Evolutionary learning of adaptation to varying environments through a transgenerational feedback

BingKan Xue^{a,1} and Stanislas Leibler^{a,b}

^aThe Simons Center for Systems Biology, Institute for Advanced Study, Princeton, NJ 08540; and ^bLaboratory of Living Matter and Center for Studies in Physics and Biology, The Rockefeller University, New York, NY 10065

Contributed by Stanislas Leibler, August 22, 2016 (sent for review June 1, 2016; reviewed by Terence Hwa and Oliver J. Rando)

Organisms can adapt to a randomly varying environment by creating phenotypic diversity in their population, a phenomenon often referred to as “bet hedging.” The favorable level of phenotypic diversity depends on the statistics of environmental variations over timescales of many generations. Could organisms gather such long-term environmental information to adjust their phenotypic diversity? We show that this process can be achieved through a simple and general learning mechanism based on a transgenerational feedback: The phenotype of the parent is progressively reinforced in the distribution of phenotypes among the offspring. The molecular basis of this learning mechanism could be searched for in model organisms showing epigenetic inheritance.

evolution | bet hedging | epigenetic inheritance | population growth | environmental fluctuations

All biological organisms have to adapt to variations in their environment. This is particularly challenging if these variations are not regular or present strong random components. In this case, adaptation has to rely on the information gathered over time, allowing organisms to anticipate such environmental changes.

One class of adaptation mechanisms depends directly on detection of undergoing changes in the environment, such as a sudden shift in temperature, chemical composition, or ecological structure. Sensing and signal transduction machinery allows the organism to change the course of development or induce a direct phenotypic transformation. Such “phenotypic plasticity” can be implemented on the level of an individual organism, at the cost of maintaining an efficient detection and signal transduction system. Examples of phenotypic plasticity are ubiquitous, e.g., the switching of metabolic pathways in microorganisms upon changes of food source (1) or the control of flowering time in plants according to temperature and photoperiod modulations (2).

Another class of adaptation mechanisms relies on “phenotypic diversification,” which acts primarily on the population level. Instead of transitioning to a new phenotype upon a detected change in the environment, adaptation is achieved by constantly sustaining a distribution of various phenotypes in the population. The latter is implemented by individual organisms having different propensities for expressing different phenotypes, which can be mathematically described by an associated probability distribution. At a given time, some of the phenotypes are better adapted to the prevailing local environment than the others. The simultaneous presence of multiple phenotypes can increase the population’s long-term growth rate in a varying environment. Such adaptive phenotypic diversification is often referred to as “bet hedging” (3, 4). Many observed cases of phenotypic diversity have indeed been suggested as examples of bet hedging (5, 6), including bacterial populations that generate a low frequency of slowly growing persisters that are tolerant to antibiotics (7), annual plants that produce a fraction of seeds that remain dormant underground for multiple years while unaffected by unpredictable weather (8–10), and insects that undergo variable lengths of diapause that halts reproductive development to cope with uncertain ecological conditions (11).

Phenotypic diversification is proposed to be useful when environmental cues are unreliable or phenotypic plastic responses are

costly or ineffective (12). (Theoretical studies of population dynamics in fluctuating environments, including bet-hedging strategies, often assume growing populations with density-independent fitness values. For recent work on the effect of environmental fluctuations on growth-limited populations, applicable to phenotypic plasticity, see, e.g., ref. 13.) It has been argued that the optimal distribution of phenotypes in the population is typically determined by the frequency of encountering different environmental conditions (14). Therefore, the bet-hedging organism requires information not only about the present environment, but also about the statistics of the past environment. Such long-term environmental information can be acquired only if the information is gathered, stored, and then transmitted over consecutive generations.

If the knowledge about the past is used to adjust the phenotypic diversity of the population in a favorable direction, then the organism may be said to have “learned” to adapt to the varying environment. Fundamentally, how can such “evolutionary learning” be achieved? Ideally, the mechanism of learning should function for a wide range of environmental variations. In particular, when the environment is stable except for rare shifts, the phenotype distribution should narrow down through this learning mechanism to the most favorable phenotype during each environmental condition. On the other hand, when the environment changes frequently and irregularly, the phenotype distribution should broaden out to allow bet hedging.

We use a theoretical model to show that the evolutionary learning of adaptation to varying environments can be achieved through a simple and very general mechanism. This mechanism acts on the level of individual organisms without relying on environmental signals. It is based on a positive feedback that enhances

Significance

Phenotypic diversification, a form of evolutionary bet hedging, is observed in many biological systems, ranging from isogenic bacteria to cell populations in a tumor. Such phenotypic diversity is adaptive only if it matches the statistics of local environmental variation. Often, the timescale of environmental variation can be much longer than the lifespan of individual organisms. Then how could organisms collect long-term environmental information to modulate their phenotypic diversity? We propose here a general mechanism of “evolutionary learning,” which could overcome the mismatch between the two timescales. The learning mechanism can in principle be realized through known molecular processes of epigenetic inheritance, suggesting experimental directions to probe the evolutionary significance of such processes.

Author contributions: B.X. and S.L. designed research; B.X. performed research; and B.X. and S.L. wrote the paper.

Reviewers: T.H., University of California, San Diego; and O.J.R., University of Massachusetts Medical Center, Worcester.

The authors declare no conflict of interest.

¹To whom correspondence may be addressed. Email: bxxue@ias.edu or livingmatter@rockefeller.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1608756113/-DCSupplemental.

[illegible]

Model

For simplicity, consider a population of isogenic asexual organisms that can express alternate phenotypes. Each individual inherits information determining the probability of expressing different phenotypes and randomly expresses a phenotype according to that probability distribution. For concreteness, we imagine that the phenotypic choice is based on a bistable epigenetic switch, which takes variable molecular inputs and generates one of two possible outputs that determines a phenotype ϕ_A or ϕ_B , respectively; once a phenotype is selected after birth, it then remains unchanged due to some irreversible developmental events. The probability of selecting each phenotype, denoted by π_A and π_B satisfying $\pi_A + \pi_B = 1$, depends on the threshold of the switch and the variability of the molecular input. We assume that each generation of individuals experiences the same environmental condition, that different generations do not overlap in time, and that the environment can change between generations. We also suppose that

the phenotypes ϕ_A and ϕ_B are favorable in one of two possible environments, ε_A and ε_B , respectively, as illustrated in Fig. 14. The expected number of offspring produced by an individual with phenotype ϕ_i in an environment ε_j ($j = A, B$) is denoted by $w_i^{(j)}$. For now we assume that environmental selection is extremely strong, so that an individual cannot survive if its phenotype does not match the environment; i.e., $w_i^{(j)} = 0$ for $i \neq j$. Generalization to the case of nonextreme selection is discussed later and shown in detail in [SI Appendix](#).

The evolutionary learning mechanism, which we propose, is based on the following positive feedback: If an individual exhibits phenotype ϕ_k ($k = A$ or B), then the frequency of this phenotype is further increased in its offspring. Mathematically, the phenotype probability distribution of the offspring is different from that of the parent by an amount $\Delta\pi_i$, which is positive for $i=k$ and negative otherwise (Fig. 1*B*). Different variants of this general learning mechanism are possible; here we consider the following simple rule:

$$\Delta\pi_i = \eta(\delta_{ik} - \pi_i) = \begin{cases} \eta(1 - \pi_k), & i = k; \\ -\eta\pi_i, & i \neq k. \end{cases} \quad [1]$$

(δ_{ik} is the Kronecker delta, which equals 1 if $i=k$ and 0 otherwise.) The parameter η is the learning rate, satisfying $0 < \eta \ll 1$. (Note that the total probability is conserved because $\sum_i \Delta \pi_i = 0$.) This rule is reminiscent of “Hebbian learning” (17), which

Fig. 1. Schematic view of the evolutionary learning mechanism. (A) Each individual (solid circle) randomly expresses a phenotype ϕ_A (blue) or ϕ_B (red) according to a probability distribution π_i ($i = A, B$). In an environment e_j ($j = A, B$; background colors), a phenotype ϕ_i has a fitness value $w_i^{(j)}$. Under extreme selection, only individuals whose phenotype matches the environment survive and produce offspring; individuals with mismatched phenotypes die (crossed out). (B) The learning rule: Each individual in a new generation $t + 1$ inherits a probability $\pi_i(t + 1)$ that equals the probability $\pi_i(t)$ carried by its parent plus a change $\Delta\pi_i$ that depends on the parent phenotype. This change makes the offspring more likely ($\Delta\pi_i > 0$) to have the same phenotype as the parent and less likely ($\Delta\pi_i < 0$) otherwise. (C) Evolutionary learning in a varying environment: Bar plots show the phenotype probability distributions $\pi_i(t)$ in each generation. After a few generations, $\pi_i(t)$ approaches the environment frequencies as a result of the learning mechanism. (D) Lineages of individuals: Solid arrows form a continuous lineage; dashed arrows are where a lineage terminates. The history of phenotypes along a continuing lineage reflects the past environment.

strengthens newly experienced states in a neural network by reinforcing synaptic connections between the coactive neurons.

Through this learning mechanism, the phenotype probability distribution π_i is updated every generation and changes over time. Individuals of different phenotypes induce different changes of the distribution π_i in their offspring. Although natural selection acts not on the distribution π_i but on the phenotypes randomly generated from π_i , the overall distribution of phenotypes in the population will nevertheless adapt to the environmental changes.

Results

To better understand how the transgenerational positive feedback functions, consider first the case where environmental selection is extremely strong. Because only individuals with the phenotype that matches the environment may survive and produce offspring, those offspring will have the same $\Delta\pi_i$ and hence the same π_i . Let $\pi_i(t)$ be the phenotype distribution of individuals in the t th generation (Fig. 1C). If the environment is ε_A and remains so for t generations, then by applying the learning rule (Eq. 1) recursively, one finds (*Materials and Methods*) that the phenotype distribution becomes

$$\begin{cases} \pi_A(t) = 1 - (1 - \eta)^t (1 - \pi_A(0)), \\ \pi_B(t) = (1 - \eta)^t \pi_B(0). \end{cases} \quad [2]$$

Therefore, as a result of the transgenerational positive feedback, the probability $\pi_A(t)$ of expressing the favorable phenotype ϕ_A increases over the generations. The adaptation timescale is given by $\tau_{ad} = -1/\log(1 - \eta) \simeq 1/\eta$ (measured in the number of generations), controlled by the learning rate η .

Now, suppose that the environment switches repeatedly during that timescale τ_{ad} ; i.e., the typical duration of an environment, τ_{env} , is much shorter than τ_{ad} . Then, through the learning mechanism, the phenotype probability distribution π_i converges to a steady distribution π_i^* and fluctuates around it. Indeed, averaging $\Delta\pi_i$ from Eq. 1 over a time period $\simeq \tau_{ad}$ and assuming $\Delta\pi_i$ is small so that $\pi_i \approx \text{constant}$, one finds (*Materials and Methods*)

$$\langle \Delta\pi_i \rangle \approx \eta (p_i - \pi_i), \quad [3]$$

where p_i is the empirical frequency of the environment ε_i during that period. The steady distribution is given by $\langle \Delta\pi_i \rangle = 0$, which yields $\pi_i^* = p_i$. This distribution, being proportional to the environment frequency, exactly recovers the optimal bet-hedging strategy (14).

We have thus shown that, in a rapidly fluctuating environment, the proposed learning mechanism enables the population to reach the optimal phenotype distribution for bet hedging. Eq. 3 also implies that the phenotype distribution π_i converges to the optimal distribution exponentially, following the direction of fastest adaptation (*Materials and Methods*). In particular, for the case of extreme selection, the convergence time is given by the adaptation timescale, $\tau_{ad} \simeq 1/\eta$ (this timescale is lengthened for the case of nonextreme selection) (*SI Appendix*).

The evolutionary learning mechanism effectively allows the organism to gather information about the past environment. This process is possible because the learning mechanism gives every individual an effective memory of its ancestral phenotypes. Starting from an individual in the t th generation and following its lineage backward, let the phenotype of its n -generation ancestor be ϕ_{i-n} , where n is the generation distance. By using Eq. 1 recursively, we find (*Materials and Methods*)

$$\pi_i(t) = \sum_{n=1}^{\infty} \eta (1 - \eta)^{n-1} \delta_{i, i-n}. \quad [4]$$

This result means that the individual may express the same phenotype as its n -generation ancestor with probability

$\eta(1 - \eta)^{n-1}$, which decays exponentially with the generation distance. Hence, effectively, the individual is able to sample the phenotypes of its ancestors from a recent time period $\simeq 1/\eta \simeq \tau_{ad}$. In the case of extreme selection, the phenotypes of one's ancestors carry complete information about the past environment, because an ancestor could survive only because its phenotype matched the environment at its time (Fig. 1D). Thus, by sampling the ancestral phenotypes, the organism is able to collect the statistics of the past environment. In particular, if the environmental variation happens much faster than the adaptation timescale τ_{ad} , the organism will be able to adapt by learning a bet-hedging strategy.

Note that in the opposite regime, in which an environment lasts for a time τ_{env} much longer than τ_{ad} , the organism is able to adapt by learning the most favorable phenotype in that environment. In the previous example where the environment remains ε_A for t generations, one finds $\pi_A(t) \rightarrow 1$ when $t \gg \tau_{ad}$ in Eq. 2, and hence the phenotype distribution becomes biased heavily toward the phenotype ϕ_A . Thus, interestingly, without having any sensors, the population effectively shows a form of plasticity that lets individuals preferentially express the favorable phenotype in the current environment. As a result, the population quickly approaches the maximum growth rate in that environment (*Materials and Methods*).

Our model can be generalized to include nonextreme environmental selection, in which case individuals with a mismatched phenotype can also produce offspring. According to the learning mechanism, those offspring will have a phenotype distribution π_i that is different from that of the offspring of the individuals with the favorable phenotype. As a result, π_i no longer remains homogeneous in the whole population. Nevertheless, if the fitness of the mismatched phenotype is much lower than that of the favorable phenotype, then the variation in π_i will be small among individuals, and the average phenotype distribution $\bar{\pi}_i$ over the whole population should behave similarly to the extreme selection case (*SI Appendix*). Note that, although we have considered here the case of only two different phenotypes, a generalization of the model leads to similar results for any number of alternate phenotypes (*Materials and Methods* and *SI Appendix*).

Those analytical results are confirmed by numerical simulations. Examples of such simulations for nonextreme selection are shown in Fig. 2. As predicted, the evolutionary learning mechanism allows the population to dynamically adjust the phenotype distribution, depending on the pattern of environmental variations. When the environment changes frequently, i.e., $\tau_{env} \ll \tau_{ad}$, the phenotype distribution quickly converges to the optimal bet-hedging solution (Fig. 2A and D). As a result, the population reaches a stable phenotypic diversity that offers the highest possible growth rate without any direct sensing of the environment. On the other hand, when the environment remains stable for long periods, i.e., $\tau_{env} \gg \tau_{ad}$, the population progressively expresses the favorable phenotype in every environment and approaches the maximum growth rate after a finite adaptation period $\simeq \tau_{ad}$ (Fig. 2C and F). Hence, effectively, the population shows a “transgenerational phenotypic plasticity.”

In addition, if the timescale of environmental changes is comparable to the adaptation timescale τ_{ad} , the population may exhibit intermittent periods of either bet-hedging or transgenerationally plastic behavior (Fig. 2B and E). This behavior happens when the actual durations of each environment vary broadly with time. As a result, there are times when an environment happens to last for a period $\tau > \tau_{ad}$ (e.g., $t \approx 50 - 100$ in Fig. 2B and E), as well as times when environmental durations are predominantly short, $\tau < \tau_{ad}$ (e.g., $t \approx 0 - 50$ in Fig. 2B and E). In the former case the population shows transgenerational plasticity, whereas in the latter case it adopts a bet-hedging strategy.

The adaptation timescale τ_{ad} , which determines the behavior of the population growth for a given pattern of environmental variations, is controlled by the learning rate η . Different learning rates may be achieved, depending on the molecular basis of the learning

Fig. 2. Simulation of a learning population in a fluctuating environment. The environment alternates between two conditions ε_A and ε_B . Each individual carries a probability distribution (π_A, π_B) and randomly selects a phenotype ϕ_A or ϕ_B . Each phenotype ϕ_i has a fitness $w_i^{(j)}$ in the environment ε_j , where $w_i^{(j)} = [[2.0, 0.2], [0.2, 2.0]]$. The learning rate is $\eta = 0.1$, which corresponds to $\tau_{ad} \approx 10$. (A–C) Blue shades represent the histograms of the probability $\pi_A(t)$ in each generation t estimated from $N = 10^5$ individuals, with normally distributed initial values $\pi_A(0) \sim \mathcal{N}(0.5, 0.05)$; yellow lines are the average probability $\bar{\pi}_A(t)$ over the population. (D–F) Blue lines are the population growth curves estimated from the $\pi_A(t)$ s in the same examples as in A–C; black lines represent the optimal bet-hedging solutions calculated from the environment frequencies; yellow lines are the maximum growth curve obtained by having the phenotype matching the environment at all times and with no cost for sensing. (A and D) The environment is independent and identically distributed ($\tau_{env} \approx 1$) with $p_A = 0.7$, $p_B = 0.3$. (B and E) The durations of the environment ε_A and ε_B are geometrically distributed with means $\tau_A = 10$ and $\tau_B = 5$, respectively. (C and F) The environment switches every $\tau_{env} = 40$ generations.

mechanism. Therefore, in principle, the learning rate could be selected over evolutionary timescales. A quantitative comparison of the performance of different learning rates is presented in *SI Appendix*, which yields the optimal learning rate η^* that maximizes population growth for given patterns of environmental variations. For very short environmental durations (e.g., $\tau_{env} \ll 10$), it is found that η^* approaches 0, and thus the phenotype distribution becomes nearly constant after reaching the optimal bet-hedging solution. In contrast, if long periods of constant environment exist (e.g., $\tau_{env} \gtrsim 10$), η^* becomes positive and finite, and hence the population can take advantage of the transgenerational plasticity during such periods. The behaviors of the phenotype distribution and the population growth for the optimal learning rate under those environmental variation patterns are similar to the examples shown in Fig. 2 (*SI Appendix*).

Note that the process of evolutionary learning cannot be thought of as adaptation through competition between groups of individuals with different and stably inherited phenotype probability distribution π_i s. In general, each lineage of individuals could have a different π_i , and modification of π_i arises constantly at every generation and in every individual. Hence the population acts as a whole, with a dynamically maintained distribution of π_i s. It is the population, not particular individuals, that adapts to the varying environment through the evolutionary learning of the optimal phenotypic diversity.

Discussion

The learning mechanism that we propose here can potentially be realized by means of epigenetic inheritance. The essential feature of the learning mechanism is the progressive reinforcement of the parent phenotype in the distribution of offspring phenotypes. Such parent-dependent changes of phenotype distribution have been characterized, for example, in the Agouti viable yellow (A^{vy}) mouse (18), a well-studied model organism for the so-called

transgenerational epigenetic inheritance. The expression of the A^{vy} allele results in a yellow coat color, as well as some other physiological conditions. However, isogenic mice carrying this allele (A^{vy}/a) show variable coat colors ranging from full yellow to wild-type agouti. Importantly, the distribution of phenotypes among the offspring is influenced by the phenotype of the dam (but not the sire)—yellow dams produce offspring that are more likely to be yellow, whereas agouti dams produce a higher percentage of agouti offspring; moreover, the percentage of agouti offspring increases even further in the next generation if both the dam and the grand-dam are agouti (18). Similar transgenerational effects are observed in axin-fused ($Axin^{Fu}$) mice showing a distribution of kinky tail phenotypes, which in this case depends on the phenotype of both parents (19). The molecular basis of these transgenerational effects has been attributed to the epigenetic inheritance of the DNA methylation pattern at the intracisternal A particle (IAP) site upstream of the A^{vy} ($Axin^{Fu}$) locus (18, 19), although the details of how the methylation pattern escapes epigenetic reprogramming are not fully understood (20).

Many other examples of such transgenerational effects are found in organisms across a wide range of taxa (21). There are well-established molecular processes that enable transgenerational epigenetic inheritance (15, 22), such as DNA methylation, histone modification, small RNA interference, and protein structural templating. Any of those epigenetic processes can in principle serve as the basis of the proposed learning mechanism. In addition, if the environmental statistics remains stable, the adaptation achieved through the learning mechanism may be further stabilized by other molecular processes acting on even longer timescales, including genetic adaptations such as contingency loci, copy number changes, and mutations (15, 16).

In conclusion, we have introduced a general evolutionary learning mechanism for adaptation to varying environments. It is based on a transgenerational feedback that can use well-established

molecular processes of epigenetic inheritance. It is our hope that this theoretical work will stimulate experimental searches for such a learning mechanism in organisms exhibiting adaptive phenotypic diversification.

Materials and Methods

Here we derive the dynamic changes of the phenotype probability distribution π_i under the learning rule (Eq. 1). In general, suppose there are multiple possible phenotypes (can be more than two) labeled by ϕ_i ($i=A, B, \dots$), and each phenotype is expressed with probability π_i , satisfying $\sum_i \pi_i = 1$. We treat the simple case with extreme selection, i.e., when the fitness matrix $w_i^{(j)}$ has only diagonal elements. Generalization to the case of nonextreme selection, i.e., when $w_i^{(j)} \neq 0$ for $i \neq j$, is presented in [SI Appendix](#).

Long-Lasting Environment and Plastic Phenotype Distribution. Under extreme selection, only individuals with a phenotype that matches the environment could survive and produce offspring. When the environment is e_j and remains so, the phenotype probability distribution π_i is updated every generation by

$$\pi_i(t+1) = (1-\eta)\pi_i(t) + \eta\delta_{ij}. \quad [5]$$

The explicit expression of $\pi_i(t)$ for all i can be found by recursively using Eq. 5 as follows:

$$\begin{aligned} \pi_i(t) &= (1-\eta)\pi_i(t-1) + \eta\delta_{ij} \\ &= (1-\eta)^2\pi_i(t-2) + (\eta + (1-\eta)\eta)\delta_{ij} \\ &\vdots \\ &= (1-\eta)^t\pi_i(0) + \sum_{n=0}^{t-1} (1-\eta)^n \eta\delta_{ij} \\ &= (1-\eta)^t\pi_i(0) + \left(1 - (1-\eta)^t\right)\delta_{ij}. \end{aligned} \quad [6]$$

Therefore, the probability for the favorable phenotype ϕ_j is given by the component

$$\pi_j(t) = 1 - (1-\eta)^t (1 - \pi_j(0)), \quad [7]$$

which increases with time and approaches 1 when $t \gg \tau_{ad} \approx 1/\eta$.

As the environment remains to be e_j , the number of individuals in the population after t generations is given by

$$N(t) = \left(\prod_{n=0}^{t-1} \pi_j(n) w_j^{(j)} \right) N(0), \quad [8]$$

where $N(0)$ is the initial population size. The growth curve of the logarithmic population size is given by

$$\log \frac{N(t)}{N(0)} = \sum_{n=0}^{t-1} \log(\pi_j(n) w_j^{(j)}), \quad [9]$$

with an instantaneous growth rate $\Lambda^{(j)}(t) = \log(\pi_j(t) w_j^{(j)})$. By the time $t \geq \tau_{ad}$, the population will approach the maximum growth rate in this environment, $\Lambda^{(j)}(t) \rightarrow \Lambda_{max}^{(j)} = \log w_j^{(j)}$.

Frequently Changing Environment and Optimal Bet-Hedging Solution. When the environment changes frequently and irregularly, the phenotype probability distribution π_i converges to a steady distribution and fluctuates around it. Let ϕ_{k_t} be the viable phenotype in the t th generation. By averaging $\Delta\pi_i$ from Eq. 1 over the adaptation timescale τ_{ad} and assuming $\Delta\pi_i$ is small so that $\pi_i \approx$ constant, we find

$$\begin{aligned} \langle \Delta\pi_i \rangle &\approx \eta (\langle \delta_{i,k_t} \rangle - \pi_i) = \eta \left(\frac{1}{\tau_{ad}} \sum_t \delta_{i,k_t} - \pi_i \right) \\ &= \eta \left(\frac{\tau_i}{\tau_{ad}} - \pi_i \right) = \eta (p_i - \pi_i). \end{aligned} \quad [10]$$

For the second equality we used the condition of extreme selection, which means k_t must match the environment at all times, and therefore δ_{i,k_t} simply counts the number of times when the environment is e_i during that period τ_{ad} , denoted by τ_i . The ratio of τ_i and τ_{ad} then gives the empirical frequency p_i of the environment e_i . Finally, setting $\langle \Delta\pi_i \rangle = 0$ yields the steady distribution, $\pi_i^* = p_i$.

To see that this steady distribution is in fact the optimal bet-hedging solution, note that the asymptotic growth rate of a population using a constant phenotype distribution $\pi = (\pi_A, \pi_B, \dots)$ is given by (14)

$$\Lambda(\pi) = \sum_j p_j \log \sum_i \pi_i w_i^{(j)} = \sum_j p_j \log(\pi_j w_j^{(j)}), \quad [11]$$

where the second equality holds under extreme selection, $w_i^{(j)} = w_i^{(j)} \delta_{ij}$. This Λ is a strictly concave function of π , which guarantees a unique maximum. The maximum can be found by using gradient ascent,

$$\Delta\pi_i \propto \sum_j g_{ij} \left(\frac{\partial \Lambda}{\partial \pi_j} \right) = \sum_j (\pi_j \delta_{ij} - \pi_i \pi_j) \left(\frac{p_j}{\pi_j} \right) = p_i - \pi_i, \quad [12]$$

where $g_{ij} = \pi_i \delta_{ij} - \pi_i \pi_j$ is the Fisher information metric on the space of probability distributions, $\{\pi | \sum_i \pi_i = 1\}$. Therefore, setting the gradient to 0, we find that the optimal bet-hedging solution is indeed given by $\pi_i^* = p_i$. Comparing Eqs. 3 and 12 shows that the learning mechanism, averaged over time, follows the direction of steepest ascent.

For consistency, we have to check that the fluctuation around the steady distribution is small. During the period of an environment that lasts for a typical time $\tau_{env} \ll \tau_{ad}$, the relative change in the probability π_i can be estimated as

$$\left| \frac{\Delta\pi_i}{\pi_i} \right| \simeq 1 - (1-\eta)^{\tau_{env}} \approx \eta \tau_{env} \approx \frac{\tau_{env}}{\tau_{ad}} \ll 1. \quad [13]$$

Because the environmental changes are random, $\Delta\pi_i$ from different time periods tend to average out. This supports our approximation that $\pi_i \approx$ constant over a period τ_{ad} .

Learning as Transgenerational Phenotypic Memory. Here we derive how an individual's phenotype probability $\pi_i(t)$ depends on the phenotypes of its ancestors. Start from an ancient ancestor with a phenotype probability $\pi_i(0)$ and follow the lineage onward with one ancestor in every generation t , whose actual phenotype is labeled by ϕ_{i_t} . We find, recursively,

$$\pi_i(1) = (1-\eta)\pi_i(0) + \eta\delta_{i,i_0}, \quad [14]$$

$$\begin{aligned} \pi_i(2) &= (1-\eta)\pi_i(1) + \eta\delta_{i,i_1} \\ &= (1-\eta)^2\pi_i(0) + (1-\eta)\eta\delta_{i,i_0} + \eta\delta_{i,i_1}, \end{aligned} \quad [15]$$

$$\begin{aligned} \vdots \\ \pi_i(t) &= (1-\eta)^t\pi_i(0) + (1-\eta)^{t-1}\eta\delta_{i,i_0} + \dots \\ &\quad + (1-\eta)\eta\delta_{i,i_{t-2}} + \eta\delta_{i,i_{t-1}}. \end{aligned} \quad [16]$$

For large t , the initial probability $\pi_i(0)$ becomes irrelevant, and hence we can write the above expression for $\pi_i(t)$ as in Eq. 4.

Simulation of the Learning Mechanism. The adaptation of the phenotype distribution through the learning mechanism is simulated as follows: A population of size N is created. Each individual carries a probability distribution $\pi = (\pi_A, \pi_B, \dots)$. At each time step, an environment e_{j_t} is chosen. Each individual randomly generates a phenotype according to its own π . The fitness of each individual is then determined by its phenotype and the current environment. At the end of each generation, a new sample of N individuals is drawn independently from all current individuals with a probability proportional to their fitness values. This step represents a process of replicating each current individual by a number proportional to its fitness value and then normalizing the sample size back to N . Each new individual inherits an updated probability distribution $\pi + \Delta\pi$ from its parent, according to the learning rule (Eq. 1). The process is repeated for every new generation.

Although the population size is constantly normalized, the instantaneous growth rate of the population can be estimated by

$$\Lambda(t) = \log \sum_i w_i^{(j_t)} \bar{\pi}_i(t), \quad [17]$$

where the average $\bar{\pi}_i$ is taken over all individuals in the current generation. Then, the logarithmic growth of the population size is estimated by

$$\log \frac{N_t}{N_0} = \sum_{s=1}^t \Lambda(s). \quad [18]$$

Those estimates are valid in the limit of a large sample size N .

(to S.L.) through Rockefeller University (Grant 345430) and the Institute for Advanced Study (Grant 345801). B.X. is funded by the Eric and Wendy Schmidt Membership in Biology at the Institute for Advanced Study.

- PNAS | October 4, 2016 | vol. 113 | no. 40 | 11271