Supply-driven evolution: mutation bias and trait-fitness distributions can drive macro-evolutionary dynamics

Zhun Ping (Julian) Xue[1*], Leonid Chindelevich[2], Frédéric Guichard[3],

1 Faculty of Medicine, McGill University, Montreal, Quebec, Canada
2 MRC Centre for Global Infectious Disease Analysis, Imperial College, London, United Kingdom
3 Department of Biology, McGill University, Montreal, Quebec, Canada

* zhunping.xue@mcgill.ca

Abstract

Many well-documented macro-evolutionary phenomena still challenge current evolutionary theory. Examples include long term evolutionary trends, major transitions in evolution, conservation of certain biological features such as *hox* genes, and the episodic creation of new taxa. Here we present a framework that may explain these phenomena. We do so by introducing a probabilistic relationship between trait value and reproductive fitness. This integration allows mutation bias to become a robust driver of long term evolutionary trends against environmental bias, in a way that is consistent with all current evolutionary theory. In cases where mutation bias is strong, such as when detrimental mutations are more common than beneficial mutations, a regime called "supply-driven" evolution can arise. This regime can explain the irreversible persistence of higher structural hierarchies, such as happens in the major transitions in evolution. We further generalize this result to the long-term dynamics of phenotype spaces. We show how mutations that open new phenotype spaces can become frozen in time. At the same time, the new possibilities may be observed as a burst in the creation of new taxa.

Introduction

One of the great and unresolved debates in evolutionary theory is the relationship between macro and micro-evolution. On the one hand, some authors provide excellent arguments for why macro-evolution should be an extension of micro-evolution [1], on the other hand, other authors have articulated why macro-evolution is a separate process that cannot be distilled to micro-evolution [2]. Among the best arguments of the second school of thought are the many macro-evolutionary phenomena that resist any micro-evolutionary explanation. These phenomena fall into four classes: the first class describes long-term evolutionary trends [3]. The trends are all empirically controversial [4], but their theoretical possibility remains a persistent question in macro-evolution. The second class are the major evolutionary transitions [5] such as the increase in structural hierarchy from RNA molecules to prokaryotes and eukaryotes, from single-cell eukaryotes to multi-cellularity, and from solitary individuals to eusociality. The seeming irreversibility of major transitions in evolution relates them to a third class of phenomena, sometimes described as "generative entrenchment" [6, 7], where developmental events are profoundly conserved over time. One example would be

the *hox* genes, which are partially responsible for the body plan of lineages. Some aspects of these body plans seems to "lock-in" over time: for example, there is no member of the tetrapod class with more than four limbs. Some developmental events in evolution simply matter more than other events, giving evolution the character of historical contingency [8] - a single such event in developmental pathways can alter the entire evolutionary trajectory of the lineage in which it occurs. A fourth class of macro-evolutionary phenomena is proposed by the proponents of punctuated equilibrium, who argue that micro-evolution provides no satisfactory explanation for the repeated bursts of novelty followed by relative quiescence [9].

These four classes of phenomena point to the inadequacies of existing micro-evolutionary theory. However, despite rich and fascinating development, modern macro-evolutionary theory still offers no satisfactory alternative [10]. Here we introduce the theory of "supply-driven" evolution (SDE), a mechanism based on mutation bias, which can explain all four classes of macro-evolutionary phenomena described above while remaining compatible with existing theories of micro-evolution. SDE integrates the probabilistic nature of the phenotype-fitness relationship with mutation bias to predict the dynamic evolution of phenotypic spaces and the locking-in of phenotypic structures.

From mutation bias to supply-driven macro-evolution

Mutation bias has historically been rejected as a significant force in evolution, starting with Fisher's early models that pitted mutation bias against natural selection. These models showed that when two alleles are in competition with each other, even a small amount of natural selection in favor of one allele easily overwhelms even large amounts of mutation bias in the opposite direction, especially when mutations are rare [11]. These predictions were based on the assumption that both alleles were already available. However, when only some variation is immediately available to selection, mutation bias can be a fundamental driver of evolution, because whichever phenotype arrives first gets first chance at success [12–17]. When combined with rugged fitness landscapes, mutation bias can have a profound effect on evolutionary trajectories because a population will discover different fitness hills depending on the frequency of different mutations.

At the same time, developmental biology unequivocally demonstrated that there are constraints on the availability of mutations. Many mutations are developmentally impossible. This means that genetic variation is biased [18–22]. Theoretical frameworks inspired by quantitative genetics have formalized verbal arguments about mutation bias and experimental observations from developmental biology [23–26]. These approaches require measuring and making assumptions about the shape of multivariate phenotype landscapes, such as the trait variance-covariance (G) matrix, as well as the fitness landscape over many traits. The complexity of these models make intuitive insight difficult.

An alternative approach consists in considering the probabilistic nature of the relationship between trait values and fitness: two organisms with the same trait value do not necessarily have the same fitness [27]. This is because changes in a trait value such as body length can be achieved via different mechanisms, such as increased cell size or cell proliferation [28]. As another example, there are many ways for different organisms to have the same complexity level, no matter how complexity is measured. There is no reason to expect two organisms, having achieved the same body length or complexity level via different means, to have the same fitness. This insight implies that there is a many-to-many mapping between trait and fitness values (Fig 1A). Because a given trait value is realizable from different phenotypes, that trait value can be associated with a range of different fitnesses.

This multiple realizability of trait values leads to a probabilistic relationship between

trait and fitness where each trait value is associated with a distribution of fitness values (Fig 1B). This contrasts with the more common assumption of a one-to-one relationship between trait and fitness, making mutation bias a weak driver of evolutionary change in the standard model of evolution. In the many-to-many case, however, mutation bias is given much more power: if a particular trait value is created more frequently through mutation bias (Fig 1C), then the possibility of a high fitness variant with that trait value also becomes more likely, because we are sampling more frequently from a distribution of fitness values.

In this paper, we begin by presenting a formal model of supply-driven evolution. We then show how the combination of mutation bias and the trait-fitness distribution can create long term trends in evolution that are not reversible by the environment. Using our framework and simulations of the NK-model, we next show how this result implies a special class of evolutionary innovations: those that create new phenotype spaces. We show how supply-driven evolution can cause these innovations to become locked-in over macro-evolutionary time. Our results suggest that the emergence of new phenotype spaces, their locking-in, and their subsequent exploration by micro-evolutionary processes can explain major evolutionary transitions, generative-entrenchment, and macro-evolutionary burst-quiescence dynamics.

Supply-driven evolution: a mathematical framework

Let us consider a homogeneous haploid population with fitness ω^* and a trait with value ϕ^*. Let a mutant m be defined by (s_m, u_m) where $s_m = \omega_m - \omega*$, the difference between the mutant and resident fitnesses, and $u_m = \phi_m - \phi^*$, the difference between their trait values. A mutant-fitness distribution, $\zeta(s, u)$, can now be defined over all mutants of the population, which is a bivariate probability distribution over s and u. When a mutant enters the population, its fitness and trait value are determined by choosing from ζ. Ordinary evolution then takes place to determine the success of this mutant through demographic drift and natural selection. Note how, in this framework, mutants with the same trait value ϕ_m can have very different fitnesses, drawn from the distribution $\zeta(s, u = u_m)$ (see Figure 1).

We will work within a classic adaptationist scenario, where there is strong selection and weak mutation (for a categorization of the different regimes of evolution, see [29]). In this case, mutations are rare. When a beneficial one comes along, it can sweep through the population if it escapes drift. In the ideal case, the only driver of evolution in this scenario are selective sweeps, which happen on a timescale that is short compared with the timescale of the generation of beneficial mutants. To simplify our analytical model, we will assume that detrimental mutations are evolutionarily inconsequential, that there is no clonal interference (that is, there is only one adaptive mutation at a time – see [30]), and that improvements in fitness are small and incremental. We will see with agent-based models that relaxing these assumptions does not change the final results.

Since improvements in fitness are small, the probability of a mutant with fitness advantage s sweeping the population is αs, where α is a constant that depends on the details of evolution ($\alpha = 2$ for Poisson distributed offspring [31], $\alpha = 1$ for a Moran process [32], $\alpha \approx 2.8$ for binary fission – see the discussion in [33]). Thus, the probability distribution that a mutant will sweep the population and change the trait value by u is a function of u:

$$p(u) \approx \alpha \int_0^\infty s\zeta(s, u)ds \tag{1}$$

We integrate over s from 0 to ∞ because we only care about the higher fitness

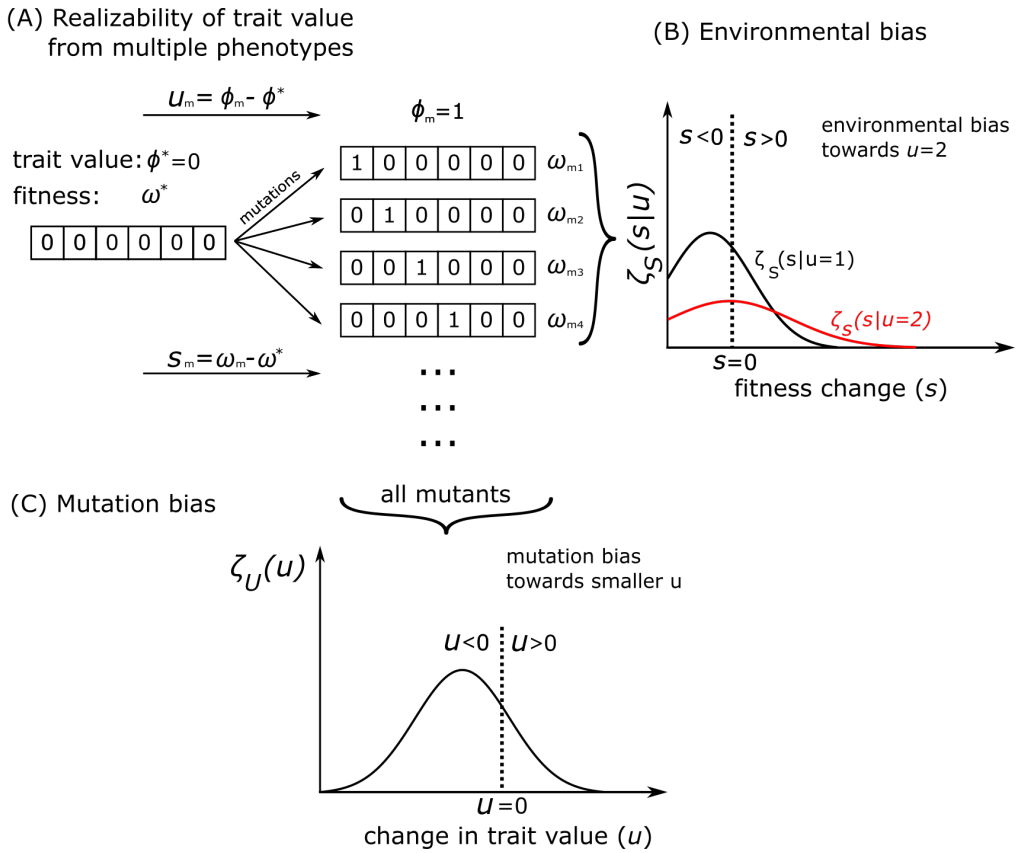


Fig 1. Components of supply-driven evolution. (A) Illustration using the NK-model of the multiple realizability of trait values from mutations. On the top left, we have the wild type, whose components are all 0's. It has fitness ω^* and trait value $\phi^* = 0$. To its right are some of its possible mutants. These mutants all have the trait value $\phi_m = 1$ with $u = 1$, but they are different phenotypes and have different fitnesses, which form (B) a fitness distribution $\zeta_S(s|u = 1)$ over s (the difference between mutant and wildtype fitnesses). The fitness distribution $\zeta_S(s|u = 2)$ for the set of all mutants $\phi_m = 2$ (red line) illustrates a case where the environment favors $u = 2$. (C) The distribution of mutations $\zeta_U(u)$ across all fitness values shows a mutation bias towards smaller u.

mutants. The probability $p(u)$ does not normalize to 1 over all u, the probability that there is no selective sweep makes up the rest of the weight.

Of course, $\zeta(s, u)$ may change over time, but for now we will consider a single selective sweep before taking a long term view. A single selective sweep, under this regime, is rapid.

To simplify this equation and to understand it better, we first can separate out the contribution of mutation bias to evolution. We can split ζ up according to its conditional probability distribution, using the standard notation where U and S are random variables for u and s:

$$\zeta(u, s) = \zeta_S(s|U = u)\zeta_U(u) \tag{2}$$

Here, $\zeta_U(u)$ is the marginal density of mutants that change trait value by u, while $\zeta_S(s|U = u)$ is the fitness distribution of all mutants with trait value difference u. In essence, $\zeta_U(u)$ is the mutation distribution: it tells us how likely a mutant with particular u is likely to arrive. On the other hand, the environment is captured by $\zeta_S(s|U = u)$: for a given trait value, the environment determines the fitness distribution of all mutants with that trait value. For example, if the environment becomes more favorable for a trait value u, the mean of the distribution $\zeta_S(s|U = u)$ may increase. Of course, the environment can change $\zeta_S(s|U = u)$ in many other ways, such as increase its variance or skew, which can also qualitatively alter evolution.

Next, we split $\zeta_S(s|U = u)$ into two parts, the detrimental and the beneficial parts

$$\zeta_S(s|U = u) = a_u\zeta_S(s|U = u, s < 0) + b_u\zeta_S(s|U = u, s > 0)$$

where

$$a_u = \int_{-\infty}^{0} \zeta_S(s|U = u)ds$$

and

$$b_u = \int_{0}^{\infty} \zeta_S(s|U = u)ds$$

That is, a_u is the probability that a mutation that changed the trait value by u is detrimental, and b_u is the probability that a mutation that changed the trait value by u is beneficial. In this case, equation (1) becomes

$$\begin{aligned} p(u) &= \alpha \int_{0}^{\infty} s\zeta_U(u)b_u\zeta_S(s|U = u, s > 0)ds \\ &= \alpha\zeta_U(u)b_u\left[\int_{0}^{\infty} s\zeta_S(s|U = u, s > 0)ds\right] \\ &= \alpha\zeta_U(u)b_u\mathbb{E}[\zeta_S(s|U = u, s > 0)] \\ &= \zeta_U(u)\beta(u) \end{aligned} \tag{3}$$

Here, $\beta(u) = \alpha b_u\mathbb{E}[\zeta_S(s|U = u, s > 0)]$ can be called the "degree of benefit" of u, which is the probability that a mutation that changed the trait value by u is beneficial, b_u, multiplied by the size of the expected fitness advantage, $\mathbb{E}[\zeta_S(s|U = u, s > 0)]$, multiplied by α. What β measures is the probability that a mutation with trait value u will selectively sweep – that is, the mutation will both be beneficial as well as escape drift. Detrimental mutations drops out of the equation because the integration on s is only from 0 to ∞, that is, in this framework only beneficial mutations will alter the evolutionary trajectory.

Equation (3) defines the probability that the population trait value will change by u through the introduction of the next mutant as the likelihood of this mutant appearing,

ζ_U, multiplied by the degree of benefit of this change, and a factor α that depends on the exact nature of the evolutionary process as described earlier. In this equation, the mutation distribution plays an equal and symmetric role to environmental selection. A trait value with half the benefit of another trait value can make up for it with double the mutation bias.

We can integrate through all trait values, u, to arrive at the expected change in trait value through the introduction of one mutant:

$$\mathbb{E}[u] = \int_{-\infty}^{\infty} u\zeta_U(u)\beta(u)du \tag{4}$$

To recap, the above equation holds in regimes where the only evolutionary events are selective sweeps by beneficial mutations, where beneficial mutations are rare and the fitness effect of beneficial mutations are small. In this formulation, selection affects evolution by changing $\beta(u)$, while mutation bias is captured by $\zeta_U(u)$. The last equation holds for the introduction of a single mutant. As time goes on, both ζ_U and β may change: the former as the structure of the organism changes, thus changing the sort of mutants it produces, and the latter will change as the environment changes.

Prediction of trait change in an agent-based model

We can first apply equation (1) to predict how a trait evolves in agent-based simulations where $\zeta_U(u)$ and $\beta(u)$ are known or can be measured. To demonstrate this, we adopt an agent-based model that was previously used to show that mutation bias can direct the evolution of a trait [34], even against environment selection [27].

The model is an agent-based NK-model [35] where agents are strings made from elements of 0's and 1's. There are interactions among these elements that create a rugged fitness landscape. Mutations flip 0's to 1's and vice-versa. The trait we measure is the sum of all the elements. The environmental bias is defined in such a way that the larger this sum, the more likely that the agent is fit. There is a mutation bias in the opposite direction, however, in the sense that mutations from 1 to 0 are more common than mutations from 0 to 1 (Fig 1A, see Appendix A for details).

In this model, ζ_U is known because it corresponds to how often mutations change 0 to 1 or vice versa. $\beta(u)$ needs to be measured, because it is a dynamic quality determined by how close the population is to a fitness local optimum. $\beta(u)$ is affected by many factors: the height of the fitness optimum, the rate of change in the environment, and the mutational load, among other things.

Since $\beta(u)$ measures the degree of benefit provided by a trait value, it will change with the environment – as a population gets close to the optimum of an environment, $\beta(u) \approx 0$ for all u, since there are much fewer mutants with improved fitness. If the environment changes constantly at a regular pace, however, we can expect $\beta(u)$ to remain relatively constant over time. This is what happens in this model: with a single measurement of $\beta(u)$, we can make good predictions for the entire time period (Figure 2). In the first plot, there is mutation bias pushing the trait value down, while in the second plot, there is no mutation bias, so there is only an environmental bias pushing the trait value up. In both plots, $\beta(u)$ was measured at $t = 2e4$, giving the model time to find an equilibrium between environmental change and mutation-selection. After that, $\beta(u)$ stayed relatively stable over time, so a single measurement allowed us to make good (although imperfect) predictions over the entire simulation both backward and forward through time. Measuring $\beta(u)$ at $t = 0$ was less useful, since the population has not settled in an equilibrium with the environment In these cases, then, the actual long term rates and directions of a trait's evolution can be known if we can obtain a measurement for $\beta(u)$.

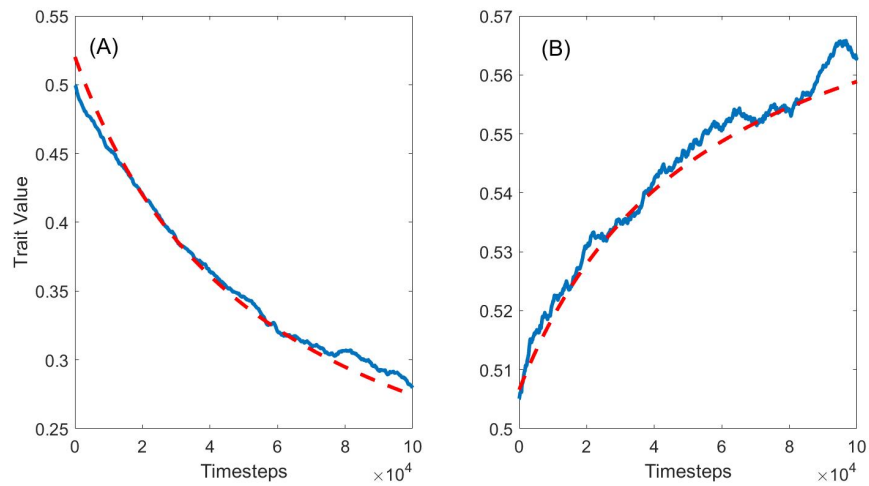


Fig 2. The blue line is the average trait evolution for 200 model runs, the red dashed line is the predicted evolution by equation (4). For the set of runs in panel A, mutation bias pressures the trait value down while the environment is biased to larger trait values. The environment changes over time. Panel B is identical, but without mutation bias, allowing environmental bias to push the trait value upwards.

The supply-driven evolution of structural hierarchies

Supply-driven evolution

We will now study scenarios where $\zeta_U(u)$ introduces a mutation bias that is unlikely to be reversed by selection from the environment. This means that the behavior of equation (4) is determined by $\zeta_U(u)$, a regime that we call "supply-driven" evolution. This happens when $\zeta_U(u)$ satisfies conditions that would guarantee $\mathbb{E}[u] > 0$ (or, conversely, $\mathbb{E}[u] < 0$) for almost any $\beta(u)$. One possible such condition is set below:

$$\begin{aligned}\mathbb{E}[u] &= \int_{-\infty}^{\infty} u\zeta_U(u)\beta(u)du \\ &= \int_{-\infty}^{0} u\zeta_U(u)\beta(u)du + \int_{0}^{\infty} u\zeta_U(u,t)\beta(u,t)du \\ &\geq \beta_{max}^{-}\int_{-\infty}^{0} u\zeta_U(u)du + \beta_{min}^{+}\int_{0}^{\infty} u\zeta_U(u)du\end{aligned}$$

In the above equation, β_{max}^{-} is the maximum value of $\beta(u)$ for all $u < 0$ and β_{min}^{+} is the minimum value of $\beta(u)$ for all $u > 0$.

Thus, $\mathbb{E}[u] \geq 0$ if (but not only if):

$$-\frac{\int_0^{\infty} u\zeta_U(u)du}{\int_{-\infty}^{0} u\zeta_U(u)du} \geq \frac{\beta_{max}^{-}}{\beta_{min}^{+}} \tag{5}$$

This is actually a very stringent condition. In particular, this condition is hard to satisfy if β_{min}^{+} is zero or very small. We can sharpen this condition by trimming out a range of u as follows:

Let $\mathbf{v}$ be the set of positive real numbers such that $\beta(v) < \epsilon, \forall v \in \mathbf{v}$. Let us define the set $\backslash\mathbf{v}$ as the set of all positive real numbers with $\mathbf{v}$ removed. Then we can guarantee $\mathbb{E}[u] \geq 0$ if

$$-\frac{\int_{\setminus \mathbf{v}} u\zeta_U(u)du}{\int_{-\infty}^{0} u\zeta_U(u)du} \geq \frac{\beta_{max}^{-}}{\beta_{min}^{\setminus \mathbf{v}}} \tag{6}$$

In the above equation, $\beta_{min}^{\setminus \mathbf{v}}$ is the minimum value of $\beta(u)$ over the set $\mathbf{v}$. This guarantees that $\beta_{min}^{\setminus \mathbf{v}} \geq \epsilon$, but reduces the value of the numerator on the left side. This is a useful condition, for example, if $\beta(u)$ drops off to zero for all u larger than some threshold, u_c. This can happen when the environment poses a hard limit to what values u can take, such that all organisms with $u > u_c$ die, rendering β_{min} for this range to be zero. In that case, in the above equation, $\setminus \mathbf{v}$ reduces to $[0, u_c]$. Mutation-bias can still drive evolution up to near u_c, so long as $\int_{\setminus \mathbf{v}} u\zeta_U(u)du$ is large enough.

Supply-driven evolution happens if the above conditions are satisfied over a long period of time. This means that the portion of ζ spread over $u > 0$ strongly outweighs the portion of ζ spread over $u < 0$, to the point where it outweighs any likely behavior of β, and this fact remains stable over time. If this is true, then mutation-bias dominates the direction of the evolution of u, regardless of the environment.

Biologically, it is easy to see what β_{max} means. This is the degree of benefit of the optimal trait value over $u < 0$. Empirically, there are few mutations that would, say, immediately double the ancestor's fitness, no matter what the new trait value is, so β_{max} is likely to be quite small, certainly less than 2. On the other hand, β_{min} is a biologically important parameter that we have never measured empirically. If β_{min} is too small, no mutation bias can drive evolutionary trends. We expect β_{min} to decrease with increasing strength of selection on the trait u compared to selection on the phenotypic realization of that trait. If the environment is directly selecting for body size, for example, with no or weak selection on how body size is phenotypically realized, then β_{min} will be small or zero for anything other than trait values that drive the lineage closer to optimal body size, and the environment alone is the driver of evolution. If, however, what really matters is the phenotype beneath the trait value, as might be the case for complexity, then β could be relatively independent of u, such that $\beta_{min} \approx \beta_{max}$. In this case, even small amounts of mutation bias can give rise to supply-driven evolution.

The origin of increase in structural hierarchies

The evolution of structural hierarchies capture major transitions in evolution [5] including transition from the RNA world [36] to the first cells, from prokaroytes to eukaryotes [37], from unicellular to multicellular eukaryotes [38], and from multicellular organisms to eusociality, such as ants or naked mole-rats, and perhaps humans [39]. What is most interesting about these major transitions is their irreversibility. As far as we know, there is no free-living unicellular organism that had an animal, plant, or fungus as its ancestor. Nearly all animals get cancer, but no cancer cell is known to survive independently of the host other than in highly controlled lab conditions. At the very limit, there are cancers in dogs, clams, and Tasmanian Devils that can infectiously spread, but even they are not known to live autonomously [40, 41]. There are a few exceptions to this irreversibility, since there are solitary organisms that had eusocial ancestors [42] and there are eukaryotes that have lost their mitochondria [43, 44], although they usually retain a remnant and are still recognizably eukaryotes [45–48]. We should also note that these exceptions are allowed by SDE and their probability are predicted to decrease with time.

Although there have been explanations for the origin of particular increases in hierarchy, we note that there is no consensus on a fitness based theory for why the increase in hierarchy should persist. Prokaryotes are, by almost any measure, the most

successful kingdom on the planet. The main other type of explanation is drift to fixation of deleterious alleles in each of the components organisms [49], rendering the hierarchy obligatory. In the next section, we show how this will occur with high probability even when there is no neutral or deleterious drift to fixation and only selection takes place.

The mutation bias that drives the locking-in of hierarchies

When two component organisms live in such close synchrony that they form one unit of selection – that is, they share a measure of fitness, and mutations in each affect this joint fitness, then mutations on either component have two effects: one on the fitness of the joint ensemble, and another on the fitness of individual components, should they separate from the whole. To see how the separation of the component organisms becomes impossible, we simply have to show a scenario of how the fitness of each component would steadily degrade over time. Components may begin as autonomous organisms, but become fully dependent on the whole over time and unlikely to have a viable fitness when severed from the whole.

We first define the fitness of the component organism as a trait with value ϕ. We assume mutation bias leads to more frequent detrimental mutations compared to beneficial mutations [50–52]. We then define ω as the fitness of the joint ensemble. We now have $u = \Delta\phi$ being the change in fitness of the components, should they become separated, while $s = \Delta\omega$ is the change in fitness of the joint ensemble, which is what selection is immediately acting on.

Since the trait variation u is also a change in the fitness of individual components, we can use Fisher's framework (see, for example, [53]) considering organisms to be points in a high dimensional space, where each dimension represents a trait. Mutations occur within a mutational volume around the organism, and the environmental optimum is another point in this space. Using this framework, theory and experiments both show that detrimental mutations always outnumber beneficial ones. This relation becoming stronger as the organism approaches the environment optimum [50].

Assuming this inequality between detrimental and beneficial mutations and its relationship with the environmental optimum, we have:

$$\int_{-1}^{0} \zeta(S = s, u)du \gg \int_{0}^{\infty} \zeta(S = s, u)du$$

as we approach the environmental optimum. Since u is a change in fitness, its range is $[-1, \infty]$.

However, this implies that

$$\int_{-\infty}^{\infty}\int_{-1}^{0} \zeta(s, u)duds \gg \int_{-\infty}^{\infty}\int_{0}^{\infty} \zeta(s, u)duds$$

Switching the order of integration,

$$\int_{-1}^{0} \zeta_U(u)du \gg \int_{0}^{\infty} \zeta_U(u)du \tag{7}$$

From this, our aim to show that $\mathbb{E}[p(u)] < 0$, that is, we expect the fitness of the component organisms to degrade at each selective sweep, such that over time they can no longer live autonomously.

In order to achieve this, we must define the relationship between trait and fitness variations. It has first been shown that the magnitude of detrimental mutations among the component organisms are, on average, larger than the magnitude of beneficial mutations [51]. That is, the average detrimental mutation is more detrimental than the

average beneficial mutation is beneficial, even if detrimental mutations have a lower bound of -1 while beneficial mutations have no upper bound. This relationship means that:

$$-\mathbb{E}[\zeta_U(u|u<0)] > \mathbb{E}[\zeta_U(u|u>0)]$$

Since

$$\mathbb{E}[\zeta_U(u|u<0)] = \frac{\int_{-1}^{0} u\zeta_U(u)du}{\int_{-1}^{0} \zeta_U(u)du} \tag{8}$$

And similarly for $\mathbb{E}[\zeta_U(u|u>0)]$, multiplying the equations (7) and (8) gives us

$$\int_{-1}^{0} u\zeta_U(u)du \gg \int_{0}^{\infty} u\zeta_U(u)du \tag{9}$$

Since detrimental mutations are always more common than beneficial ones, the above equation is true over time. Thus, we are within striking distance of being in the regime of supply-driven evolution, as we described in the previous section in equation (5). What matters are the values of β, especially β^-_{min}. Biologically, β^-_{min} is asking a deep question: is it likely that mutations that decrease the fitness of a component organism, still are beneficial for the joint organism? If the answer is never, then $\beta^-_{min} = 0$ and supply-driven evolution is impossible. However, if it is yes, then β^-_{min} will be a reasonable value, and we may be in the regime of supply-driven evolution, such that

$$\frac{\int_{-1}^{0} u\zeta_U(u)du}{\int_{0}^{\infty} u\zeta_U(u)du} \geq \frac{\beta^+_{max}}{\beta^-_{min}} \tag{10}$$

In this case, mutation bias will steadily drive u down to zero regardless of what the environment does. This means that, over evolutionary time, the component organism will no longer be able to separate from the joint organism, as its independent fitness become negligible.

Evolutionary lock-in: an agent-based model

We apply the above theory to a NK model, so agents are strings made from elements of 1's and 0's that interact with each other. However, mutations do not change the elements; instead, they change interactions. Each mutation causes a number of elements to interact with different elements than before. We also have a rarer class of mutations: occasionally an agent will be joined with another agent, to form a much longer string. We call this a joining mutation. Initially, the elements of the two agents have only a few links with each other, but with mutations they eventually can be more heavily linked to each other. Lastly, long agents can also be cleaved in half to form two individual agents. The interactions that were present between the elements of the two agents when they were a single organism will then be lost.

The environment is constructed such that the longer the agent, the less likely it is to be fit. This is a conservative assumption, to show that joining mutations can be locked-in even when the environment is biased against it. There is a 2% penalty to fitness each time the length of the organism doubles. Increasing this penalty increases the amount of time before evolutionary lock-in holds; the system simply has to wait for a joining mutation that increases the fitness for the resulting joined organism, despite its increased length. This is possible because there are different ways for agents to join together, some of which cause an increase in fitness that can overcome the 2% penalty. For details on the model, see Appendix B.

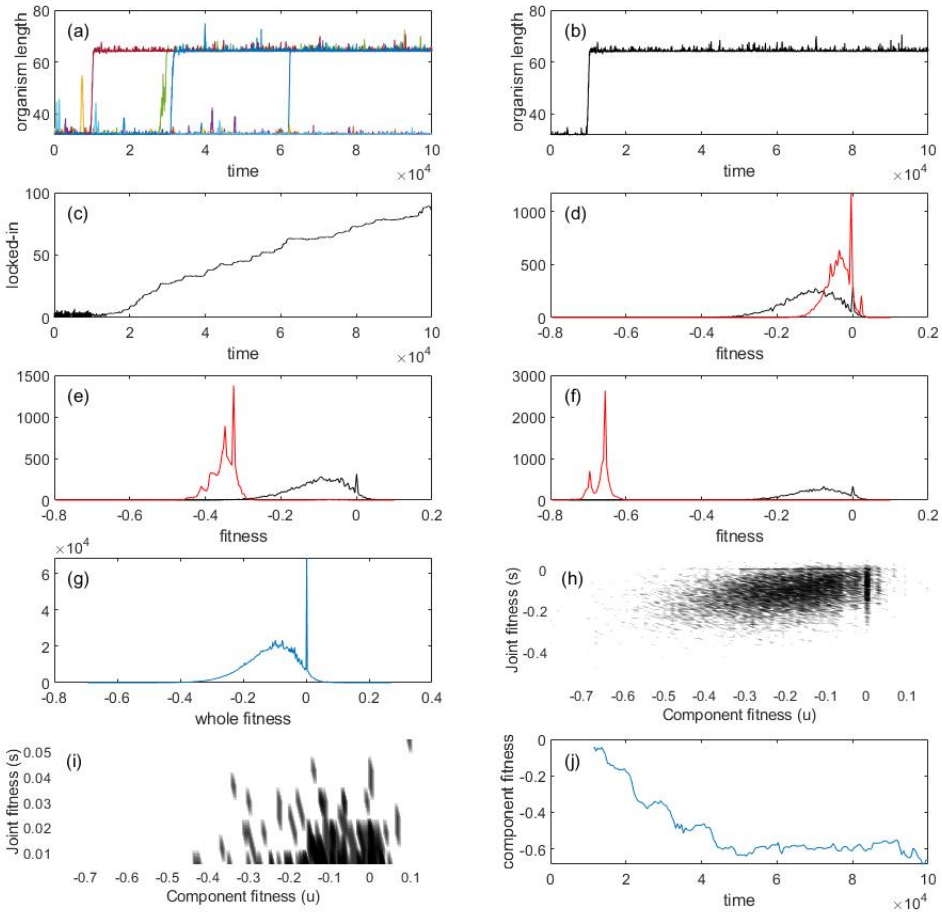


Fig 3. Results from NK simulations. (a) Time series of organism length for eight simulations. (b-j) Results from a single simulation where locking-in of a joining mutation happens early on. See text for more details.

Time series of organism length over multiple simulations (Figure 3a) illustrate how many joining mutations almost immediately reverse themselves, but some persist. This persistence means that the number of hierarchies rises over time, despite an explicit fitness penalty for doing so. Focusing on a single simulation where locking-in is observed (Figure 3b), we can show that the increased hierarchy becomes irreversible when two components become increasingly enmeshed with each other as measured by the number of links between the first and second halves of the joint organism (Figure 3c). This is a measure of how much effect a cleaving mutation will have on the organism. This steadily increases with time, which means that a cleaving mutation will have increasing effect on phenotype over time, reducing its chances of being beneficial. On the other hand, u, the fitness of the component organisms, steadily declines over time (Figure 3d-f). We can see how the component organisms are becoming less and less fit and how much more frequent detrimental mutations are, compared to beneficial mutations (Figure 3g), providing a direct comparison with equation (7). We can see a spike in the number of mutations that lead to no change in fitness of the whole organism (Figure 3g), which are all the neutral mutations (0.36%), but there are many more detrimental mutations (92.7%) than beneficial mutations (6.94%).

In this experiment, we can also directly measure $\zeta(s, u)$ right after the joining event and directly see that there is a large class of mutations that benefit the larger agent, while being detrimental to the component agents (Figure 3h,i). This shows the mutation bias that drives the enmeshment: most mutations that are beneficial to the whole organism are detrimental to the component agent. As these mutations become fixed, the fitness of the component agents, u, steadily declines over time (Figure 3j). These results are compatible with our earlier prediction that due to the strength of the mutation bias, which may be over an order of magnitude in one direction versus the other, no naturally observed level of selection is likely to reverse this decline in fitness of the component organisms.

Generalized evolutionary lock-in

We now generalize this mechanism to a theory of evolutionary lock-in beyond the specific case of joining mutations. In the above example, joining mutations become irreversible (that is, lock-in) because they create a phenotypic space that was not previously possible - the joint organisms. If this new space is large enough, then supply-driven evolution will push the trajectory of the lineage into the new space, since there will be many more mutations that link the two organisms together than mutations that separate them. Once the lineage has remained in this new phenotypic space (the joint organism) for a while, return to the old space becomes impossible, because component organisms have had their fitnesses degraded beyond return.

We now generalize this example by defining σ as a "potentiation" structure, after [54]. This is a structure that creates new phenotypic possibilities. To be specific, σ makes N new mutations possible, and if σ is lost, then these N mutations become impossible again. We assume this mutation space is unstructured, meaning that the N mutations are independent from each other. σ becomes locked in when mutations that cause σ to be lost become almost certainly detrimental. This means a very long time must pass before σ is lost. If this time is long compared to the duration of the lineage, then the lineage is more likely to become extinct than to lose σ.

We now show that σ's lifespan scales exponentially in N (to the order of $\Omega(a^N)$, where a is a constant larger than 1), so that even for moderately large values of N, σ can lock in (that is, remain conserved for such a long period of time that the lineage is most likely to end before σ is lost). Immediately after σ is gained, we track the number of new mutations n, made possible by σ, which become real. Clearly, n will vary from

initially 0 to a maximum of N. At each selective sweep, the organism can either gain or lose such a mutation, so n can either increase or decrease. We do not track the selective sweeps that neither produce a gain nor a loss of a mutation in this space, because such selective sweeps do not change the ultimate outcome in this system. In addition, there is a certain number of mutations, B, that cause σ to be lost. If σ is lost, all n of the newly realized mutations are also lost. In the previous NK model, there is only one such mutation (so, $B = 1$), the cleaving mutation that split the joint organism back into two, and all the interactions between the two are lost.

Our task to to show that, for these B mutations that cause σ to be lost, the probability that they sweep the population becomes vanishing small over time. Let us say that, if $n = 0$ (that is, no new mutations were lost), then the class of all mutations that cause the loss of σ has benefit β, where benefit is defined as in equation (3): the probability that this class of mutations have of sweeping (both being advantageous and escaping drift). In the NK model, this would be the likelihood that a cleaving mutation sweeps the population if there were no links between the first and second halves of the joint organism.

We re-emphasize that, when σ is lost, all the mutations that were gained in the new mutation space are lost. The more mutations that are lost, the less likely that the loss of σ would be beneficial. If n new mutations were lost, we assume that the probability that the reversal mutation is advantageous is μ^n for some $0 < \mu < 1$. We justify this as follows: let's say that for the n mutations lost, each lost mutation has a probability ν of being advantageous. Conservatively, let's say that more than $n/2$ of the mutations lost must be advantageous for the mutation that reverses σ to be, overall, advantageous. (The choice of $n/2$ is arbitrary, any fraction would work – we would just make a variable substitution in n.) This corresponds to the cumulative binomial density function, $\Pr(B(n, 1-\nu) \geq \lfloor n/2 \rfloor)$. Since ν is the probability of an advantageous mutation, we can reasonably assume that $1 - \nu > 1/2$, which means that Hoeffding's inequality holds:

$$\Pr(B(n, 1-\nu) \leq \lfloor n/2 \rfloor) \leq \exp\left(-2n\left(\nu - \frac{1}{2}\right)^2\right)$$

The probability that a mutation that reverses σ is beneficial is at most μ^n, where

$$\mu = \exp\left(-2\left(\nu - \frac{1}{2}\right)^2\right)$$

Thus, the probability that a mutant which lost σ can sweep the population is simply $\beta\mu^n$.

At time t, let there be $n(t)$ new mutations that have been actualized. At this point, there are n new mutations that can be lost and $N - n$ new mutations that can be gained. If all mutations have the same probability, this corresponds to the strength of the mutation bias that will increase or decrease $n(t)$. Initially, there are N mutations that will increase $n(t)$ and 0 that will decrease it, leading to a very strong mutation bias. At $n = N/2$, there will be equal numbers of mutations that increase or decrease $n(t)$, so there will be no more mutation bias that favors the increase of $n(t)$.

There can, of course, be an environmental side to this, such that increases or decreases in $n(t)$ can be favored or disfavored by the environment. However, this will be unimportant in our case for even moderately large values of N, since at the beginning of the process of exploring the new mutation space, n will be very small and so $N \gg n$, a difference that will be very hard for the environment to overcome. By the intuition we have developed in this paper, we can see that although the environment can alter the specific values of equilibrium $n(t)$, it is unlikely to alter the ultimate outcome of whether σ can be locked-in.

In this case, let us assume that the fitness benefit of increasing n is p, and of decreasing n is q. The benefit of a mutation that reverses σ is β. With these notations, we can now model the system with the Markov chain transition matrix, $\mathbf{S}$, of size $N+2$:

$$\mathbf{S} = \begin{bmatrix} 0 & \frac{Np}{D_0} & ... & ... & ... & \frac{B\beta}{D_0} \\ \frac{q}{D_1} & 0 & \frac{(N-1)p}{D_1} & ... & ... & \frac{B\beta\mu}{D_1} \\ 0 & \frac{2q}{D_2} & 0 & \frac{(N-2)p}{D_2} & ... & \frac{B\beta\mu^2}{D_2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & ... & \frac{(N-1)q}{D_{N-1}} & 0 & \frac{p}{D_{N-1}} & \frac{B\beta\mu^{N-1}}{D_{N-1}} \\ 0 & ... & ... & \frac{Nq}{D_N} & 0 & \frac{B\beta\mu^N}{D_N} \\ 0 & ... & ... & ... & 0 & 1 \end{bmatrix} \tag{11}$$

Here $D_i = (N-i)p + iq + B\beta\mu^i$.

We can show that the expected time to losing σ, or T_B, for $p = q = \beta$ and $B = 1$, is (see Appendix C)

$$\mathbb{E}[T_B] = \Omega\left(\frac{2}{1+\mu}\right)^N \tag{12}$$

This does not qualitatively change if we alter the values of p, q or β, for any constant B, once a potentiation structure comes about, the expected time before it is lost increases exponentially with its amount of phenotypic novelty.

We should also note that this process does not only involve the opening of new phenotypic spaces by the introduction of σ, it also closes off old phenotypic spaces, because all the mutations that cause σ to be lost (there are B of them) are now rendered evolutionarily impossible. There is thus an opening, but also a closing of evolutionary space over time. This mechanism works generally: all evolutionary innovations that open sufficiently large new phenotypic space can become locked-in. σ can represent the formation of eukaryotes, multicellularism or eusociality, but it can also be associated with new ways of patterning that are not necessarily hierarchical, such as the formation of body axes, or of limbs.

Discussion

Evolutionary biologists have long had an intuition that macro-evolution is a different kind of process from micro-evolution [55–59]. This intuition has been criticized from the view that macro-evolution is "nothing but" successive rounds of micro-evolution. Balanced against the latter view, some see evolutionary biology to be an essentially historical science [60–62], contingent upon historically significant events such as major evolutionary transitions [5, 63]. These problems still resist standard explanations of modern theory. The mechanism presented here may answer some of these problems. We explain these macro-evolutionary phenomena in a way that is compatible with micro-evolutionary processes with an evolutionary regime called "supply-driven evolution" (SDE). In SDE, micro-evolution drives macro-evolution via natural selection, but mutation bias changes the phenotypes that are present for natural selection. We show that this process can dynamically open or close phenotypic spaces. Our analysis and simulation results show how SDE provides an appropriate framework to study long

term evolutionary trends, some of the major transitions in evolution, the profound conservation (locking-in) of certain structures, as well as the episodic creation of new taxa.

Our framework describes macro-evolutionary patterns where micro-evolutionary changes come about primarily by the selective sweep of rare beneficial mutations. The next, and most important, requirement is that a given trait value must correspond to a phenotypic space that includes many different phenotypes, each with their own fitness, and collectively resulting in distribution of fitness for any given trait value. For example, two very different organisms can have the same height or weight. Secondly, we assume a distribution of mutation probabilities allowing for a bias in the order of introduction of mutants. If a particular trait value is introduced more frequently, then one of its associated phenotypes is more likely to succeed.

Applying selective sweeps to these fitness and mutation distributions, we show how there are scenarios where mutation bias can be so strong that no environment is likely to reverse it. We called these regimes "supply-driven" evolution (SDE). The most obvious use for SDE is to show how mutation bias can create long term evolutionary trends. However, it can also show how the major evolutionary transition resulting from increases in structural hierarchies become irreversible, from self-replicating RNA molecules to prokaryotes, to eukaryotes, to multicellularity, to eusociality. Mutations that leads to such hierarchies are those that join organisms into a new joint organism. We showed that these mutations create new phenotypic space as the new joint organism. This space defines traits with their respective fitness distributions. One such trait is the likelihood of breaking the joint organism into its components, which requires components to be more fit than the joint organism. This trait is part of a phenotypic space created along with the joint organism and is realizable in many different phenotypes. We showed that a strong mutation bias drives this likelihood down over time because most of the mutations that benefit the joint organism will be detrimental to the component organisms.

We then showed that this mechanism generalizes to any structure that makes possible new phenotypes. If a new structure came about that makes available a new set of possible phenotypes that were not previously evolutionarily accessible, then we showed that this structure can be evolutionarily locked-in over time such that no descendant of the lineage of this organism will break it. The measure we consider is the likelihood that this structure will break in a way that is selectively favorable; but there is a strong mutation bias driving this likelihood lower over time. The bias is because there are, at least initially, many more ways to enter into the new set of possibilities than to leave it, and once enough new possibilities are realized, breaking the structure that made it all possible is almost certainly detrimental. The new structure thus opens new space, but closes off old ones.

The theory of SDE expands macro-evolutionary theories beyond the Modern Synthesis, while not rejecting any of its elements. The fact that eukaryotism has persisted might have nothing to do with whether it is evolutionarily advantageous for the lineage (except for that short moment in which it came about); but rather because eukaryotism was impossible before and possible afterwards. Once a large enough swathe of novelty becomes *possible*, SDE shows that there is a chance the structure behind the novelty becomes necessary. SDE also contributes to reconciling the role of structures with the Modern Synthesis [57]. After all, what determines the innovations produced by mutations is simply the existing structures within organisms. For example, exaptation has a natural home in SDE: a structure that creates a new space of possibility is also creating a new role for itself. Feathers, for example, originally selected for temperature homeostasis, also create a new space of possibility, which includes flight and, from there, the entire evolutionary space of birds, which were impossible prior to feathers [64]. SDE

shows that when such a new space is created, the innovation locks in and its continued existence becomes divorced from the original reason the innovation was selected for.

Another important phenomenon can be made sense of in the light of this theory: the empirical insight that the most evolutionarily impactful innovations are actually quite silent in the moment of their invention [65]. There is evidence, for example, that a long period of cryptic evolution has been present before the Cambrian explosion [66]. Closer to home, this phenonmenon was seen in the Long Term Evolution Experiment (LTEE) when aerobic citrate utilization evolved among lineages of *E. coli.* [54, 67–69]This ability to use citrate requires several prior mutations, or "potentiation", before it can be "actualized" by later mutations [54]. Our study shows that a potentiation structure, if it opens a large enough space, will likely lock-in and form a part of historical contingency [69]. However, Blount *et al.*'s work also shows the current experimental challenges to SDE: potentiation structures can be evolutionarily very quiet. Whatever the size of the space they have opened up, that space remains silent until the actualization, and observers only see the actualization.

The macro-evolutionary impact of an evolutionary change is not the ecological effect it has in the moment, but the size of the novel evolutionary space the lineage gains. A large space "sucks" the lineage into it, and its future ecological effects take place inside the new set of possibilities. In this way, the opening of new evolutionary space becomes the macro-evolutionary equivalent of fitness; a new structure that opens new evolutionary space has a probability of "locking-in", in the same way that a mutation that increases fitness has a probability of sweeping. This is happening even when all that observers can see are selective sweeps of populations through differences in fitness. What the observer cannot observe, however, are the enormous spaces of mere potential that might be opening or closing at each selective sweep. On the other hand, these invisible spaces exert a real evolutionary force, metaphorically similar to a vacuum, because a large potential space will "suck in" evolutionary trajectories. Such a large space forms the underlying and invisible universe of the possible from which the actual and observed are chosen. Their appearance, and the subsequent entry of the lineage into that new space, could be observed as punctuated equilibrium: the sudden creation of new taxa, composed of phenotypes that were previously impossible.

As an alternative to SDE, the theory of constructive neutral evolution proposes that complex structures can be created by duplication, each duplicate accumulating deleterious mutations [70, 71]. We similarly explain the evolution of hierarchies when components join together and can no longer depart because of the accumulation of deleterious mutations. Both theories are based on the idea that there are more ways to be different than to be the same – and that this can form a force to drive organisms to become more complex ([59] made a very similar point). Here we show that such macro-evolutionary dynamics can be extended from neutral evolution to regimes of high natural selection based on selective sweeps. The strength of supply-driven evolution is even predicted to increase with selection strength by reducing the number of evolutionarily relevant mutants, making the order of mutations increasingly important.

There are a plethora of future directions. For example, we have asserted the existence of mutations that make new evolutionary space possible. This is equivalent to assuming a certain topology of permissible evolutionary space, one that resembles a tree - such that a particular direction opens new fields and closes old ones. This type of topology was hypothesized to drive "epochal evolution" (*sensu* [72, 73]) and was shown to be satisfied by several fitness functions [74]. However, it is still unclear why this topology would be found in nature.

Anther future direction would be the experimental study of SDE. That would first involve constructing a mutation library of as many mutants as possible of a certain lineage, then classifying this mutation library according to the phenotype of each

mutant: its fitness and whichever trait we are interested in (e.g. [52, 75]). This library can then be used to construct ζ, the fitness distribution of mutants and of their trait value. One can then test predictions from SDE, *i.e.* long term trends in evolution through mutation bias, against the observed lineage evolution. Constructing such a library at multiple moments would further allow testing the stationarity of ζ, an open empirical question. The next step, the search for potentiation structures, is challenging. From mutation libraries, we would have to sequence and understand changes in structure that can be mapped to changes in these libraries. This would finally allow the identification of the emergence (newly available) or loss (newly lethal) of phenotypes.

Conclusion

We hope that SDE can significantly shift the discourse on highly contentious topics of evolutionary biology, such as the evolution of altruism, culture, intelligence, and other traits of human interest – or that make humans special. Instead of taking a selectionist view that carefully studies the environmental conditions in which certain structures (or behaviors) are promoted and maintained, we might instead want to ask how new mutations are generated, and what new possibilities are created. Such a view of evolution, and indeed, of culture, would give novelty and innovation their proper place in evolutionary theory. Perhaps evolution is biased not only in favor of survival, but also in favor of creativity.

Beyond that, we can gain a scientific understanding of historical contingency, instead of relegating this vast phenomenon under the rubric of "chance", or simple path-dependency. There is the hope of finally understanding two classes of change, one that is potentially historically significant, and one that is not. Potentially historically significant changes might be those where new structures come about that open up large spaces of new possibility. These are the structures that have a chance to be frozen in time – not to be changed for all descendants of a lineage. Historically insignificant structures do not open up new space. However useful they are for the moment, insignificant structures can be discarded when the environment changes. These and other possibilities would certainly enrich modern evolutionary theory.

Appendix A: Details of the NK model for supply-driven evolution

We define an NK model that is similar to, but different from, the model described in [27]. Agents are now strings of L bits, either 0 or 1. Each bit is an "element". Each element is linked to K other element in the agent, forming L substrings of $K+1$ length. Before running the model, a "fitness table" is generated. In this table, a fitness value is associated with each possible substrings of $K+1$ length. The agent's fitness is simply the average fitness of all its L substrings, by looking up each substring in the fitness table. Since fitness in this landscape depends on elements that are linked together, the effect of any one component value depends on other element values. In the original NK model, the fitness table is populated by fitnesses that are randomly chosen from 0 to 1.

Unless stated otherwise, the population consists of 5000 agents, each of which is a string of $N = 40$ elements. K is set to 3. The reason for this is is detailed in the next paragraph. All agents are wired identically in a wiring diagram that is chosen at the beginning of each model run. Each element has equal chance of being connected to any other element in any order. Element values is either 0 or 1. The fitness table represents the environment and will be specified below. When the environment changes, this change is represented by choosing a number of entries from the fitness table and

changing the fitness values associated with these substrings. The rate of environmental change is the same for all models, at each timestep, each entry in the fitness table has 10^{-5} probability of being changed.

The reason K is set to 3 is because if K is very small, the fitness table is also very small (it is, after all, simply 2^{K+1}). For $K = 2$, the fitness table is only 8 in size. This means that any environmental change is very large in scope, as the fitness of $\frac{1}{8}$ of all substrings are changed. This introduces a wild source of noise into our runs, which can be seriously altered with any change to the fitness table. With $K = 3$, this noise is much reduced. Higher levels of K slow down the run. For all values of $K > 1$, however, the outcomes are qualitatively the same.

For the run itself, at each timestep, N agents are chosen with replacement for update. An agent is selected with probability proportional to its fitness to reproduce one other agent. Another agent is then randomly selected to die. This is a Moran process. Thus, we can set α everywhere in our equations to be equal to 1. The descendent is identical to agent k, except for mutations. Each element of the parent agent k has 10^{-5} probability of being different in the child. If a element mutates between parent and child, the value of that element will either increase by flipping from 0 to 1, staying as 1 if it already was 1, or decrease by flipping from 1 to 0, staying as 0 if it already was 0.

We will study the evolution of the trait $\phi = \sum_{l=1}^{N} C_l$. We generate a fitness table such that phenotypes with higher trait values have, on average, higher fitnesses. For any substring $s_1 s_2 \ldots s_{K+1}$, its entry in the fitness table is a random number chosen from the exponential distribution with the mean $= 1.1^{\sum_{i=1}^{K+1} s_i}$. Thus, each increment of 1 in the substring leads to a 10% increase in fitness. Overall, then, an agent with larger values in each of its elements will also have higher fitness. However, the variable fitness assumption is satisfied because many different agents can have the same trait value, but have different fitnesses.

A mutation bias is here introduced where there is 90% chance that, if a mutation happens, the element decreases. There is thus a 10% chance that, if a mutation happens, the element increases.

Each time an entry from the fitness table changes due to environmental change, it is re-sampled from the same distribution, $\lambda e^{-\lambda x}$. Here x is not the same as ω as in the previous models, since x denotes only the fitness contribution of a substring to the whole organism's fitness, not ω, which is the organism's whole fitness. This does not happen if there is no environmental change.

This specifies the NK model used in the paper.

We ran this model with the above specified parameters for 200 runs, for 1e5 time steps each. We ran this model for another 200 runs without mutation bias. Final results can be seen Figure 2.

We also present Figure S1, a figure for individual results of each of 200 runs with mutation bias.

An interesting side effect of mutation pressure depressing the trait value over time is that the average population fitness over time can fluctuate and have long periods of decrease over time (Figure S2).

To test our theoretical predictions, let ϕ be our trait value, which is the number of 1's in an organism, while $N - \phi$ is the number of 0's in an organism. We observe the following: excepts for rare spurts of selective sweeps, the population usually has a dominant, homogeneous wildtype (see Figure S1). The ϕ value of the wild type is more or less the population ϕ value. In this population, then, ζ_U is simply made from two parts: probability $\dfrac{0.9\phi}{N}$ to increase the element value, and probability $\dfrac{0.1(N-\phi)}{N}$ to decrease the element value. As the population evolves and ϕ changes, ζ_U also changes.

The distribution of benefit $\beta(u)$ at any one point can be simply obtained. At pre-defined times (in the runs, they are measured 10 times over the entire run at equal

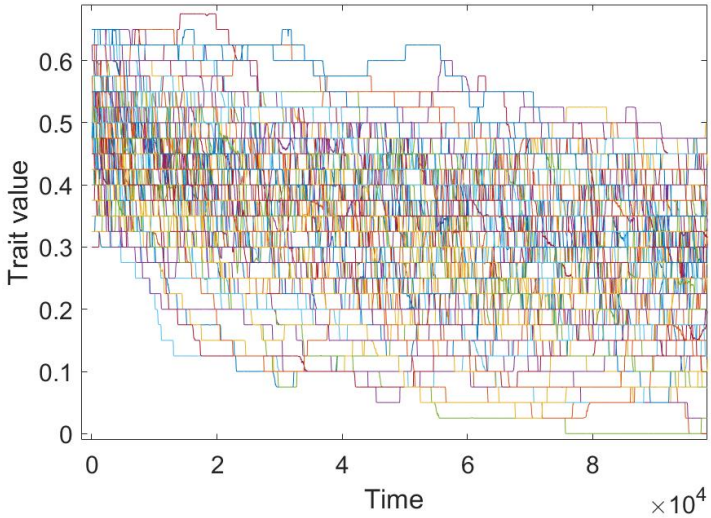


Fig S1. All 200 runs of the model, with trait value plotted against time. Casual examination shows that trait value decreases across time.

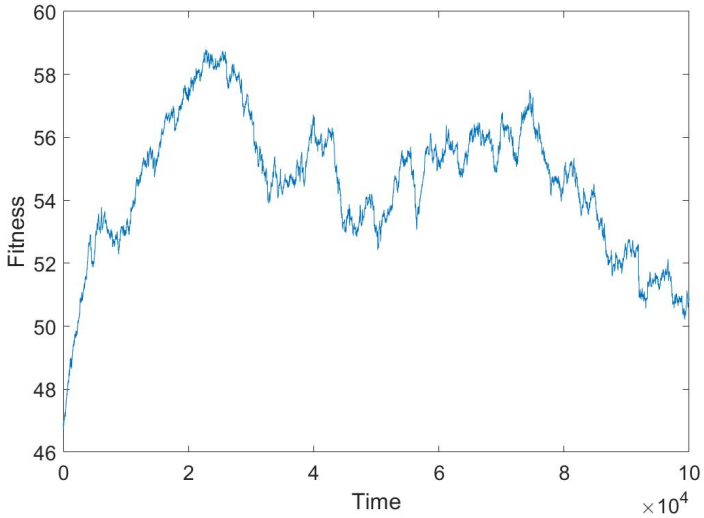


Fig S2. This figure shows how mean fitness changes over time, averaged over all 200 runs of the model

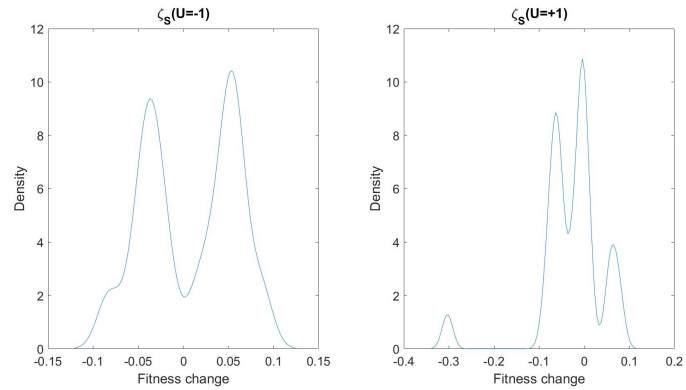


Fig S3. An example of the ζ_S distribution of one of the model runs, at t=1000. The panel on the left is the fitness distribution of all the mutants that decrease trait value by 1, the panel to the right is the fitness distribution of all the mutants that increase trait value by 1.

intervals), 10000 mutants of the wildtype are examined, without mutation bias – 5000 of these mutants decrease trait value and 5000 of them increase trait value (mutants that does not change trait value are not examined). These mutants are not real in that they are not a part of the population, they are only used for calculating β. The fitness of each of these mutants are noted and $\zeta_S(s|U=u)$ is constructed. An example of a ζ_S is shown below, for one of the runs at time = 10000. Since there are only two possibilities for trait value change in our model, we present the two distributions $\zeta_S(s|U=-1)$ and $\zeta_S(s|U=+1)$:

From this ζ_S, it is simple to compute

$$\beta(u) = \alpha \int_0^\infty \zeta_S(s|U=u) ds \mathbb{E}\zeta_S(s|U=u, s>0) \tag{13}$$

per our definition of β, using $\alpha = 1$ since we implemented a Moran process.

By computing β is this way at a certain time step, we can attempt to predict how the trait value will evolve, since we know how ζ_U changes. We assume β remain relatively constant. We can then use these parameters to predict the evolution of the trait value ϕ in the model. We choose $t = 10000$ as the time to measure β, as that is the moment where the population had settled down to some sort of equilibrium with the environmental change. The results are seen also in Figure 2. The fit between theory and agent-based model is good rather than perfect, showing that β does change over time, but better work in predicting how it changes over time will have to be in the future.

Appendix B: Details of the NK model for the rise in hierarchies

We will detail here the NK model used to show how increases in hierarchical organization can be irreversible.

We start with a population similar NK model as in Appendix A, with a population of 2000 agents, evolving in a Moran process. The initial N for all these agents, however, is 32 (we use exponentials of 2 for N so they can be readily cleaved or doubled). The initial K is much larger than before, $K = 10$, we will see the reason for this below. The environment still changes as above, there is a probability of 10^{-5} of any entry in the fitness table being changed at any timestep. The fitness table is now a lot larger, of course, due to K being so large.

Mutations in the previous model changed the element value of the agent from 0 to 1 or vice versa. Here, there are no mutations that change element value - whether an element is 0 or 1 is set at the very beginning of the run with 50% chance of each. Instead, mutations in this model change the wiring between two elements. As we have seen, a NK model is defined by its elements as well as the wiring among those elements. Thus, a mutation in our model will take a link between two element i and j, and throw the second element to a random other element, $k \neq j$. This potentially changes a substring in the organism and thus potentially changes its fitness.

There is an issue with implementing mutation in this straight forward way, however, because we will see that the number of links in this model can change over time through joining and cleaving mutations (detailed below). Thus, the effect size of a mutation changes based on the number of wiring that the agent has. In order to normalize this effect, we define a probability, $probRewire$, in our model, such that when an agent is to be mutated, all of its links have this probability to be rewired. We set $probRewire = 0.005$ for our model runs. This means approximately 2.5% of all the substrings in an agent is changed whenever the agent is mutated.

There are two other types of mutation that happens in this model, joining and cleaving mutations. The probabilities of these two types of mutations are set at 0.025 each, such that 0.95 of all mutations changing the wiring of an agent. Thus, when a mutation happens, there is 0.025 probability that two organisms join together and form a single joint organism, and the same probability that an organism is cleaved into its halves, the two component organisms. When a cleaving mutation happens, all wiring between the two components are lost. In order to keep K the same, these wiring are thrown to random other elements in the component organism.

The three types of mutations are mutually exclusive. In order to keep the population size the same, when there is a cleaving mutation, one of the components is randomly removed, and when there is a joining mutation, one of the two original organisms is randomly retained.

Because of the way that the cleaving mutation is implemented, its effects are mitigated

There is one other issue of implementing joining and cleaving mutations in the straightforward way. The problem is that, given an existing wild type, there is exactly one possible outcome of a joining mutation, and exactly one possible outcome of a cleaving mutation. This is a violation of the multiple fitness assumption, that there are many different ways of joining or cleaving in length, each with a different outcome in fitness. This means that there is no mutation bias effect possible in the joining or cleaving events, since at any one point, the fitness outcome of a joining or a cleaving event is predetermined. It also makes several of the results below difficult to analyze. To overcome this, we added a process whereby after any joining mutation or cleaving mutation, there is also a rewiring mutation, where 0.5% of all wires of the joint or component organisms are mutated. This effectively means that there are many ways for joining mutations to happen, and many different ways for cleaving mutations to happen. This probably corresponds to biological reality: there is more than one way for increased hierarchies to evolve, and there is also more than one way for hierarchies to dissolve.

This completes our description of this model. To generate the results we see in Figure 3, we run this model 8 times for 100,000 time steps. For panels 3d-f, we took one of the runs in which joining happened and was locked in place, and we measured the fitness distribution of all the mutants of the joint organism (black line), as well as the fitnesses of the component organisms that would result from a cleaving mutation (red line). We do so by examining 10,000 mutants of the joint organism (wiring only), and 10,000 component organisms that results from a cleaving mutation after the wiring mutations. We do so three times: 100 time steps after the joining mutation manages to

sweep the population (Figure 3d), 1000 time steps after the joining mutation has swept (Figure 3e), and 10000 time steps after the joining mutation has swept (Figure 3f). We see that the red distribution becomes less and less fit as time goes on. We see that after sufficient time, the fitness of the component organisms decay so much that it becomes basically impossible for any component organism to sweep the population. The decay of the average fitness of the component organism comes can be seen in Figure 3j.

Figure 3g-i show the mutation bias that drive the system. Figure 3g show the fitness distribution of 10,000 mutants of a joint organism (wiring only). Many of the mutations are neutral, but the vast majority are detrimental and less than 7% are beneficial. Figures 3h and 3i show the empirical ζ in the system. We examined 10,000 mutants (wiring only) of the joint organism and their effects on the fitness of the wild-type (y-axis), as well as their effect on the component organisms (x-axis). Figures 3h shows the overall ζ distribution, Figure 3i shows only those mutants that are advantageous for the joint organism. We can see that most of these mutations are disadvantageous for the component mutant, which means that the fitnesses of the component organisms decrease over time.

Appendix C: Evolutionary locking-in by the discovery of new evolutionary space

We start with equation (11). For simplicity's sake, we first give the proof of the case where $p = q = \beta$ and $B = 1$. This gives the basic methods of the proof that can be generalized later. In that case, equation (11) becomes:

$$\mathbf{S} = \begin{bmatrix} 0 & \frac{N}{N+1} & ... & ... & ... & ... & \frac{1}{N+1} \\ \frac{1}{N+\mu} & 0 & \frac{N-1}{N+\mu} & ... & ... & ... & \frac{\mu}{N+\mu} \\ 0 & \frac{2}{N+\mu^2} & 0 & \frac{N-2}{N+\mu^2} & ... & ... & \frac{\mu^2}{N+\mu^2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & ... & ... & \frac{N-1}{N+\mu^{N-1}} & 0 & \frac{1}{N+\mu^{N-1}} & \frac{\mu^{N-1}}{N+\mu^{N-1}} \\ 0 & ... & ... & ... & \frac{N}{N+\mu^N} & 0 & \frac{\mu^N}{N+\mu^N} \\ 0 & ... & ... & ... & ... & 0 & 1 \end{bmatrix}$$

In $\mathbf{S}$, all ... represent zeros. The first $N+1$ rows represent organisms with n new mutations ($0 \leq n \leq N$) that are realized. We discount mutations that simply return n to n, hence the zeroes on the diagonals. The probability of increasing n is simply $\frac{N-n}{N+\mu^n}$, of decreasing n is $\frac{n}{N+\mu^n}$, and of successfully (that is, selectively advantageous) breaking σ and destroying all n mutations is $\frac{\mu^n}{N+\mu^n}$. $\beta = p = q$ cancels out through the numerators and denominators. In reality, β becomes a time scaling factor – it measures the time between two selective sweeps, so that time flows faster for larger β and slower for smaller β.

The last row in $\mathbf{S}$ is the absorbing state. If σ is destroyed, we make the conservative assumption that it stays destroyed and cannot come about again. This makes our system a discrete phase-type distribution and is entirely characterized by $\mathbf{S}$ minus its

last row and last column, knowing that all the rows sum up to 1 and the last row is the absorbing state:

$$M_N := \begin{bmatrix} 0 & \frac{N}{N+1} & 0 & \dots & \dots & \dots & 0 \\ \frac{1}{N+\mu} & 0 & \frac{N-1}{N+\mu} & \ddots & & & \vdots \\ 0 & \frac{2}{N+\mu^2} & 0 & \frac{N-2}{N+\mu^2} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \frac{N-2}{N+\mu^{N-2}} & 0 & \frac{2}{N+\mu^{N-2}} & 0 \\ \vdots & & & \ddots & \frac{N-1}{N+\mu^{N-1}} & 0 & \frac{1}{N+\mu^{N-1}} \\ 0 & \dots & \dots & \dots & 0 & \frac{N}{N+\mu^N} & 0 \end{bmatrix}$$

and its starting state is the vector $\alpha = [1, 0, \dots, 0]$. By the standard theory of absorbing Markov chains, the quantity of interest, the time until the system reaches the absorbing state, is given by

$$\mathbb{E}[T_B] := \alpha(I - M_N)^{-1}\mathbf{1}, \tag{14}$$

where I is the identity matrix and $\mathbf{1}$ is the vector of all ones.

The upper limit case: $\mu = 1$

The simplest case of (14) arises in the limiting case $\mu = 1$. When this occurs, the matrix M_N can be written as $M_N = \frac{1}{N+1}Q_N$, with $Q_N\mathbf{1} = N\mathbf{1}$. Hence, the equation

$$(I - M_N)x = \mathbf{1}$$

has the obvious solution $x = (N+1)\mathbf{1}$, meaning that $\mathbb{E}[T_B] = N + 1$ (in fact, this is true no matter what the initial state is). This means that, when the probability of breaking σ is a constant independent of the number of mutations n, the expected time to reversal is proportional to the number of possible mutations.

0.1 The lower limit case: $\mu = 0$

In the second simplest case of (14), when $\mu = 0$, we also have identical denominators throughout M_N, except for the first row:

$$M_N := \begin{bmatrix} 0 & \frac{N}{N+1} & 0 & \dots & \dots & \dots & 0 \\ \frac{1}{N} & 0 & \frac{N-1}{N} & \ddots & & & \vdots \\ 0 & \frac{2}{N} & 0 & \frac{N-2}{N} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \frac{N-2}{N} & 0 & \frac{2}{N} & 0 \\ \vdots & & & \ddots & \frac{N-1}{N} & 0 & \frac{1}{N} \\ 0 & \dots & \dots & \dots & 0 & \frac{N}{N} & 0 \end{bmatrix}$$

The strategy we will adopt is to solve the system $(I - M_N)x = \mathbf{1}$ from the bottom up. Numbering the entries of x from 0 to N, the last equation in the system simplifies to

$$x_N - x_{N-1} = 1,$$

the first one simplifies to

$$x_0 - \frac{N}{N+1} x_1 = 1,$$

and the remaining ones take the form

$$x_k - \frac{k}{N} x_{k-1} - \frac{N-k}{N} x_{k+1} = 1 \ \ \forall\, 1 \le k \le N-1.$$

While solving the system in decreasing order of k values, it is easy to see that in fact, the solution satisfies

$$x_{N-k} - x_{N-k-1} = \frac{P_{N-k}(N)}{Q_{N-k}(N)} \ \ \forall\, 1 \le k \le N-1,$$

where both P_{N-k} and Q_{N-k} are polynomials of the same degree k. In fact, the explicit form for P_{N-k} and Q_{N-k} is given by

$$P_{N-k}(N) = k! \sum_{j=0}^{k} \binom{N}{j}; \ Q_{N-k}(N) = k! \binom{N-1}{k}.$$

We now show both of these statements simultaneously by induction on k. The base case, $k = 0$, results in $P_N(N) = 1 = Q_N(N)$, which is consistent with the initial equation $x_N - x_{N-1} = 1$. Suppose that this result holds for some value $k - 1$ with $k > 0$.

Substituting it into the $(N-k)$-th equation in the system yields

$$\begin{aligned} x_{N-k} - \frac{N-k}{N} x_{N-k-1} - \frac{k}{N} x_{N-k+1} = 1 \implies \\ x_{N-k} - \frac{N-k}{N} x_{N-k-1} - \frac{k}{N}\left(x_{N-k} + \frac{P_{N-k+1}(N)}{Q_{N-k+1}(N)}\right) = 1 \iff \\ \frac{N-k}{N}\left(x_{N-k} - x_{N-k-1}\right) = 1 + \frac{kP_{N-k+1}(N)}{NQ_{N-k+1}(N)} \iff \\ x_{N-k} - x_{N-k-1} = \frac{NQ_{N-k+1}(N) + kP_{N-k+1}(N)}{(N-k)Q_{N-k+1}(N)}. \end{aligned}$$

The desired statement now follows from observing that the denominator is

$$(N-k)Q_{N-k+1}(N) = (N-k)(k-1)!\binom{N-1}{k-1} = \frac{(N-1)!}{(N-k-1)!} = k!\binom{N-1}{k},$$

and that, by standard properties of the binomial coefficients, the numerator is

$$\begin{aligned} NQ_{N-k+1}(N) + kP_{N-k+1}(N) = N(k-1)!\binom{N-1}{k-1} + k(k-1)!\sum_{j=0}^{k-1}\binom{N}{j} = \\ = k!\binom{N}{k} + k!\sum_{j=0}^{k-1}\binom{N}{j} = k!\sum_{j=0}^{k}\binom{N}{j}, \end{aligned}$$

which completes the inductive step and proves the explicit formulas for P and Q. Lastly, by setting $k = N-1$, we obtain the equation

$$x_1 - x_0 = \frac{(N-1)!\sum_{j=0}^{N-1}\binom{N}{j}}{(N-1)!\binom{N-1}{0}} = \sum_{j=0}^{N-1}\binom{N}{j} = 2^N - 1,$$

and upon substituting this into the first equation, we obtain the final result

$$x_0 - \frac{N(x_0 + 2^N - 1)}{N+1} = 1 \iff \frac{x_0}{N+1} = 1 + \frac{N(2^N-1)}{N+1} \iff x_0 = N2^N + 1.$$

It follows that in this extreme case, the time required to break σ grows exponentially with the number of possible mutations N.

The generic case: $0 < \mu < 1$

We now show that the value of interest, x_0, exhibits exponential, rather than polynomial, growth that when μ is bounded away from 1, by repeating the reasoning in the previous section in this regime. More specifically, we show that

$$\mathbb{E}[T_B] \in \tilde{\Theta}\left(\left(\frac{2}{1+\mu}\right)^N\right),$$

for all $0 < \mu < 1$, where $\tilde{\Theta}$ denotes the asymptotic growth order with logarithmic factors neglected.

Because of the more complex structure of the equations involved and the need to extract a single coefficient from the solution of equation (14), we use Cramer's rule [76] to compute $\mathbb{E}[T_B]$ as a ratio of two determinants, the denominator being the

determinant of the coefficient matrix $I - M_N$, and the numerator being the determinant of the same matrix, but with the first column replaced by the right-hand side vector $\mathbf{1}$:

$$\mathbb{E}[T_B] := \frac{|[\mathbf{1}; (I - M_N)_{-1}]|}{|I - M_N|}, \tag{15}$$

where X_{-1} means the matrix X without its first column and $X; Y$ denotes the vertical concatenation of the matrices X and Y.

Computing the denominator of (15)

To compute the denominator, we use the standard recurrence used for computing the determinant of a tridiagonal matrix, but start it from the bottom right, rather than the top left. Specifically, if a_i, b_i, c_i denotes the i-th element of the diagonal, sub-diagonal, and super-diagonal, respectively, the recurrence is

$$\Delta_0 = 1; \Delta_1 = a_{N+1}; \Delta_i = a_{N+2-i}\Delta_{i-1} - b_{N+2-i}c_{N+2-i}\Delta_{i-2}, 2 \le i \le N+1 \tag{16}$$

and Δ_{N+1} is then the determinant of the matrix (in fact, Δ_i is the determinant of the matrix consisting of the last i rows and columns of the original matrix).

In our case, we have

$$a_i = 1, 1 \le i \le N+1;\ b_i = -\frac{i}{N+\mu^i}, 1 \le i \le N;\ c_i = -\frac{N+1-i}{N+\mu^{i-1}}, 1 \le i \le N.$$

The recurrence then becomes

$$\Delta_0 = 1;\ \Delta_1 = 1; \Delta_i = \Delta_{i-1} - \frac{(i-1)(N+2-i)}{(N+\mu^{N+1-i})(N+\mu^{N+2-i})}\Delta_{i-2},\ 2 \le i \le N+1.$$

We begin by solving it in the extreme cases $\mu = 0$ and $\mu = 1$. When $\mu = 0$, the recurrence simplifies to

$$\Delta_0 = 1;\ \Delta_1 = 1; \Delta_i = \Delta_{i-1} - \frac{(i-1)(N-i+2)}{N(N+[i=N+1])}\Delta_{i-2},\ 2 \le i \le N+1,$$

and we will prove by induction that its solution is given by

$$\Delta_i = \frac{N(N-1)\dots(N-i+1)}{N^i} = \frac{N!}{(N-i)!N^i}, 0 \le i \le N.$$

The base cases with $i = 0$ and $i = 1$ correctly produce $\Delta_0 = 1 = \Delta_1$, and then

$$\Delta_i = \Delta_{i-1} - \frac{(i-1)(N-i+2)}{N^2}\Delta_{i-2} = \frac{N!}{(N-i+1)!N^{i-1}} - \frac{N!(i-1)(N-i+2)}{(N-i+2)!N^{i-2}N^2} =$$
$$= \frac{N!}{(N-i+1)!N^i}[N-(i-1)] = \frac{N!(N-i+1)}{(N-i+1)!N^i} = \frac{N!}{(N-i)!N^i},$$

completing the proof by induction.

This means that $\Delta_{N-1} = \frac{N!}{N^{N-1}}, \Delta_N = \frac{N!}{N^N}$ and finally, the recurrence with $i = N+1$ results in $\Delta_{N+1} = \Delta_N - \frac{1}{N+1}\Delta_{N-1} = \frac{N!}{N^N(N+1)}$. By using Stirling's approximation, we conclude that, asymptotically for N large,

$$\det(I - M_N)|_{\mu=0} \sim \sqrt{2\pi/N}e^{-N}. \tag{17}$$

To resolve the case $\mu = 1$, we proceed slightly differently; namely, we first scale the matrix M_N by a factor of $N+1$, and show that the eigenvalues of this scaled matrix,

which we denote by Q_N, are equally spaced between $-N$ and N with a step size of 2 (which includes 0 when N is even and does not include it when N is odd). It will then follow that the eigenvalues of $I - M_N = I - \frac{1}{N+1}Q_N$ have the form $\frac{2k+1}{N+1}$ for $0 \le k \le N$, and the determinant will follow immediately.

Let us consider the scaled matrix $Q_N := (N+1)M_N$, given below.

$$Q_N := \begin{bmatrix} 0 & -N & 0 & \dots & \dots & \dots & 0 \\ -1 & 0 & 1-N & \ddots & & & \vdots \\ 0 & -2 & 0 & 2-N & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 2-N & 0 & -2 & 0 \\ \vdots & & & \ddots & 1-N & 0 & -1 \\ 0 & \dots & \dots & \dots & 0 & -N & 0 \end{bmatrix}$$

The fact that its eigenvalues are equal to $-N + 2k, 0 \le k \le N$, is a classical result first published (without proof) by Sylvester in 1854 [77]. The first known proof is due to Cayley [78], and this matrix was also considered by Schrödinger [79] who determined the correct eigenvalues, but was unable to prove the result. Here we present the classical treatment due to Taussky and Todd [80], as well as the more sophisticated treatment due to Mark Kac, as reported by Munarini and Torri [81], since it is the latter that we will use to solve the general case.

The elementary approach for factoring the characteristic polynomial of Q_N, given (up to sign) by the determinant $\det(\lambda I - Q_N)$, consists of a sequence of steps that do not change its value, followed by a "deflation" procedure that isolates one of the eigenvalues and reduces the dimension by 1. This process is illustrated here for $N = 3$ [80]. First, we have

$$\begin{vmatrix} \lambda & 3 & 0 & 0 \\ 1 & \lambda & 2 & 0 \\ 0 & 2 & \lambda & 1 \\ 0 & 0 & 3 & \lambda \end{vmatrix} = \begin{vmatrix} \lambda+3 & 3 & 0 & 0 \\ \lambda+3 & \lambda & 2 & 0 \\ \lambda+3 & 2 & \lambda & 1 \\ \lambda+3 & 0 & 3 & \lambda \end{vmatrix} = \begin{vmatrix} \lambda+3 & 3 & 0 & 0 \\ 0 & \lambda-3 & 2 & 0 \\ 0 & -1 & \lambda & 1 \\ 0 & -3 & 3 & \lambda \end{vmatrix} = (\lambda+3)\begin{vmatrix} \lambda-3 & 2 & 0 \\ -1 & \lambda & 1 \\ -3 & 3 & \lambda \end{vmatrix}.$$

Here, denoting by R_i and C_j the i-th row and the j-th column, respectively, the first step replaces C_1 with $C_1 + C_2 + C_3 + C_4$, the second step subtracts R_1 from each of the remaining rows, and the third step expands the determinant along the first column.

Next, we get

$$\begin{vmatrix} \lambda-3 & 2 & 0 \\ -1 & \lambda & 1 \\ -3 & 3 & \lambda \end{vmatrix} = \begin{vmatrix} \lambda-3 & 2 & 0 \\ -1 & \lambda+1 & 1 \\ -3 & \lambda+3 & \lambda \end{vmatrix} =$$

$$\begin{vmatrix} \lambda-1 & 2 & 0 \\ \lambda & \lambda+1 & 1 \\ \lambda & \lambda+3 & \lambda \end{vmatrix} =$$

$$\begin{vmatrix} \lambda-1 & 2 & 0 \\ \lambda & \lambda+1 & 1 \\ 0 & 2 & \lambda-1 \end{vmatrix} =$$

$$\begin{vmatrix} \lambda-1 & 2 & 0 \\ 1 & \lambda-1 & 1 \\ 0 & 2 & \lambda-1 \end{vmatrix}.$$

Here, the first step replaces C_2 with $C_2 + C_3$, the second step replaces C_1 with $C_1 + C_2$, the third step replaces R_3 with $R_3 - R_2$, and the final step replaces R_2 with $R_2 - R_1$. This shows that the original matrix Q_3 has -3 as an eigenvalue, and the remaining ones are obtained by adding 1 to those of Q_2, which is easily turned into an inductive proof of the desired statement in the general case.

Since the set $2k+1, 0 \leq k \leq N$ forms the complete collection of eigenvalues of Q_N, it follows that the determinant of $I - M_N$ is exactly

$$\frac{(2N+1)!!}{(N+1)^{N+1}} := \frac{\prod_{k=0}^{N}(2k+1)}{(N+1)^{N+1}} = \frac{(2N+2)!}{(N+1)!(2N+2)^{N+1}},$$

and thus, by Striling's approximation, asymptotically for large N,

$$\det(I - M_N)|_{\mu=1} \sim (2\sqrt{2}/e)(2/e)^N. \tag{18}$$

Next, we illustrate the more complex method used by Mark Kac in his approach towards the computation of determinants such as those of Q_N. Towards this end, we define a quantity which is essentially the Cayley continuant [81], but with the signs reversed on the sub- and super-diagonals, which does not change its value as these entries only enter the determinant of a tridiagonal matrix in matching pairs as seen in the recurrence:

$$U_{n+1}(\theta, x) := \begin{vmatrix} \theta & -x & 0 & \dots & \dots & \dots & 0 \\ -1 & \theta & 1-x & \ddots & & & \vdots \\ 0 & -2 & \theta & 2-x & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 2-n & \theta & n-2-x & 0 \\ \vdots & & & \ddots & 1-n & \theta & n-1-x \\ 0 & \dots & \dots & \dots & 0 & -n & \theta \end{vmatrix}.$$

Note that the quantity we defined is indexed by its dimension (hence, the index is $n+1$ rather than n). Clearly, the desired quantity in our case is given by $U_{N+1}(\theta = 0, x = N)$.

We increment the index by 1 which adds the entries $-(n+1)$ and $(n-x)$ at positions $(n+2, n+1)$ and $(n+1, n+2)$, respectively. We then write down the recurrence in (16) in the form

$$U_{n+2}(\theta, x) = \theta U_{n+1}(\theta, x) + (n+1)(n-x)U_n(\theta, x), \tag{19}$$

and convert it into a differential equation over the exponential generating series

$$U(\theta, x; t) := \sum_{n \geq 0} U_n(\theta, x) \frac{t^n}{n!}.$$

We first multiply each term of (19) by $\frac{t^{n+1}}{(n+1)!}$ and sum over all $n \geq 0$. We drop the dependence on θ and x and use the initial conditions $U_0 = 1, U_1 = \theta$ to get

$$\sum_{n \geq 0} U_{n+2} \frac{t^{n+1}}{(n+1)!} = \theta \sum_{n \geq 0} U_{n+1} \frac{t^{n+1}}{(n+1)!} + \sum_{n \geq 0} (n+1)(n-x) U_n \frac{t^{n+1}}{(n+1)!}.$$

For the left-hand side, note that

$$U - U_0 = \sum_{n \geq 0} U_{n+1} \frac{t^{n+1}}{(n+1)!} \implies \frac{dU}{dt} = \sum_{n \geq 0} U_{n+1} \frac{t^n}{n!} = U_1 + \sum_{n \geq 0} U_{n+2} \frac{t^{n+1}}{(n+1)!}$$

For the right-hand side, the first term simplifies to

$$\sum_{n \geq 0} U_{n+1} \frac{t^{n+1}}{(n+1)!} = U - U_0.$$

The coefficient of the second term can be written as $(n+1)n - x(n+1)$, and the first part simplifies to

$$\sum_{n \geq 0} (n+1) n U_n \frac{t^{n+1}}{(n+1)!} = t^2 \sum_{n \geq 1} U_n \frac{t^{n-1}}{(n-1)!} = t^2 \sum_{n \geq 0} U_{n+1} \frac{t^n}{n!} = t^2 \frac{dU}{dt},$$

while the second part simplifies to

$$-x \sum_{n \geq 0} (n+1) U_n \frac{t^{n+1}}{(n+1)!} = -xt \sum_{n \geq 0} U_n \frac{t^n}{n!} = -xtU,$$

which, in combination with the previous terms, yields the overall result

$$\frac{dU}{dt} - U_1 = \theta(U - U_0) + t^2 \frac{dU}{dt} - xtU \implies (1 - t^2) \frac{dU}{dt} = (\theta - xt)U, \tag{20}$$

where we used the initial conditions $U_0 = 1, U_1 = \theta$. The resulting differential equation is separable, so its solution is readily obtained by standard techniques:

$$\frac{dU}{U} = \frac{\theta - xt}{1 - t^2} = \frac{1}{2} \left[\frac{\theta - x}{1 - t} + \frac{\theta + x}{1 + t} \right] \implies$$

$$\log(U) = \frac{\theta + x}{2} \log(1 + t) - \frac{\theta - x}{2} \log(1 - t) \implies$$

$$U(\theta, x; t) = \frac{(1+t)^{(\theta+x)/2}}{(1-t)^{(\theta-x)/2}} = \sum_{k \geq 0} \binom{(\theta + x)/2}{k} t^k \sum_{\ell \geq 0} \binom{(\theta - x)/2 + \ell - 1}{\ell} t^\ell.$$

where we use the initial condition $U(t=0) = U_0 = 1$ to identify the integration constant, and the definition of the binomial coefficients as well as the property

$$\binom{-(z+1)}{k} = (-1)^k \binom{z+k}{k}$$

to get the product expression. By expanding the convolution and recalling that U is the exponential generating function for the $U_n(\theta, x)$, we finally get

$$U_n(\theta, x) = n! \sum_{\{k+\ell=n\}} \binom{(\theta+x)/2}{k} \binom{(\theta-x)/2+\ell-1}{\ell} = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{\theta+x}{2}\right)^{\underline{k}} \left(\frac{\theta-x}{2}\right)^{\overline{n-k}},$$

where the falling factorial $x^{\underline{k}}$ and rising factorial $x^{\overline{k}}$ are respectively defined by

$$x^{\underline{k}} := x(x-1)\dots(x-k+1); \quad x^{\overline{k}} := x(x+1)\dots(x+k-1).$$

In particular, this implies the desired result

$$\begin{aligned} U_{N+1}(\theta=0, x=N) &= \sum_{k=0}^{N+1} \binom{N+1}{k} \left(\frac{N}{2}\right)^{\underline{k}} \left(\frac{-N}{2}\right)^{\overline{N+1-k}} = \\ &\frac{1}{2^{N+1}} \sum_{k=0}^{N+1} \binom{N+1}{k} \prod_{j=0}^{k-1} (N-2j) \prod_{\ell=0}^{N-k} (-N+2\ell) = \\ &\frac{\sum_{k=0}^{N+1} \binom{N+1}{k}}{2^{N+1}} \prod_{t=0}^{N} (N-2t) = \prod_{t=0}^{N} (N-2t), \end{aligned}$$

in agreement with our earlier result.

We will now proceed with computing the determinant in the general case. We will proceed similarly to the above, but alter the diagonal to contain powers of μ in addition to the constant, defining V_n as the resulting determinant:

$$V_{n+1}(\mu, \theta, x) := \begin{vmatrix} \theta+1 & -x & 0 & \dots & \dots & \dots & 0 \\ -1 & \theta+\mu^1 & 1-x & \ddots & & & \vdots \\ 0 & -2 & \theta+\mu^2 & 2-x & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 2-n & \theta+\mu^{n-2} & n-2-x & 0 \\ \vdots & & & \ddots & 1-n & \theta+\mu^{n-1} & n-1-x \\ 0 & \dots & \dots & \dots & 0 & -n & \theta+\mu^n \end{vmatrix}$$

and the value of interest to us is $V_{N+1}(\mu=\mu, \theta=N, x=N)/D_{N+1}(\mu)$, where

$$D_{N+1}(\mu) := (N+1)(N+\mu)\dots(N+\mu^N) = \prod_{j=0}^{N} (N+\mu^j).$$

It is immediate that $D_{N+1}(\mu)$ is an increasing function of μ with $D_{N+1}(\mu=0) = N^N(N+1)$ and $D_{N+1}(\mu=1) = (N+1)^{N+1}$, and it follows that in general, for all $0 \le \mu \le 1$, we have $D_{N+1}(\mu) = c(N,\mu)N^N(N+1)$, with

$$1 \le c(N,\mu) \le \left(\frac{N+1}{N}\right)^N \le e,$$

so $D_{N+1}(\mu)$ may be treated as independent of μ up to a multiplicative constant.

Rather than solving for V_{N+1} via a differential equation (which becomes a functional-differential equation due to the presence of μ) we simplify our task by expanding it as a polynomial in powers of μ and truncating the result.

Since V_{N+1} as a function of μ is a polynomial of degree $N(N+1)/2$, let us write $V_{N+1} := \sum_{k=0}^{N(N+1)/2} c_k \mu^k$. Let M be the matrix used to define V_{N+1}. Since the determinant of M is obtained as a (signed) sum of $M_{i,\pi(i)}$ over all permutations π of $N+1$ elements, we can also write this differently by grouping the permutations according to the set of diagonal terms (the fixed points of π) from which we select powers of μ (as opposed to N). If $S \subseteq [N+1]$ is this subset, the remaining products in aggregate equal the determinant of the matrix $M_{-S,-S}$ obtained by crossing out the rows and columns corresponding to S and setting $\mu = 0$ in what is left (here, we denote by $[L]$ the set of all integers between 1 and L inclusively). In other words, defining $|S| := \sum_{s \in S} s$,

$$V_{N+1} := \sum_{S \subseteq [N+1]} |M_{-S,-S}|_0 \mu^{|S|} = \sum_{k=0}^{N(N+1)/2} \mu^k \sum_{\{S \subseteq [N+1] | |S|=k\}} |M_{-S,-S}|_0. \tag{21}$$

In particular, note that the coefficient of μ^0 in V_{N+1} is obtained by using the expression for Δ_{N+1} when $\mu = 0$ and multiplying it by $N^N(N+1)$, the scaling factor. It follows that this coefficient is given by

$$c_0 := N^N(N+1)\frac{N!}{N^N(N+1)} = N!$$

We make two observations about the remaining coefficients c_k. First, the terms corresponding to μ^k with $k > N$ do not make a major contribution to the total because they require sets S of size at least 2, and the corresponding terms involve selecting several times the powers of μ from the diagonal terms, which are much smaller than the other component of the diagonal term, N, since $0 < \mu < 1$. Although there are many such sets, we will see that their contribution can be neglected.

Second, the coefficients corresponding to μ^k with $1 \le k \le N$ are dominated by the term corresponding to the singleton $S = \{k\}$, by the same reasoning. We will therefore obtain c_k to leading order with respect to N by computing the (k,k) cofactor of V_{N+1}, i.e. the determinant of the matrix obtained by crossing out its k-th row and column, with $\mu = 0$.

It is easy to see that this cofactor splits into the product of two terms, say $c_k = d_k e_k$. The first of these, d_k, is the determinant of the submatrix formed by the first k rows and columns of the matrix defining V_{N+1}, and the second one, e_k (up to a scaling factor of N^{N-k}), equals Δ_{N-k} when $\mu = 0$, i.e. the determinant of the submatrix formed by the last $N-k$ rows and columns of M_N when $\mu = 0$, computed previously, so

$$e_k := N^{N-k}\Delta_{N-k} = N^{N-k}\frac{N!}{k!N^{N-k}} = \frac{N!}{k!}. \tag{22}$$

It remains to compute d_k, obtained as V_k evaluated at $\mu = 0, \theta = x = N$. This can be achieved by altering the initial conditions in the original equation for U, (19), to

$U_0 = 1, U_1 = \theta + 1$. After an identical transformation this results in the inhomogeneous equation

$$(1-t^2)\frac{dU}{dt} = x(1-t)U + 1, \tag{23}$$

where we use the fact that $\theta = x = N$ in our case of interest here. Using the homogeneous solution as an ansatz, we set $U(t) = (1+t)^x f(t)$ and this results in a simpler differential equation for $f(t)$,

$$f'(t) = \frac{1}{(1-t)(1+t)^{x+1}},\ f(0) = 1,$$

whose solution is obtained by a partial fraction expansion of the right-hand side:

$$\begin{aligned} f'(t) &= \frac{1/2}{(1+t)^{x+1}} + \frac{1/2}{(1-t)(1+t)^x} = \frac{1/2}{(1+t)^{x+1}} + \frac{(1/2)^2}{(1+t)^x} + \frac{(1/2)^2}{(1-t)(1+t)^{x-1}} \\ &= \cdots = \sum_{j=1}^{x+1} \frac{(1/2)^{x+2-j}}{(1+t)^j} + \frac{(1/2)^{x+1}}{1-t} = \frac{1}{2^{x+1}}\left[\frac{1}{1-t} + \frac{1}{1+t} + \sum_{j=1}^{x} \frac{2^j}{(1+t)^{j+1}}\right] \\ &\implies f(t) = \frac{1}{2^{x+1}}\left(\log(1+t) - \log(1-t) + \sum_{j=1}^{x} \frac{2^j}{j}\left[1 - \frac{1}{(1+t)^j}\right]\right) + 1, \end{aligned}$$

where the last term and the square brackets account for the integration constant.

The expression $U(t) = (1+t)^x f(t)$ with $x = N$ finally becomes

$$U(t)|_{x=N} = (1+t)^N + \frac{1}{2^{N+1}}\left[\log\left(\frac{1+t}{1-t}\right)(1+t)^N + \sum_{j=1}^{N} 2^j \frac{(1+t)^N - (1+t)^{N-j}}{j}\right].$$

We note the following fact: since $U(t) = \sum_{k\geq 0} \frac{d_k}{k!} t^k$ and $e_k = \frac{N!}{k!}$, we have

$$\sum_{k\geq 0} c_k t^k = \sum_{k\geq 0} d_k e_k t^k = \sum_{k\geq 0} N! \frac{d_k}{k!} t^k = N! \sum_{k\geq 0} \frac{d_k}{k!} t^k = (N!)U(t),$$

and therefore, the expression of interest equals $(N!)U_{\leq N}(t = \mu)$, where $U_{\leq N}(t)$ denotes the truncated series

$$U_{\leq N}(t) := \sum_{k\leq N} \frac{d_k}{k!} t^k,$$

noting that the two expressions agree at 0 since $c_0 = N!$ (as shown previously).

To estimate this truncated series we note that the only part of $U(t)$ whose expansion extends beyond t^N is $\big(\log(1+t) - \log(1-t)\big)(1+t)^N$. We compute the series expansion of $\log(1+t) - \log(1-t)$, which is given by $\sum_{k\geq 0} \frac{2}{2k+1} t^{2k+1}$. Since all the coefficients of both $\log(1+t) - \log(1-t)$ and $(1+t)^N$ are non-negative, for $0 \leq t \leq 1$ we have

$$T_N[\big(\log(1+t) - \log(1-t)\big)(1+t)^N] \leq T_N[\log(1+t) - \log(1-t)]T_N[(1+t)^N],$$

where T_N is the truncation operator on a power series that returns its first $N+1$ terms; this holds for the convolution product of any two power series with non-negative coefficients since

$$\begin{aligned} T_N\left[\sum_{i\geq 0} u_i t^i \sum_{j\geq 0} v_j t^j\right] &= \sum_{i,j\geq 0} u_i v_j t^{i+j}[i+j \leq N] \leq \sum_{i,j\geq 0} u_i v_j t^{i+j}[i \leq N][j \leq N] \\ &= \sum_{i\geq 0} u_i t^i [i \leq N] \sum_{j\geq 0} v_j t^j [j \leq N] = T_N\left[\sum_{i\geq 0} u_i t^i\right] T_N\left[\sum_{j\geq 0} v_j t^j\right], \end{aligned}$$

where $[E]$ is the Iverson bracket whose value is 1 is expression E is true and 0 otherwise.

We also see that, on $0 \le t \le 1$, since $2(N-1)+1 \ge N$ for all $N \in \mathbb{N}$, we get

$$0 \le T_N[\log(1+t) - \log(1-t)] \le \sum_{k=0}^{N-1} \frac{2}{2k+1} \le 2\sum_{k=1}^{N} \frac{1}{k} \le 2\log(N).$$

It follows that

$$U_{\le N}(t) = (1+t)^N \left(1 + \frac{C(N,t)}{2^N}\right) + \frac{1}{2^{N+1}} \left[\sum_{j=1}^{N} 2^j \frac{(1+t)^N - (1+t)^{N-j}}{j}\right],$$

where $0 \le C(N,t) \le \log(N)$ uniformly for all N and $0 \le t \le 1$.

To estimate the second term of $U_{\le N}(t)$, we let $Z_N := \frac{\sum_{j=1}^{N} 2^j/j}{2^{N+1}}$ and compute

$$Z_N = \frac{1}{2} Z_{N-1} + \frac{1}{2N} \implies Z_N - \frac{1}{N} = \frac{1}{2}\left(Z_{N-1} - \frac{1}{N-1}\right) + \frac{1}{2N(N-1)},$$

and since Z_N is bounded above by 1 and below by $1/N$, it follows that $Z_N - 1/N$ must converge to 0 as N increases. It therefore follows that

$$U_{\le N}(t) = (1+t)^N \left(1 + \frac{C(N,t)}{2^N} + D(N,t)\right) - E(N,t),$$

where $D(N,t) \ge 0, D(N,t) - 1/N \to 0$ as $N \to \infty$ and $0 \le E(N,t) \le \log(N)$ as

$$E(N,t) := \frac{1}{2^{N+1}} \sum_{j=1}^{N} \frac{2^j (1+t)^{N-j}}{j} \le \frac{2^N}{2^{N+1}} \sum_{j=1}^{N} \frac{1}{j} \le \frac{\log(N)}{2} \le \log(N),$$

uniformly for all N and $0 \le t \le 1$, so that finally, $U_{\le N}(t) \approx (1+1/N)(1+t)^N$.

When we compute $(N!)U_{\le N}(t=1)$, using this approximation results in $2^N(N+1)!/N$, giving the denominator of (15) as $\frac{2^N(N+1)!}{N(N+1)^{N+1}}$, which Stirling's formula shows to be in agreement with the previously found result, equation (18), up to a factor of $\sqrt{N}$, which is therefore the cost of approximating the polynomial in (21) using only sets of size at most 1.

Computing the numerator of (15)

To compute the numerator of (15) we need to evaluate

$$W_{N+1}(\mu,\theta,x) := \begin{vmatrix} \theta+1 & -x & 0 & \dots & \dots & \dots & 0 \\ \theta+\mu^1 & \theta+\mu^1 & 1-x & \ddots & & & \vdots \\ \theta+\mu^2 & -2 & \theta+\mu^2 & 2-x & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \theta+\mu^{n-2} & & \ddots & 2-n & \theta+\mu^{n-2} & n-2-x & 0 \\ \theta+\mu^{n-1} & & & \ddots & 1-n & \theta+\mu^{n-1} & n-1-x \\ \theta+\mu^n & \dots & \dots & \dots & 0 & -n & \theta+\mu^n \end{vmatrix}.$$

The value of interest to us will then be $W_{N+1}(\mu = \mu, \theta = N, x = N)/D_{N+1}(\mu)$, where $D_{N+1}(\mu)$ remains exactly as defined in the previous subsection.

To evaluate W_{N+1} we begin by expanding it along the first column to get

$$W_{N+1}(\mu, \theta, x) = \sum_{j=0}^{N} (-1)^j (\theta + \mu^j) |M_{-(j+1),-1}|,$$

where M is the matrix whose determinant is W_{N+1}. Furthermore, we note that because M is tridiagonal, all the terms with the exception of the one with $j = 0$ split into a product of two terms, the first of which is the product of the first $j - 1$ super-diagonal entries and the second of which is the determinant of the last $N - j$ rows and columns of M. More precisely:

$$W_{N+1}(\mu, \theta, x) = \sum_{j=0}^{N} (-1)^j (\theta + \mu^j) \prod_{k=0}^{j-1} (k - x) |M_{-[j+1],-[j+1]}|,$$

where, as before, $[L]$ denotes the set of all integers from 1 to L.

This is because the shape of the matrix remaining after crossing out the $j + 1$-th row (using the numbering from 1 to $N + 1$) and the first column of M is a tridiagonal matrix with non-zero entries on the diagonal, sub-diagonal and sub-sub-diagonal (rather than the super-diagonal) above the original $j + 1$-th row, and a "standard" tridiagonal matrix below it.

The critical entry to consider is the entry $M_{j+2,j+1}$ in the original matrix, which becomes entry $(j + 1, j)$ in the truncated matrix. We show that any permutation $\pi \in S_N$ that contains this entry (i.e. that has $\pi(j + 1) = j$) produces a zero product for structural reasons, and does not contribute to the determinant. Indeed, crossing out the $j + 1$-th row and the j-th column in the truncated matrix leaves an imbalance - the top left corner is a possibly empty $j \times (j - 1)$ submatrix, the bottom right corner is a possibly empty $(N - j - 1) \times (N - j)$ submatrix, and the bottom left and top right corners both only contain 0's, meaning that any permutation yields a zero product, as claimed.

It now follows that the determinant of the truncated matrix factors into the product of two determinants. The first one is the determinant of the $j \times j$ tridiagonal matrix with non-zero entries only on and below the diagonal, which is therefore also lower-triangular and thus, its determinant equals the product of its diagonal entries, which were the first j super-diagonal entries of M. The second one is the determinant of the $N - j \times N - j$ tridiagonal matrix containing the last $N - j$ rows and columns of M, leading to the desired conclusion.

We now recall that, by the reasoning in the previous section, the dominant contributions to various terms involving μ in the definition of W_{N+1} come from permutations where μ^k is a monomial arising from the selection of μ^k in the k, k-th entry of M while the remainder is approximated by setting $\mu = 0$ in the corresponding determinant. It follows that

$$\begin{aligned} W_{N+1}(\mu, \theta, x) &\approx \sum_{j=0}^{N} (\theta + \mu^j) \prod_{k=0}^{j-1} (x - k) |M_{-[j+1],-[j+1]}|_{\mu=0} \\ &= (\theta + 1)\Delta_N + \theta \sum_{j=1}^{N} \prod_{k=0}^{j-1} (x - k)\Delta_{N-j} + \sum_{j=1}^{N} \prod_{k=0}^{j-1} (x - k)\Delta_{N-j}\mu^j \\ &= N! \left[(\theta + 1) + \theta \sum_{j=1}^{N} \frac{1}{j!} \prod_{k=0}^{j-1} (x - k) + \sum_{j=1}^{N} \frac{1}{j!} \prod_{k=0}^{j-1} (x - k)\mu^j \right], \end{aligned}$$

where we distributed the negative sign across the product in the first line and used the definitions $e_j = N^{N-j}\Delta_{N-j} = |M_{-[j+1],-[j+1]}|_{\mu=0}$ in the second line and their known values from (22) in the third line.

Upon substituting $\theta = x = N$ in the above, we finally obtain

$$W_{N+1}(\mu, N, N) \approx N! \left[(N+1) + N \sum_{j=1}^{N} \frac{1}{j!} \prod_{k=0}^{j-1} (x-k) + \sum_{j=1}^{N} \frac{1}{j!} \prod_{k=0}^{j-1} (x-k)\mu^j \right]$$
$$= N! \left[(N+1) + N \sum_{j=1}^{N} \binom{N}{j} + \sum_{j=1}^{N} \binom{N}{j} \mu^j \right] = N!(N \cdot 2^N + (1+\mu)^N),$$

where we used the Binomial theorem in the last line. We verify that this yields correctly yields $N!(N \cdot 2^N + 1)$ when $\mu = 0$ and $(N+1)! \cdot 2^N$ when $\mu = 1$, the latter value being once again within a factor of $\sqrt{N}$ from the correct one, $(2N+1)!! \cdot (N+1)$, according to Stirling's theorem and the work in equation (18).

The overall conclusion is that, asymptotically for N large and $0 \leq \mu \leq 1$, the expected time scales as

$$\mathbb{E}[T_B] \sim \frac{N!(N \cdot 2^N + (1+\mu)^N)}{N!(1+1/N)(1+\mu)^N} = \frac{N\left(\frac{2}{1+\mu}\right)^N + 1}{1+1/N} \approx N\left(\frac{2}{1+\mu}\right)^N + 1.$$

Note that, although the right-hand side happens to be exact for both $\mu = 0$ and $\mu = 1$, it may be off by as much as a factor of N in either direction for values strictly between 0 and 1 due to approximation errors; however, the rate of growth remains exponential in N provided that μ is bounded away from 1. However, numerical experiments suggest that the true value tends to be much closer to this estimate in practice, with relative errors on the order of $1/N$, meaning that in fact, the errors in the approximation of the numerator and of the denominator may be in the same direction.

Acknowledgments

We are indebted to Artem Kaznatcheev, Andre Costopoulos, Colin Wren, Jennifer Bracewell, and members of the Guichard lab for crucial discussions.

LC acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union.

References

1. Dietrich MR. Microevolution and macroevolution are governed by the same processes. Francisco J Ayala and Robert Arp. 2009; p. 169.

2. Erwin DH. Macroevolution is more than repeated rounds of microevolution. Evolution & development. 2000;2(2):78–84.

3. McShea DW. Mechanisms of large-scale evolutionary trends. Evolution. 1994;48(6):1747–1763.

4. Gregory TR. Evolutionary trends. Evolution: Education and Outreach. 2008;1(3):259–273.

5. Szathmary E, Smith JM. The major evolutionary transitions. Shaking the tree: readings from Nature in the history of life. 2000; p. 32.

6. Wimsatt WC. Developmental constraints, generative entrenchment, and the innate-acquired distinction. Integrating scientific disciplines. 1986; p. 185–208.

7. Schank JC, Wimsatt WC. Generative entrenchment and evolution. In: PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association. vol. 1986. JSTOR; 1986. p. 33–60.

8. Powell R. Contingency and convergence in macroevolution: a reply to John Beatty. The Journal of Philosophy. 2009;106(7):390–403.

9. Gould SJ. Punctuated equilibrium. Harvard University Press; 2009.

10. Calcott B, Sterelny K. Introduction: A dynamic view of evolution. In: The Major Transitions in Evolution Revisited. MIT Press; 2011. p. 1–14.

11. Fisher SRA. The genetical theory of natural selection: a complete variorum edition. Oxford University Press, USA; 1999.

12. Fontana W, Buss LW. "The arrival of the fittest": Toward a theory of biological organization. Bulletin of Mathematical Biology. 1994;56(1):1–64.

13. Stoltzfus A. Mutationism and the dual causation of evolutionary change. Evolution & Development. 2006;8(3):304–317.

14. Stoltzfus A, Yampolsky LY. Climbing mount probable: mutation as a cause of nonrandomness in evolution. Journal of Heredity. 2009;.

15. Stoltzfus A. On the possibility of constructive neutral evolution. Journal of molecular evolution. 1999;49(2):169–181.

16. Yampolsky LY, Stoltzfus A. Bias in the introduction of variation as an orienting factor in evolution. Evolution & Development. 2001;3(2):73–83.

17. Schaper S, Louis AA. The Arrival of the Frequent: How Bias in Genotype-Phenotype Maps Can Steer Populations to Local Optima. PLOS ONE. 2014;9(2):e86635.

18. Alberch P, Gale EA. A developmental analysis of an evolutionary trend: digital reduction in amphibians. Evolution. 1985;39(1):8–23.

19. Klingenberg CP. Developmental constraints, modules, and evolvability. In: Variation. Academic Press; 2005.

20. Arthur W. The interaction between developmental bias and natural selection: from centipede segments to a general hypothesis. Heredity. 2002;89(4):239–246.

21. Müller GB, Newman SA. Origination of organismal form: beyond the gene in developmental and evolutionary biology. MIT Press; 2003.

22. Hall BK, Pearson RD, M
"uller GB. Environment, development, and evolution: toward a synthesis. MIT Press Cambridge, Massachusetts; 2004.

23. Rice SH. A geometric model for the evolution of development. Journal of Theoretical Biology. 1990;143(3):319–342.

24. Rice SH. The evolution of canalization and the breaking of von Baer's laws: modeling the evolution of development with epistasis. Evolution. 1998;52(3):647–656.

25. Rice SH. The place of development in mathematical evolutionary theory. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 2012;318(6):480–488.

26. Klingenberg CP. Evolution and development of shape: integrating quantitative approaches. Nature Reviews Genetics. 2010;11(9):623–635.

27. Xue JZ, Costopoulos A, Guichard F. A Trait-based framework for mutation bias as a driver of long-term evolutionary trends. Complexity. 2015;.

28. Arthur W. Developmental drive: an important determinant of the direction of phenotypic evolution. Evolution & development. 2001;3(4):271–278.

29. Sniegowski PD, Gerrish PJ. Beneficial mutations and the dynamics of adaptation in asexual populations. Philosophical Transactions of the Royal Society B: Biological Sciences. 2010;365(1544):1255–1263.

30. Kim Y, Orr HA. Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. Genetics. 2005;171(3):1377–1386.

31. Haldane JBS. A mathematical theory of natural and artificial selection, part V: selection and mutation. In: Mathematical Proceedings of the Cambridge Philosophical Society. vol. 23. Cambridge University Press; 1927. p. 838–844.

32. Nowak MA. Evolutionary dynamics: exploring the equations of life. Belknap Press; 2006.

33. Johnson T, Gerrish PJ. The fixation probability of a beneficial allele in a population dividing by binary fission. Genetica. 2002;115(3):283–287.

34. Stoltzfus A. Mutation-biased adaptation in a protein NK model. Molecular biology and evolution. 2006;23(10):1852.

35. Kauffman S, Levin S. Towards a general theory of adaptive walks on rugged landscapes*. Journal of theoretical Biology. 1987;128(1):11–45.

36. Orgel LE. Some consequences of the RNA world hypothesis. Origins of Life and Evolution of the Biosphere. 2003;33(2):211–218.

37. Lang BF, Gray MW, Burger G. Mitochondrial genome evolution and the origin of eukaryotes. Annual Review of Genetics. 1999;33(1):351–397.

38. Bonner JT. The origins of multicellularity. Integrative Biology Issues News and Reviews. 1998;1(1):27–36.

39. Wilson EO, Hölldobler B. Eusociality: origin and consequences. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(38):13367.

40. Metzger MJ, Reinisch C, Sherry J, Goff SP. Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. Cell. 2015;161(2):255–263.

41. McCallum H. Tasmanian devil facial tumour disease: lessons for conservation biology. Trends in Ecology & Evolution. 2008;23(11):631–637.

42. Wcislo WT, Danforth BN. Secondarily solitary: the evolutionary loss of social behavior. Trends in Ecology & Evolution. 1997;12(12):468–474.

43. Clark CG, Roger AJ. Direct evidence for secondary loss of mitochondria in Entamoeba histolytica. Proceedings of the National Academy of Sciences. 1995;92(14):6518.

44. Horner DS, Embley TM. Chaperonin 60 phylogeny provides further evidence for secondary loss of mitochondria among putative early-branching eukaryotes. Molecular Biology and Evolution. 2001;18(10):1970–1975.

45. Tovar J, Fischer A, Clark CG. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite Entamoeba histolytica. Molecular microbiology. 1999;32(5):1013–1021.

46. Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, van der Giezen M, et al. Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. Nature. 2003;426(6963):172–176.

47. Tachezy J, Sánchez LB, Müller M. Mitochondrial type iron-sulfur cluster assembly in the amitochondriate eukaryotes Trichomonas vaginalis and Giardia intestinalis, as indicated by the phylogeny of IscS. Molecular Biology and Evolution. 2001;18(10):1919–1928.

48. Hampl V, Simpson AG. Possible mitochondria-related organelles in poorly-studied "amitochondriate" eukaryotes. In: Hydrogenosomes and mitosomes: mitochondria of anaerobic eukaryotes. Springer; 2007. p. 265–282.

49. Moran NA. Tracing the evolution of gene loss in obligate bacterial symbionts. Current opinion in microbiology. 2003;6(5):512–518.

50. Eyre-Walker A, Woolfit M, Phelps T. The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics. 2006;173(2):891.

51. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nature Reviews Genetics. 2007;8(8):610–618.

52. Monroe J, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, et al. Mutation bias reflects natural selection in Arabidopsis thaliana. Nature. 2022; p. 1–5.

53. Orr HA. The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. Journal of theoretical biology. 2006;238(2):279–285.

54. Blount ZD, Barrick JE, Davidson CJ, Lenski RE. Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature. 2012;489(7417):513–518.

55. Vrba ES. What is species selection? Systematic Zoology. 1984;33(3):318–328.

56. Jablonski D. Species selection: theory and data. Annual review of ecology, evolution, and systematics. 2008;39:501–524.

57. Gould S. The evolution of life on earth. Special Editions. 2004;.

58. Eldredge N, Thompson JN, Brakefield PM, Gavrilets S, Jablonski D, Jackson JBC, et al. The dynamics of evolutionary stasis. Paleobiology. 2005;31(2_Suppl):133.

59. McShea DW, Brandon RN. Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems. University Of Chicago Press; 2010.

60. Gould SJ. Wonderful life. Norton; 1989.

61. Beatty J. The evolutionary contingency thesis. In: Concepts, theories, and rationality in the biological sciences: the Second Pittsburgh-Konstanz Colloquium in the Philosophy of Science, University of Pittsburgh, October 1-4, 1993. Univ of Pittsburgh Pr; 1995. p. 45.

62. Beatty J. Chance variation and evolutionary contingency: Darwin, Simpson, the Simpsons, and Gould. na; 2008.

63. Szathmáry E, Calcott B, Sterelny K. The major transitions in evolution revisited. MIT Press; 2011.

64. Foth C, Tischlinger H, Rauhut OW. New specimen of Archaeopteryx provides insights into the evolution of pennaceous feathers. Nature. 2014;511(7507):79–82.

65. Erwin DH. A conceptual framework of evolutionary novelty and innovation. Biological Reviews. 2021;96(1):1–15.

66. Zhang X, Shu D. Current understanding on the Cambrian Explosion: questions and answers. PalZ. 2021; p. 1–20.

67. Blount ZD, Borland CZ, Lenski RE. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proceedings of the National Academy of Sciences. 2008;105(23):7899.

68. Wiser MJ, Ribeck N, Lenski RE. Long-term dynamics of adaptation in asexual populations. Science. 2013;342(6164):1364–1367.

69. Blount ZD, Lenski RE, Losos JB. Contingency and determinism in evolution: Replaying life's tape. Science. 2018;362(6415).

70. Gray MW, Lukeš J, Archibald JM, Keeling PJ, Doolittle WF. Irremediable Complexity? Science. 2010;330(6006):920.

71. Force M A Lynch, Pickett FB, Amores A, Yan Y, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999;151(4):1531.

72. Crutchfield JP, Van Nimwegen E. The evolutionary unfolding of complexity. In: Evolution as computation. Springer; 2002. p. 67–94.

73. Crutchfield JP. When evolution is revolution—origins of innovation. Evolutionary Dynamics—Exploring the Interplay of Selection, Neutrality, Accident, and Function, Santa Fe Institute Series in the Sciences of Complexity. 2003; p. 101–133.

74. Gavrilets S. Evolution and speciation on holey adaptive landscapes. Trends in ecology & evolution. 1997;12(8):307–312.

75. Menda N, Semel Y, Peled D, Eshed Y, Zamir D. In silico screening of a saturated mutation library of tomato. The Plant Journal. 2004;38(5):861–872.

76. Robinson SM. A short proof of Cramer's rule. Mathematics Magazine. 1970;43(2):94–95.

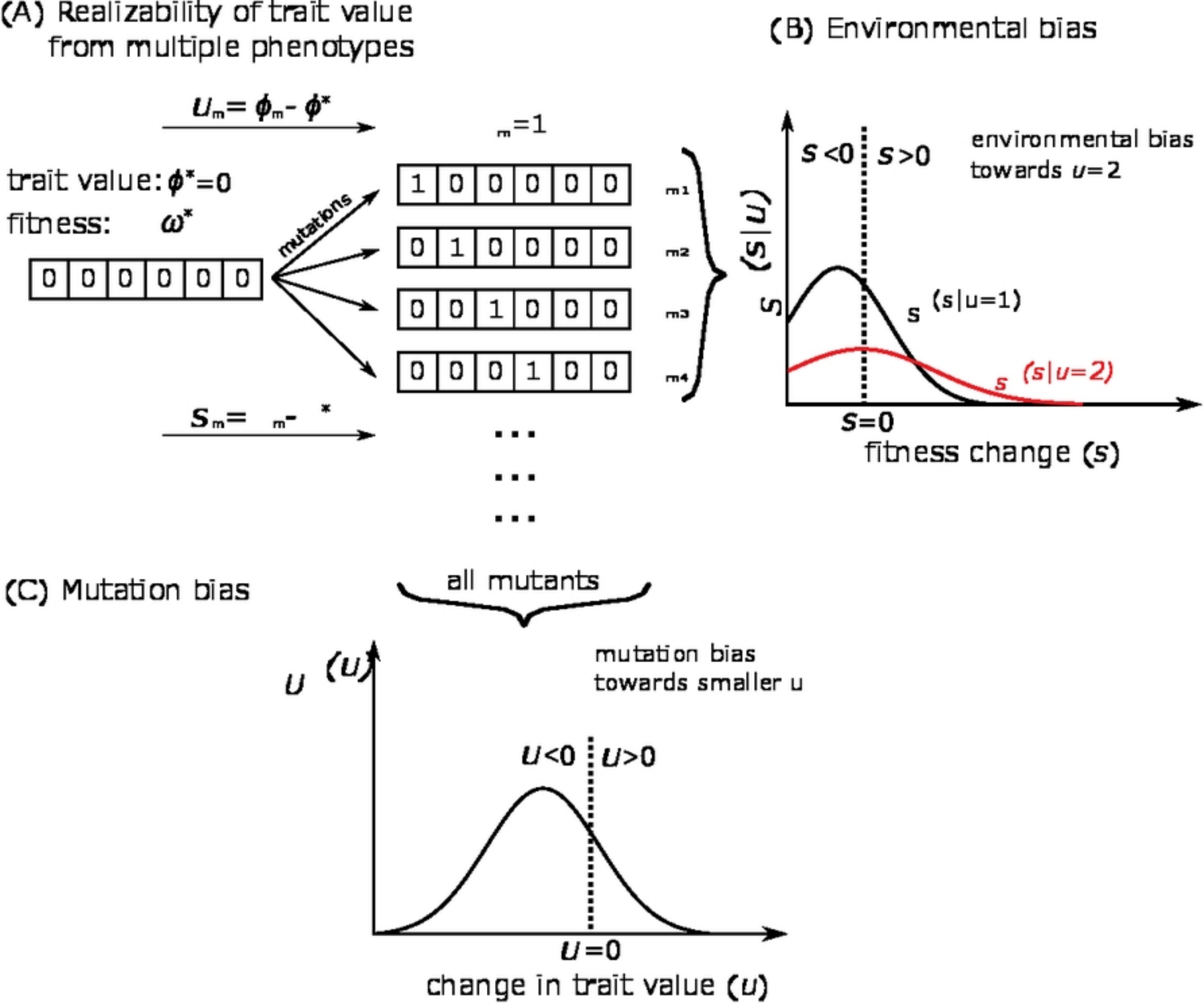
77. Sylvester JJ. Théorème sur les déterminants de M. Sylvester. Nouv Ann Math. 1854;13:305.

78. Cayley A. On the determination of the value of a certain determinant. Quart J Math. 1857;2:163–166.

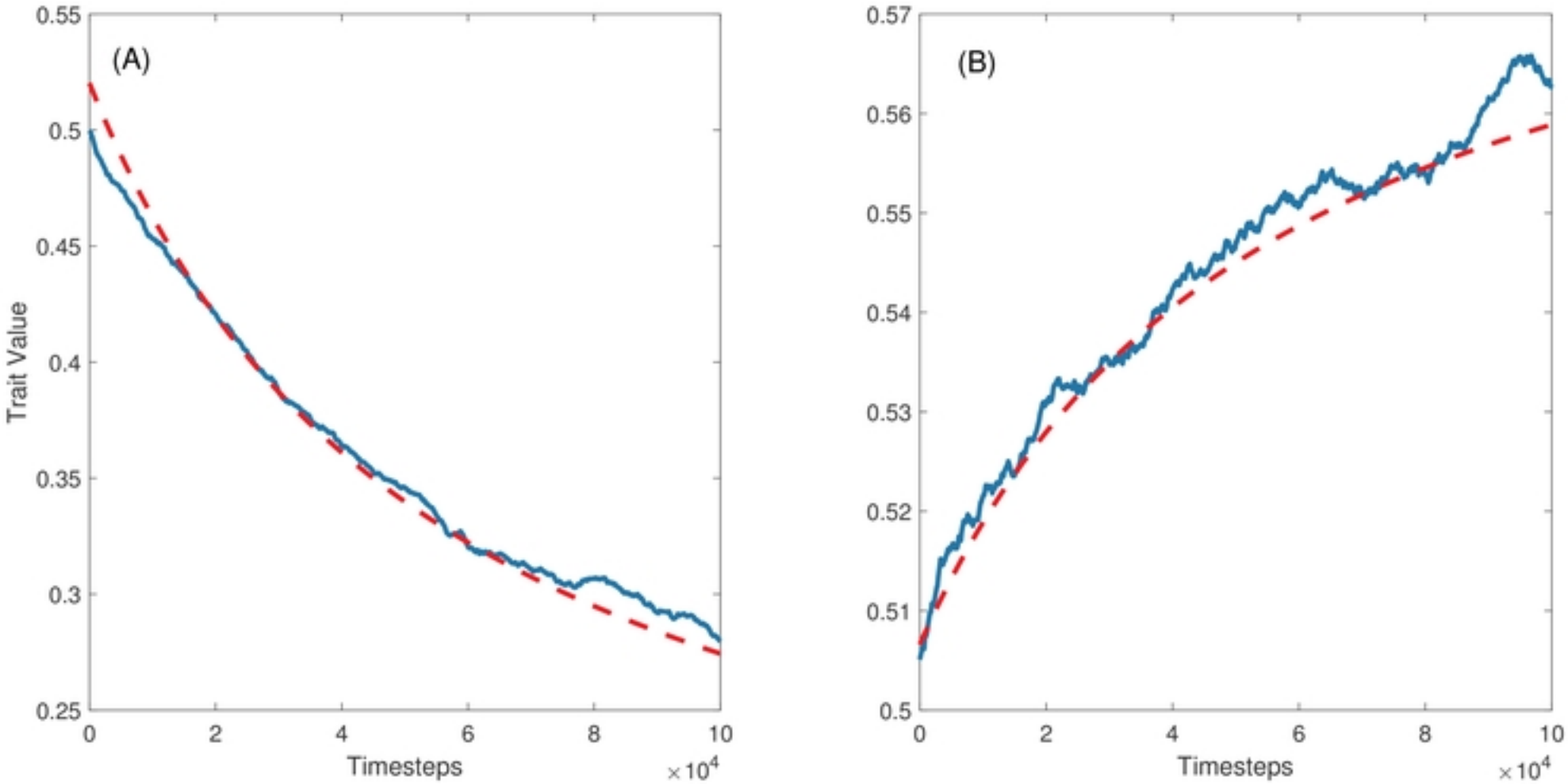
79. Schrödinger E. Quantisierung als eigenwertproblem. Annalen der physik. 1926;385(13):437–490.

80. Taussky O, Todd J. Another look at a matrix of Mark Kac. Linear Algebra and Its Applications. 1991;150(C):341–360.

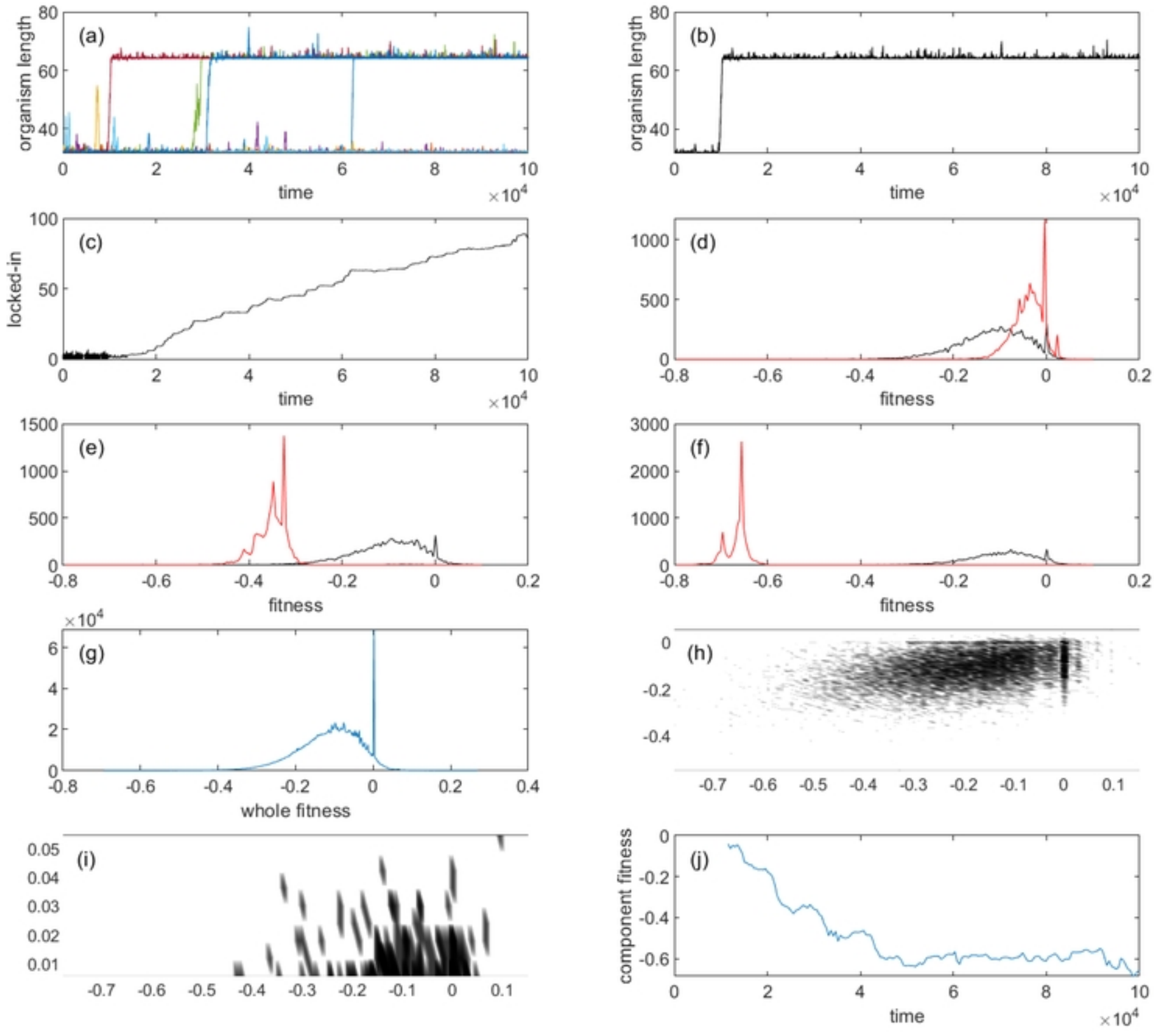
81. Munarini E, Torri D. Cayley continuants. Theoretical computer science. 2005;347(1-2):353–369.


(A) Realizability of trait value from multiple phenotypes
$u_m = \phi_m - \phi^*$
trait value: $\phi^*=0$
fitness: ω^*
0 0 0 0 0 0
mutations
$_m=1$
1 0 0 0 0 0
0 1 0 0 0 0
0 0 1 0 0 0
0 0 0 1 0 0
m1
m2
m3
m4
$s_m = {}_m - {}^*$
...
(B) Environmental bias
$s<0$
$s>0$
environmental bias towards $u=2$
$_s(s|u)$
$_s(s|u=1)$
$_s(s|u=2)$
$s=0$
fitness change (s)
(C) Mutation bias
all mutants
$_u(u)$
mutation bias towards smaller u
$u<0$
$u>0$
$u=0$
change in trait value (u)

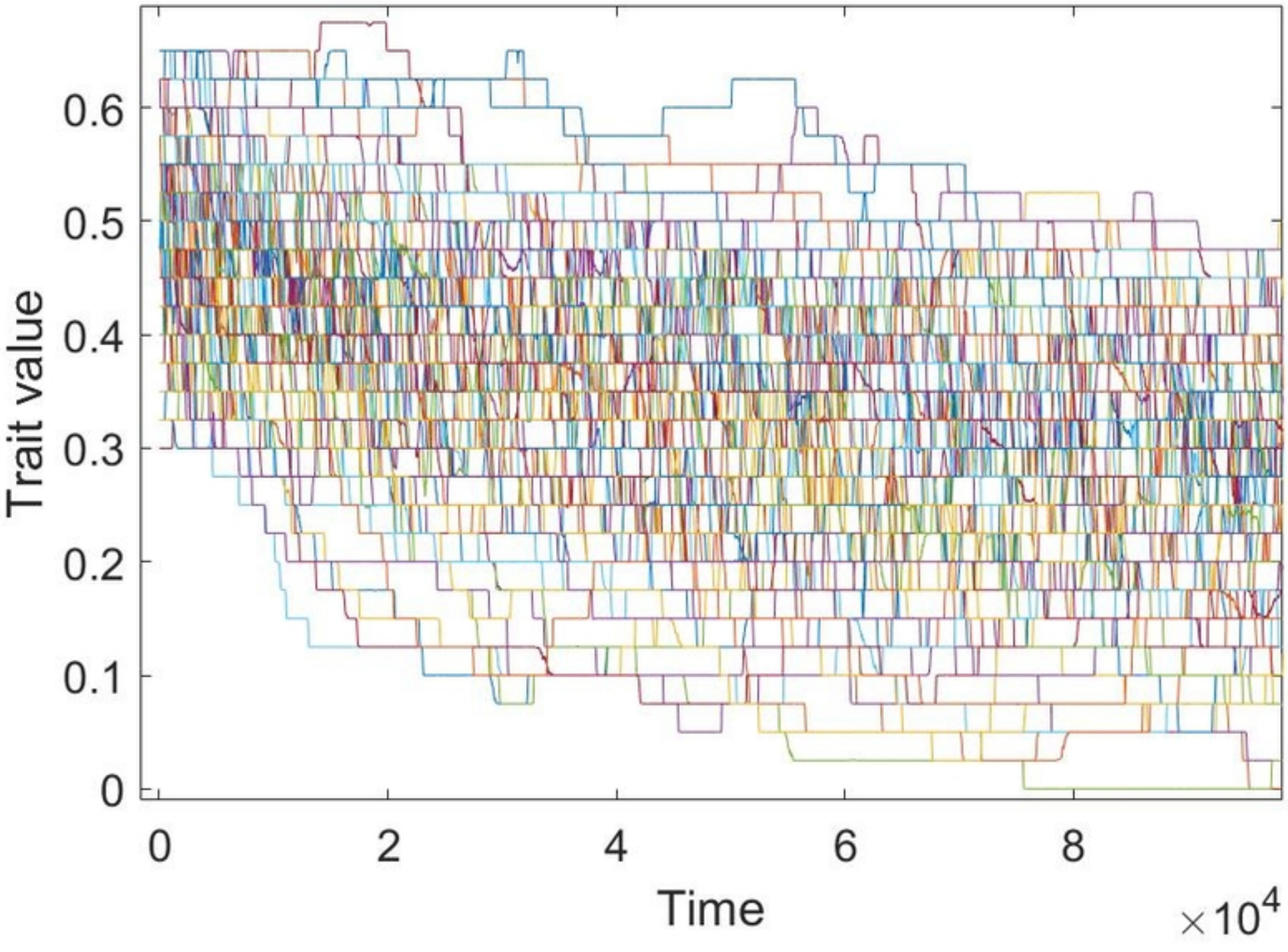
Figure


(A)
Trait Value
Timesteps
×10^4
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0
2
4
6
8
10
(B)
Timesteps
×10^4
0.57
0.56
0.55
0.54
0.53
0.52
0.51
0.5

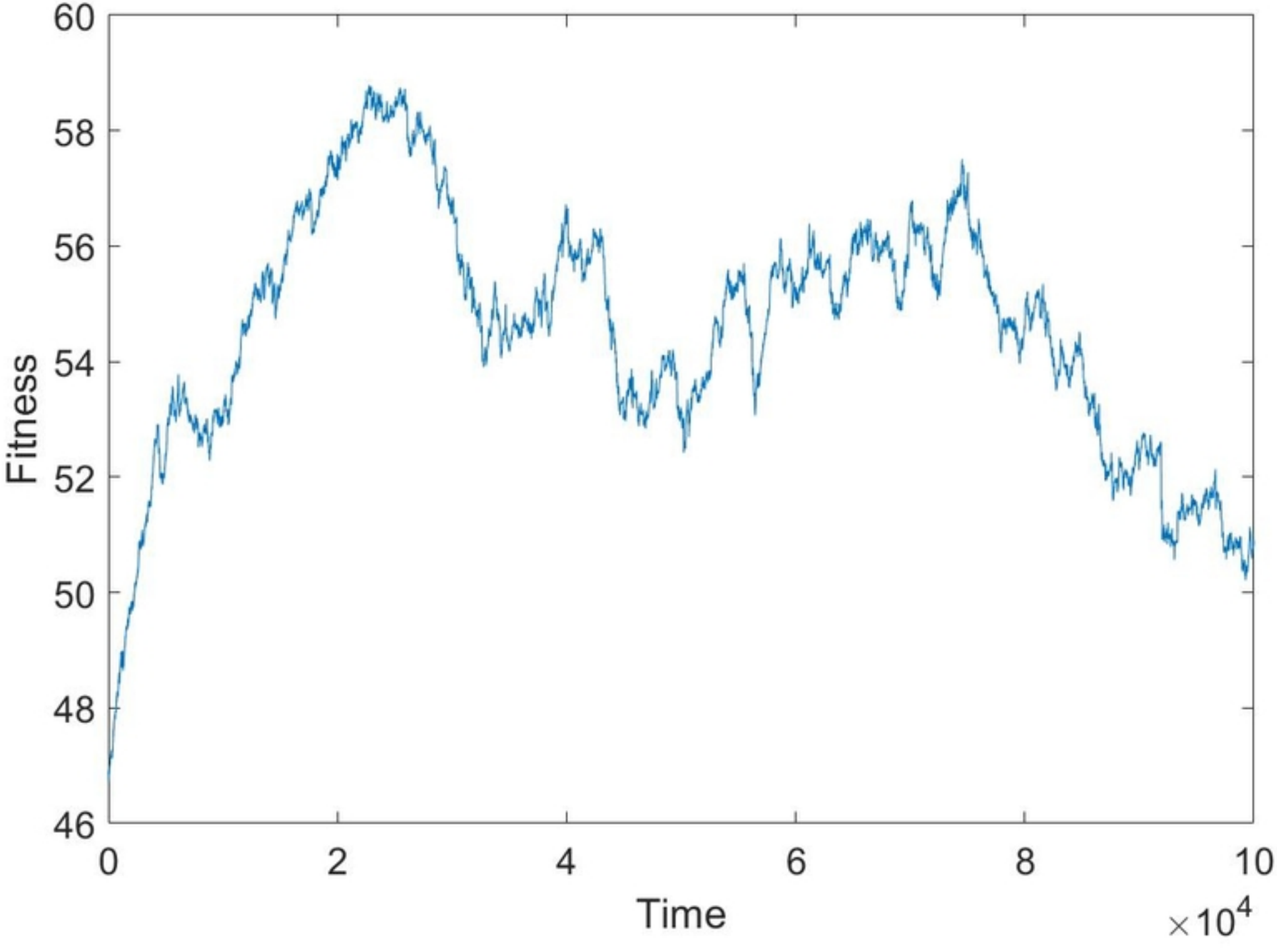
Figure


(a)
organism length
time
×10^4
(b)
organism length
time
×10^4
(c)
locked-in
time
×10^4
(d)
fitness
(e)
fitness
(f)
fitness
(g)
whole fitness
(h)
(i)
(j)
component fitness
time
×10^4

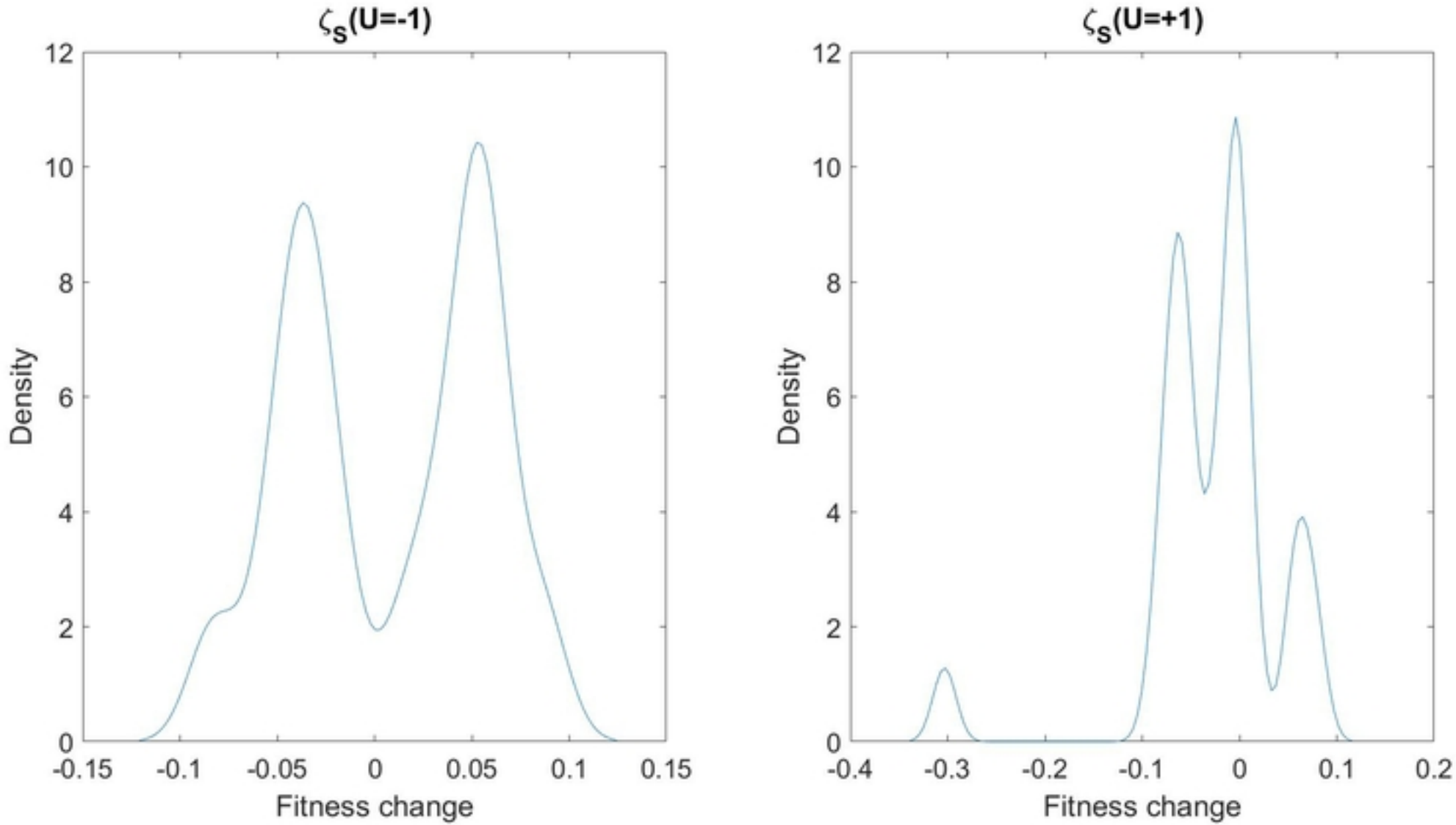
Figure



Figure



Figure


ζ_s(U=-1)
ζ_s(U=+1)
Density
Fitness change
12
10
8
6
4
2
0
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2

Figure