFP-marked cells alone, and the rightmost plot shows the correspondence of density estimates for RFP-marked cells alone. In all cases, the correlation of cytometry- and plating-based estimates of culture density is high.

Uncertainty Intervals

Error bars in Figure 2 are pointwise 95% intervals representing technical uncertainty in flow cytometry enumeration. For each of 10,000 bootstrap replicates, GFP, RFP, and double-positive counts were resampled jointly within each competition, time point, and dilution; a Poisson count bootstrap was used when only one complete enumeration was available. Replicate means were calculated within dilutions and positive dilution estimates were averaged matching the original analysis. The resulting resampled time series were propagated through calculations of strain frequencies, growth rates of the population and optimal WT strain, information gain, and load. For the fixed-comparator analysis in Figure 2c,d, the overall mean WT growth rate was re-estimated in each bootstrap replicate. Published point estimates were retained, with bootstrap perturbations applied around them to preserve the original transfer-volume and timing corrections.

E.3 Supplemental results of selection experiments

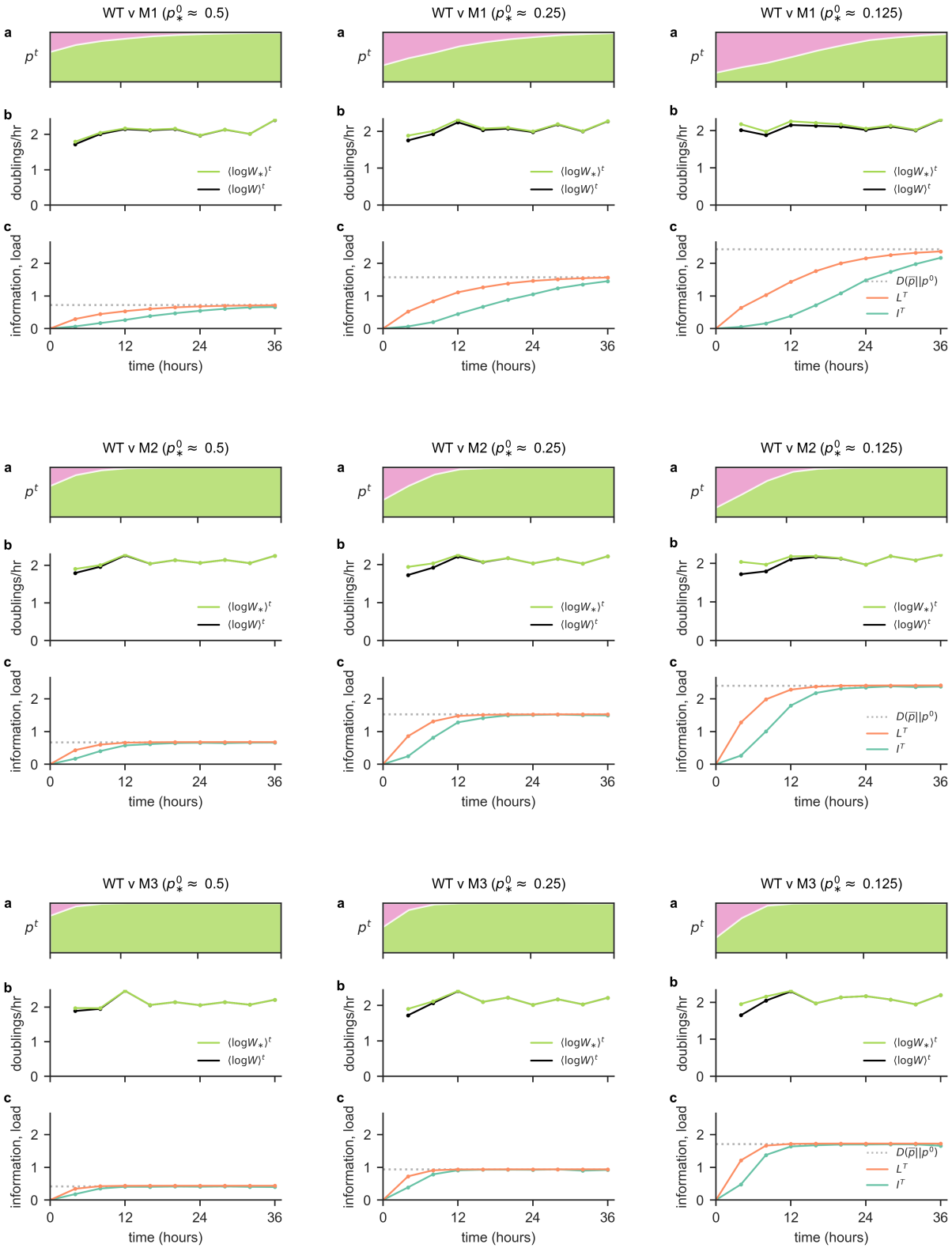


Fig. S8. Supplemental results from selection experiments. Here we show the changes in strain frequencies (**a** plots) and growth rates (**b** plots), and information and load measures (**c** plots) over time for each of the 9 selection competitions presented in main text Figure 2. The title of each sub-panel indicates the strain combination and approximate initial frequency of the optimal WT strain in the respective competition.