

Supplementary Material

Appendix S1: Mean First-Passage Times of a Multidimensional Random Walk in an Arbitrary Environment

In contrast to a one-dimensional random walk process, a multidimensional walk, propagating from a predefined starting point to another predefined end point, can take any of numerous pathways. Clearly, the characteristics of each pathway, and in particular the walk probabilities assigned to each location, determine both the dynamics of the walk along this pathway and the likelihood of choosing it. Consider all the possible pathways from this starting point to the end point on a given landscape. Each of these pathways can be conceived as a simple one-dimensional “landscape”, which in turn induces a specific drawdown value. We argue that in arbitrary environments, these drawdown values will significantly differ from each other, exhibiting a roughly exponential distribution. The essential uniqueness of drawdown trajectories is similar in nature to that obtained for upper bounds by Deuschel and Zeitouni (1999) in the problem they handle, bearing some rough similarity to ours. This argument stems directly from the strong exponential effect of the pathway drift (manifested by the values of the odd-ratios ρ_i along the pathway) on the drawdown value. We will term the pathway with the minimal drawdown value ‘the *Principal-Pathway*’ and the drawdown value it induces ‘the *Principal-Pathway drawdown*’. Considering the correlation between the first-passage time and the drawdown value in one-dimensional landscapes, we thus conjecture that all pathways apart from the *Principal-Pathway* are in fact irrelevant. Essentially, the random-walk process will take place along a “sausage”-like region around that minimal drawdown pathway and can be regarded as “almost” a one-dimensional process. The expected first-passage time will thus be dominated by the *Principal-Pathway* drawdown in an analogous manner to that shown in the one-dimensional case.

To generate rugged multidimensional fitness landscapes with a tunable drawdown value we

use the following modified versions of the generalized Rastrigin and Schwefel functions, widely used multimodal benchmark functions (Mühlenbein et al., 1991; Salomon, 1996; Ballester and Carter, 2004):

$$F_{Rastrigin}(\vec{x}) = -C_r \cdot d - \sum_{i=1}^d x_i^2 + C_r \cdot \cos(2\pi x_i) \quad C_r \in [1, 4]$$

$$x_i \in \{-5.0, -4.8, \dots, 0\}$$

$$F_{Schwefel}(\vec{x}) = -C_s \cdot \sum_{i=1}^d (-x_i \sin(\sqrt{|x_i|})) \quad C_s \in [0.001, 0.015]$$

$$x_i \in \{-500, -460, \dots, 420\}$$

where d denotes the dimension (see also Fig. 2a). These modified versions are designed so that the global optimum is located at one corner of the hypercube on which the function is defined and the starting point can be positioned on the opposite corner, at a relatively low-fitness value. C_r and C_s provide a simple way to control the ruggedness of the landscape (and consequently, the drawdown value it induces).

We validate our conjecture regarding the drawdown values distribution across various pathways numerically, using the modified Rastrigin function defined above. A set of two-dimensional and three-dimensional landscapes with varying drawdown values was generated. For each such landscape, a sample of 1000 pathways from the starting point $(-5,-5)$ to the global optimum point $(0,0)$ was randomly selected with uniform probability and the drawdown value of each pathway was calculated. For simplicity, we limit our sampling to forward oriented pathways (i.e. pathways composed of exactly $d(n-1)$ steps, wherein d denotes the dimension, n denotes x_i resolution in each dimension and each step is a positive step in one of the dimensions). Although actual pathways may be more complex and can include also negative steps, this subset is sufficient to demonstrate the drawdown values distribution. Furthermore, considering only forward oriented pathways, we can calculate the Principal-Pathway drawdown using dynamic programming whereby for each point in the multidimensional space, the Principal-Pathway

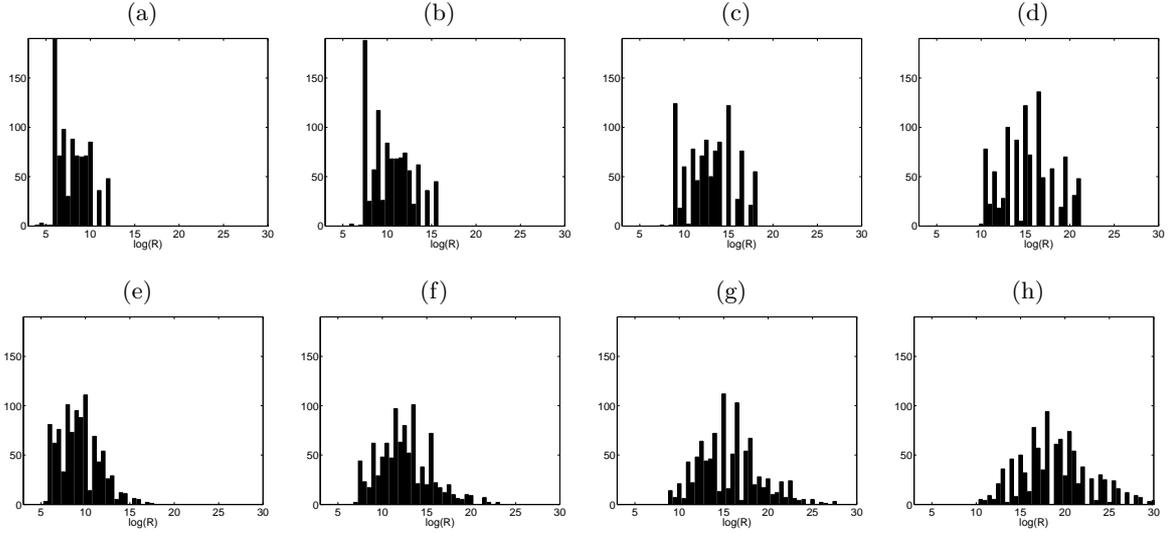


Fig. S1: The distribution of drawdown values across randomly selected pathways on fitness landscapes with varying ruggedness. **(a-d)** Drawdown values distribution on two-dimensional landscapes with increasing ruggedness. The landscapes are generated with the modified Rastrigin function described above with C_r coefficient of 2.4, 3.0, 3.5 and 4 respectively. The logarithm of the Principal-Pathway drawdown for these landscape is 3.702, 4.787, 5.692 and 6.596 respectively. **(e-h)** Drawdown values distribution on three-dimensional landscapes with increasing ruggedness. The C_r values and Principal-Pathway drawdowns are similar to those described in the two-dimensional case.

drawdown to that point can be obtained according to the values calculated for points that may preceded it in the pathway. Figure S1 illustrates the resulting distribution of drawdown values for several landscapes. Evidently the variation in drawdown values induced by randomly selected pathways is extremely large (note that the distribution is drawn on a logarithmic scale), and increases markedly with the ruggedness of the landscape. Moreover, as predicted above, the Principal-Pathway drawdown is markedly smaller than the drawdown of a randomly selected pathway.

Appendix S2: Methods

S2.1: The Evolutionary Process

A one-dimensional innate fitness function was defined on the interval $[1, 200]$ as a sum of several Gaussian functions, yielding a continuous, multi-peaked function $F(x)$ (Fig. 5a, solid line). Various plasticity schemes were then applied (see below) to produce the corresponding effective fitness functions.

The evolutionary process was simulated as a simple random walk. The RW probabilities in each location i are calculated as: $p_i = \mathcal{B}_T(F_i^+) = 1/[1 + e^{(F_i^- - F_i^+)/T}]$, $q_i = \mathcal{B}_T(F_i^-) = 1/[1 + e^{(F_i^+ - F_i^-)/T}]$ where $F_i^+ = F(i + 1)$, $F_i^- = F(i - 1)$ (using the effective fitness function in the plastic mode), and $\mathcal{B}_T(x)$ denotes the Boltzmann scaling (Mitchell, 1996) with fixed temperature $T = 0.1$. The genetic configuration x in the first generation of each evolutionary trial was set to 1.

To evaluate the convergence rate of an evolutionary process numerically, two measures are considered. First, the first-passage time of each genetic configuration provides a direct measure of the progress rate of the evolutionary process. Second, the population mean *innate* fitness value in each generation provides a good estimate for the genetic quality of the evolving individuals and allows us to track the extent of the Baldwin effect, i.e., how well did the genetically encoded solution approach the optimal one (Mayley, 1997). Clearly, evolving individuals in the plastic mode are ultimately evaluated according to their effective fitness value (that is, their fitness after phenotypic plasticity takes place), which may be higher than their innate fitness value. However, here we focus on the rate of the evolutionary process, tracking down the genetic assimilation.

S2.2: Ideal and Partial Deterministic Learning

Ideal deterministic learning is implemented as a simple iterative hill climbing process. Hill climbing iterations are performed until reaching the local maxima and no further improvement is possible. Partial deterministic learning was examined using a learning scheme which applies a

limited number of hill climbing iterations. The resulting effective fitness functions are illustrated in Figure 5a.

S2.3: Stochastic Learning

Each individual employs 100 SA (simulated annealing) iterations during its lifespan to determine its effective fitness value. Let S_i denote the individual's current configuration. In each iteration, one of the adjacent configurations (± 1) is selected at random as a candidate for a new configuration S_j . If $F(S_j) > F(S_i)$ the new configuration becomes the current configuration. Otherwise, S_j becomes the current configuration with probability $e^{(F(S_j)-F(S_i))/T}$ where T denotes a temperature parameter that cools from 1 to 0 throughout the learning process.

S2.4: Random Phenotypic Variation

Given an individual with genetic configuration x , we define $F_{ec}(x) = F(x_v)$ where x_v is randomly chosen with a uniform distribution from the interval $[x - \Delta d, x + \Delta d]$. Δd thus represents the range of phenotypic variation allowed by some developmental process. This random perturbation is performed anew in each generation throughout the simulation run.

S2.5: Random Walk with Static Periods

The random walk process described in Section S2.1 is extended to allow staying in the same genetic configuration, i , with a probability z_i . The RW probabilities in each location i are calculated as: $p_i = \mathcal{B}_T(F_i^+) = \frac{e^{F_i^+/T}}{C}$, $q_i = \mathcal{B}_T(F_i^-) = \frac{e^{F_i^-/T}}{C}$, $z_i = \mathcal{B}_T(F_i) = \frac{e^{F_i/T}}{C}$ where $F_i = F(i)$ and C is a normalization factor (letting $p_i + q_i + z_i = 1$).

S2.6: Random Walk with Kimura's Fixation Probabilities

For each genetic configuration i , we calculate the selection coefficient s for each of the two mutants ($i + 1$ and $i - 1$), $s_i^\pm = F_i^\pm / F_i - 1$. We then calculate the fixation probability, u , of each of the two mutants as follows: If $|s|$ is small (i.e., $|s| \leq 1/N_e$), we use Kimura's fixation

probability for neutral mutations, $u = 1/(2N_e)$. Otherwise $u = (1 - e^{-4N_e sm})/(1 - e^{-4N_e s})$, where N_e denotes the effective population size and $m = 1/2N_e$ denotes the initial frequency of the mutant. In the simulations described in this study we use $N_e = 100$. We set the RW probabilities, $p_i = u^+$, $q_i = u^-$ and $z_i = 1 - p_i - q_i$.

References

- Ballester, P. and Carter, J. 2004. An effective real-parameter genetic algorithm for multimodal optimization. In: *Adaptative Computing in Design and Manufacture VI* (I. Parmee, ed.), pp. 359–364. Springer, London.
- Deuschel, J. and Zeitouni, O. 1999. On increasing subsequences of i.i.d. samples. *Comb. Probab. Comput.*, 8(3):247–263.
- Mayley, G. 1997. Guiding or hiding: Explorations into the effects of learning on the rate of evolution. In: *Proceedings of the Fourth European Conference on Artificial Life* (P. Husbands and I. Harvey, eds), pp. 135–144. Bradford Books/MIT Press, Cambridge, UK.
- Mitchell, M. 1998. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, UK.
- Mühlenbein, H., Schomish, M., and Born, J. 1991. The parallel genetic algorithm as function optimizer. *Parallel Computing*, 17:619–632.
- Salomon, R. 1996. Re-evaluating genetic algorithm performance under coordinate rotation of benchmark functions: A survey of some theoretical and practical aspects of genetic algorithms. *BioSystems*, 39:263–278.