

Fitness and Overfitness: Implicit Regularization in Evolutionary Dynamics

Hagai Rappeport ^{*}

School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel

Mor Nitzan [†]

*Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem, Israel;
School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel;
and Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel*



(Received 31 July 2025; accepted 12 May 2026; published 17 June 2026)

A common underlying assumption in evolutionary thought is that adaptation generally drives an increase in biological complexity. However, the rules governing the evolution of complexity appear more nuanced, and the forces that drive the increase or decrease in organismal complexity are not fully understood. Evolution is deeply connected to learning, where complexity, as well as its origins and consequences, are much better understood, with established results on the optimal complexity appropriate for a given learning task in various settings. In this work, we suggest a mathematical framework for studying the relationship between organismal complexity and the complexity of the environment in which the organisms evolve by leveraging an existing mathematical isomorphism between evolutionary dynamics and learning theory, specifically an isomorphism between the replicator equation and sequential Bayesian learning, with evolving types corresponding to competing hypotheses, and fitness in a given environment corresponding to the likelihood of observed evidence. In Bayesian learning, implicit regularization prevents overfitting and drives the inference of hypotheses whose complexity matches the learning challenge. We show how these results naturally carry over to the evolutionary setting, where they are interpreted as organism complexity evolving to match the complexity of the environment, with organisms that are too complex or too simple for their environment suffering from effects we term *overfitness* and *underfitness*, respectively. Other aspects, peculiar to evolution and not to learning, reveal additional trends. One such trend is that frequently changing environments decrease selected complexity, a result with potential implications in both evolution and learning. Together, our results suggest that the balance between complex organisms that may overadapt to transient environmental features, and simple organisms that may be insufficiently flexible in responding to environmental challenges, drives the emergence of optimal complexity that reflects environmental structure. This mathematical framework offers new ways to think about biological complexity, and it suggests new potential causes for an increase or decrease in complexity in different selective environments.

DOI: [10.1103/zrqb-cj47](https://doi.org/10.1103/zrqb-cj47)

I. INTRODUCTION

Living systems, across scales and levels of organization, are complex. This complexity, manifested in the morphological, organizational, and behavioral aspects of organisms, is hailed by many as one of life's defining features [1]. But while our detailed understanding of various complex adaptations continually expands, a comprehensive theory of the nature and origins of biological complexity itself remains elusive [2]. As with virtually any other feature of living systems, biological complexity may make more sense in light of its evolution [3–5].

Early thinkers on evolution assumed that lineages tend to progressively increase in complexity as they adapt to environmental challenges [2,6,7], and much work has been dedicated to elucidating the evolutionary forces behind this supposed trend [8,9]. However, biological reality is more nuanced. Parasites, simpler in many aspects than their free-living relatives [10], are a standard example of environments favoring reduced complexity. In general, however, we do not fully understand the selective pressures affecting the increase or decrease of complexity and their interplay in various contexts and environments [2,11].

The existing literature on selection for decreased complexity has tended to focus on two points. The first is a supposed maintenance cost of complexity and the potential fitness advantages gained from forfeiting it and diverting resources elsewhere [12,13]. The second point relates to the entropic force generated on genomes by the constant barrage of mutations which accumulate unless purged by stabilizing selection. Presumably, any complex structure not under strong stabilizing selection is thus expected to erode over time [14–16]. While these two evolutionary forces are certainly

^{*}Contact author: Hagai.Rappeport@mail.huji.ac.il

[†]Contact author: Mor.Nitzan@mail.huji.ac.il

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

relevant, it remains unclear what their relative contribution is and whether they are the only ones at play [17]. Complexity may also theoretically hinder adaptation rates due to pleiotropy [18], but empirically this seems not to be the case [19] (see also [8]).

An additional hurdle to developing a theory of biological complexity is the lack of an agreed-upon definition of the subject matter. Multiple formalisms have been suggested, but none has met universal acceptance. Biological complexity can be defined as the amount of information a genome carries about its environment [20]. But under such a definition, complexity necessarily increases until genomes encode a maximal amount of information on their environment, thus not capturing the phenomenon of potentially decreasing complexity in simple environments. One path to a stand-alone definition of organismal complexity has been to transfer to a biological context the idea, well understood in theoretical computer science, of complexity as a measure of a system's minimal description [21–24]. Other definitions appeal to a notion of complexity as related to the “number of parts” that an organism is composed of [25], be they genes [26–28], cell types [9], developmental pathways [29], or morphological traits [30].

The drawback of all the aforementioned definitions is that they are external to evolutionary theory, and consequently their causal interaction with selection is unclear. In particular, they do not allow for making predictions of evolutionary trends in complexity depending on different types of environments. To bridge this gap, a notion of complexity that is directly emergent from evolutionary dynamics is required.

The “evolutionary dynamics” in question can be captured by many different models, but arguably the simplest [31] is the discrete-time replicator equation:

$$x_i^{(t+1)} \propto x_i^{(t)} f_i^{(t)}, \quad (1)$$

with $x_i^{(t)}$ representing the frequency of type i at generation t , and $f_i^{(t)}$ denoting the fitness of type i given the environment experienced at that generation. The form of this dynamical equation is analogous (and indeed mathematically equivalent) to a different system where types change in frequency over time in response to external constraints favoring some over others. Namely, in the context of Bayesian inference, we write

$$p(H_i | E) \propto p(H_i)p(E | H_i), \quad (2)$$

where $p(H_i | E)$ is the posterior of hypothesis i given its prior $p(H_i)$ and the likelihood $p(E | H_i)$ that H_i assigned to the observed evidence E . Sequential inference is then described by the updated equation

$$p^{(t+1)}(H_i) \propto p^{(t)}(H_i)p(E^{(t)} | H_i), \quad (3)$$

where $p^{(t)}(H_i)$ represents the probability associated with hypothesis i at time step t , and $p(E^{(t)} | H_i)$ is the likelihood under hypothesis i of the evidence observed at that time step.

Equations (1) and (3) describe the same exact dynamics, only with a different interpretation of the quantities being updated [32]. However, facts about the dynamics which may be obvious in one context may be completely obscure in the other, and vice versa. In particular, Bayesian

learning has a well-developed theory of optimal complexity for a given learning task. Over the past decade, a rich literature has emerged on connections between evolution and learning theory, porting insights from the latter to better understand the former [33–37]. Here we will attempt to expand this intellectual enterprise to the study of the evolution of complexity.

Learning theory has multiple operational notions of complexity [22,38–40], relating a model's (hypothesis') complexity to the number of degrees of freedom it possesses. A common practice in computational learning is to penalize, or *regularize*, models to prevent them from becoming too complex, as models that are complex relative to the underlying task are likely to *overfit*. That is, they may fit noise in the training data rather than the underlying patterns [41]. In analogy to evolution, such penalties correspond to direct fitness costs of maintaining complex structures, as discussed above. In addition to explicit regularization, many forms of learning also possess implicit forms of regularization emergent from their learning dynamics [41,42]. The question naturally arises whether there exist analogs of implicit regularization in evolutionary dynamics, and if so, what is their significance to the evolution of complexity?

II. MODEL

Denoting type i 's phenotype at generation t as $y_i(t)$, we assume that fitness decreases monotonically with distance to an optimal phenotype $y^*(t)$, which is therefore the unique fitness peak for that generation. Our goal is to follow how selective pressures vary for organisms of different complexity depending on the complexity of their environment. To this end, we define organismal and environmental complexity as the number of degrees of freedom in the processes generating organism phenotypes $y_i(t)$ and optimal phenotypes $y^*(t)$, respectively. Measuring model complexity via the number of free parameters required to specify it is standard in learning theory [38,39] and closely corresponds to algorithmic complexity [22,43]. We defer further treatment of this measure and its place in the literature on biological complexity to the Discussion section.

The most straightforward way to model environments and genotypes in this setting is as generative distributions over phenotype space. However, phenotype distributions provide only a statistical summary of outputs without relating to the underlying processes that generate these responses. To better capture the mechanistic relationship between sensing, processing, and phenotypic expression, a more natural setting is populations of *functions*. This aligns naturally with computational learning, where one typically evaluates competing functions on their ability to map inputs to outputs rather than just examining their output distributions.

Organisms infer features of the environment they find themselves in by various sensing apparatus, and they adjust their phenotypes accordingly. Thus, we consider genotypes as coding for input-output maps ϕ_i [Fig. 1(a)], where inputs are environmental cues (either measurements of external conditions such as temperature, nutrient availability, etc., or signals conveyed by other organisms) and outputs are phenotypes (any variant morphologies or behaviors that may affect

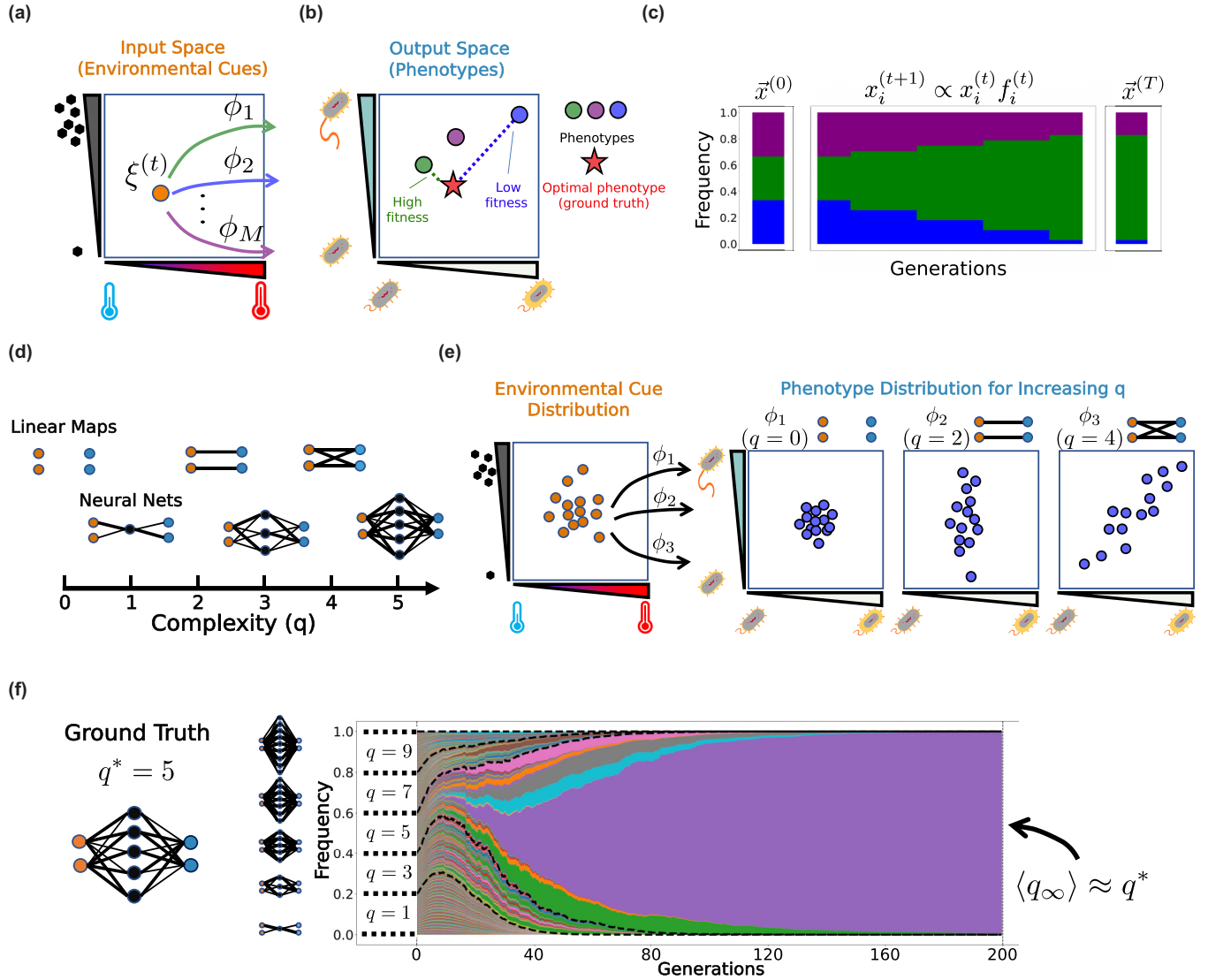


FIG. 1. A model for selection of complexity in environments of a given complexity. (a) Populations consist of functions $\{\phi_i\}$ mapping environmental cues $\xi^{(t)}$ (e.g., temperature or nutrient availability) received at each timepoint t to phenotypes (e.g., membrane thickness or flagellum size). (b) Fitness in a given environment is defined via a ground truth function ϕ^* (optimal by definition), and each member's fitness f_i is an increasing function of the distance between its phenotype to the optimal phenotype $\phi^*(\xi^{(t)})$. (c) Selection follows the discrete time replicator equation, and we follow relative frequencies of each member. Illustrated is a Muller plot following the dynamics of a population composed of three members starting from a uniform distribution $\bar{x}^{(0)}$ and evolving into a population with distribution $\bar{x}^{(T)}$. (d) Complexity is defined as the number of tunable parameters in a function, here illustrated for two function classes—linear mapping (top) or one-hidden-layer neural networks (bottom). (e) Different members ϕ_i may have different associated complexities q_i , resulting in different levels of intricacy for their mappings from environment to phenotype space. The environmental complexity q^* is defined as the complexity of ϕ^* . (f) Muller plot for a typical simulation of the replicator dynamics, following a population of functions divided into five complexity classes (dashed lines) and starting from a uniform distribution. The simulation culminates with the dominance of a member from the $q=5$ class, which is equal in this case to q^* .

fitness). The optimal phenotype in a given environment is defined via a “ground truth” map ϕ^* , which encodes the underlying environmental regularities, i.e., the optimal response for any possible environmental cue [Fig. 1(b)], and fitness decreases monotonically with distance to the optimal phenotype.

In this setting, we follow pure selection dynamics (with no mutation, thereby isolating the effect of selection on existing variability) according to the replicator equation over populations of parametrized functions $\{\phi_i\}_{i=1}^M$, with

parameters representing genomes [Fig. 1(c)]. The population is divided into *complexity classes* $\{C_q\}_{q \in \mathcal{Q}}$, where each class is defined by the number of parameters required by its functions [Figs. 1(d) and 1(e)]. More parameters allow for capturing more intricate patterns in mapping environmental cues into phenotypes [Fig. 1(e)], such as in, e.g., the degree of intricacy required for advanced combinatorial sensing [44]. The *environmental complexity* is then defined as the complexity of ϕ^* , matching the intuition that the complexity of an environment should match the intricacy of response required to thrive in it.

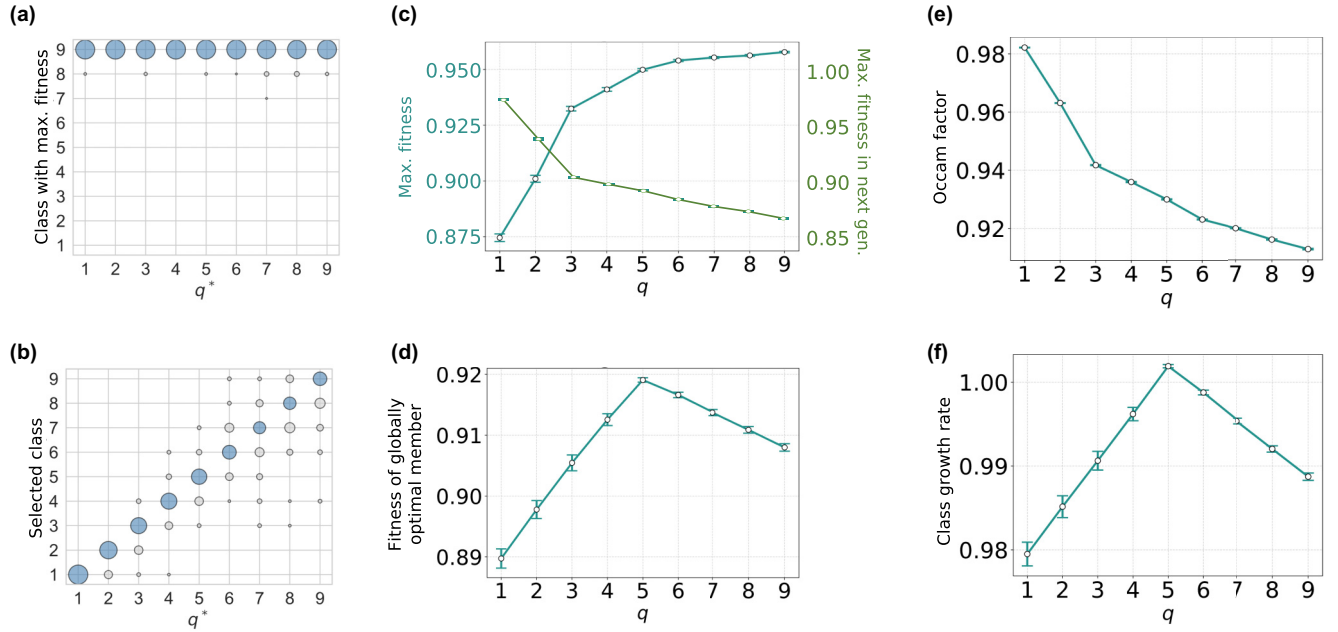


FIG. 2. Selected complexity depends both on a class' maximal per-time-step fitness as well as on fitness stability over time. (a) The maximal fitness is typically found in the highest complexity class, independent of q^* . Circle sizes represent, for a given q^* (x-axis), the fraction out of 100 simulations in which time-averaged max fitness was found in a given complexity class (y-axis), with the highest fraction marked in blue. (b) Selected complexity (y-axis), defined as the average population complexity at the simulation's end. The selected class tends to correspond to the environmental complexity (q^* ; x-axis) in which selection occurred. Circle sizes and colors as in (a). (c)–(f) Illustration for a specific q^* ($q^* = 5$) of time-averaged metrics. (c) While maximal fitness is strictly increasing with the complexity of the class (y-axis, left), high fitness at a given time step does not necessarily predict high fitness at the next, illustrated by plotting fitness at time $t + 1$ of the member that achieved best fitness at time t (divided by its fitness at time t) (y-axis, right). (d) The best globally optimal member (with highest time-averaged fitness) is found in q^* . (e) The Occam factor (y-axis) decreases with complexity (x-axis). (f) Average class growth rate (which is proportional to the maximal fitness times the Occam factor) is peaked at q^* , which thus eventually dominates. Results in all panels are based on 100 simulation realizations with linear maps. Population size was 2000 per complexity class, $\sigma = 1$, $\gamma = 0.05$, and $T = 1000$. In (c)–(f) circles represent mean values of metrics time-averaged over simulated trajectories, and error bars show 95% confidence intervals (1.96 standard errors) over these means.

In the interest of focusing on selection, this setting does not consider *generating* novel complex phenotypes [18], an issue we leave for future work.

At each generation t , all functions $\{\phi_i\}_{i=1}^M$ receive a randomly sampled environmental cue $\tilde{\xi}^{(t)}$. In the setting considered throughout the main text, cues are sampled independently across generations from an isotropic standard normal distribution, $\tilde{\xi}^{(t)} \sim \mathcal{N}(0, I_n)$. The fitness of a given member ϕ_i is given by a Gaussian selection term $f_i^{(t)} = e^{-\gamma \|\phi_i(\tilde{\xi}^{(t)}) - \phi^*(\tilde{\xi}^{(t)}) + \epsilon^{(t)}\|_2^2}$ [45]. The prefactor γ determines selection strength; smaller values of γ lead to smaller fitness differences and hence smaller population changes per time step. When generations are measured in units of $\frac{1}{\gamma}$, the dynamics become independent of γ , in the limit $\gamma \rightarrow 0$ converging to the continuous-time stochastic replicator equation (methods). Lastly, $\epsilon^{(t)} \sim \mathcal{N}(0, \sigma^2)$ is a noise term, modeling features relevant for fitness that are not encoded by the provided environmental cues.

III. RESULTS

For the sake of simplicity, and to best illustrate the underlying principles, we start by considering $\{\phi_i\}_{i=1}^M$ and ϕ^*

to be linear mappings. We note, however, that all results in this section hold qualitatively for other types of mappings, such as polynomials or feed-forward neural networks; see Appendix D for further details. In addition, in what follows we consider the setting in which all complexity classes start with the same representation in the population, as the effects in play are clearer in this setting. See Appendix G for a demonstration of the insensitivity of the main findings to this assumption.

A. Implicit regularization in the replicator equation prevents overfitness

In all function classes considered, at any given time step the maximal fitness achieved is found in the class of highest complexity $q_{\max}$ [Fig. 2(a)]. (See Appendix F for a discussion on the effect of population size on maximal fitness.) Nevertheless, the complexity class which eventually comes to dominate the distribution (hereafter the *selected complexity*, denoted $\langle q_{\infty} \rangle$), is typically not $q_{\max}$, but rather corresponds to environmental complexity, i.e., $\langle q_{\infty} \rangle \approx q^*$ [Fig. 2(b)].

To understand this apparent discrepancy, we find it illustrative to start by considering, for each complexity class C_q , the member of C_q with maximal fitness at time step t , call it $i_q^*(t)$,

and follow its fate at the next time step $t + 1$ (for the following discussion, we focus on a fixed, intermediate, environmental complexity $q^* = 5$). While the fitness of $i_q^*(t)$ indeed increases with complexity, its fitness at $t + 1$ (normalized to its fitness in time step t) decreases with complexity [Fig. 2(c)]. When considering the fitness of each class's globally best performing member (i.e., the one with the highest mean log fitness over the whole simulation), we indeed find the best performing member in q^* [Fig. 2(d)]. As a whole, this analysis suggests that the apparent increased success of complex classes is only due to their ability to better fit specific environmental cues or noise realizations rather than the underlying environmental challenge.

To formalize and better understand this effect, we next turn to a mathematical construct that is closely related to the aforementioned ratio of fitness values, known in learning theory as the *Occam factor* [43]. In terms of the replicator equation [Eq. (1)], for a setting such as ours, dividing our population into N subpopulations (classes) $\{C_q\}_{q \in \mathcal{Q}}$, we can write a replicator equation for the frequencies of classes in the form

$$X_q^{(t+1)} \propto X_q^{(t)} F_q^{(t)}, \quad (4)$$

where $X_q^{(t)}$ is the frequency of class q at generation t (relative to other classes), and F_q serves as class fitness, i.e., the growth rate of class C_q once normalized to other classes' fitness. Following the notation introduced above, the frequency of the highest-fitness member of class C_q at generation t is $x_{i_q^*(t)}^{(t)}$, and

we denote $\tilde{x}_{i_q^*(t)}^{(t)} = \frac{x_{i_q^*(t)}^{(t)}}{X_q^{(t)}}$, i.e., $\tilde{x}_{i_q^*(t)}^{(t)}$ is the within-class relative frequency. We can now write (see the derivation in the Methods section)

$$F_q^{(t)} = f_{i_q^*(t)}^{(t)} \frac{\tilde{x}_{i_q^*(t)}^{(t)}}{\tilde{x}_{i_q^*(t)}^{(t+1)}}. \quad (5)$$

That is, the fitness of class q at time t is a product of the fitness of its best member at that time step, $f_{i_q^*(t)}^{(t)}$, and a factor of magnitude less than 1, the *Occam factor*, which is the reciprocal of the change in frequency in favor of the current best member $i_q^*(t)$. A smaller Occam factor corresponds to a larger “collapse” of a class's internal distribution in favor of its current best member, and is associated with decreased class growth at that time step.

The fitness of members in complex classes tends to change more substantially over time compared to simple classes [Fig. 2(c)]. Thus complex classes collapse on new “optimal” members, every generation anew, and overall have lower average Occam factors [Fig. 2(e)]. Consequently, they possess suboptimal class growth rates [Fig. 2(f)], explaining the fact that increased complexity beyond environmental complexity does not lead to increased selection. In analogy to the term used in learning theory, we term this drawback of complexity “overfitness.”

To summarize this section, we find that even without imposing any external cost on complexity, the dynamics of the replicator equation are sufficient to select, over time, complexity that matches the underlying environmental complexity. This occurs via two opposing forces. On the one hand,

overfit classes that are too complex relative to environmental complexity favor a different member for each time step and therefore do not rise to dominance. On the other hand, classes that are too simple are unable to match nuances in the optimal mapping (*underfit* for the environmental challenge) and are therefore disfavored.

B. Simple classes are transiently favored in the initial phase of selection

Thus far we have considered only the long-term success of different complexity classes. In an evolutionary setting, short-term selection dynamics may prove relevant as well. In a typical simulation realization, before the $q = q^*$ complexity class eventually comes to dominate, we observe a transient period where simple complexity classes are more successful [Fig. 3(a)]. A first hint towards explaining this transient is that we may write a class's fitness [Eq. (4)] as the average fitness of its members, weighted by their current representation in the class:

$$F_q^{(t)} = \frac{\sum_{i \in C_q} x_i^{(t)} f_i^{(t)}}{X_q^{(t)}}. \quad (6)$$

Our simulations begin from a uniform composition for all populations, and therefore at $t = 0$ we have $F_q^{(0)} \propto \langle p_q \rangle$, i.e., the fitness of class q is equal to the mean of the fitness distribution p_q from which it drew its values, and this mean value is highest for simple classes [Fig. 3(b)]. Thus, at $t = 0$ the simplest classes exhibit the largest growth rate [Fig. 3(f)]. Of note, this effect reflects transient dynamics, and the eventual steady-state complexity still typically converges to q^* .

At $t > 0$ class fitness begins to change due to selection, and via a close analog of Fisher's fundamental theorem of natural selection [46] (see Methods for a derivation),

$$\Delta F_q(0) = \frac{\text{Var}[p_q]}{\langle p_q \rangle}. \quad (7)$$

Since complex classes have higher values for $\text{Var}[p_q]$ and lower values for $\langle p_q \rangle$ [see also Fig. 3(b)], their rate of class fitness increase outpaces simple classes and starts overtaking them.

In the long term, a class C_q 's fitness equals the fitness of its optimal member i_q^* (see Methods). Since $f_{i_q^*}$ is highest for the $q = q^*$ class [Fig. 2(d)], after sufficient time this class dominates the overall population composition, as demonstrated above.

A different way of understanding this effect is by following the dynamics of the Occam factor. A class's growth rate is a product of the fitness of its best member and its Occam factor. While maximal fitness does not change with time [Fig. 3(d)], the Occam factor increases for all classes as their composition becomes concentrated around their respective best members [Fig. 3(e)], at which point the growth rate of all classes reaches its steady-state value [Fig. 3(e)]. Thus, the transient in question represents the initial process of locating optimal members for the given environment, where simple classes possess an advantage.

To summarize, selection starts with a transient phase where complex classes have yet to promote optimal

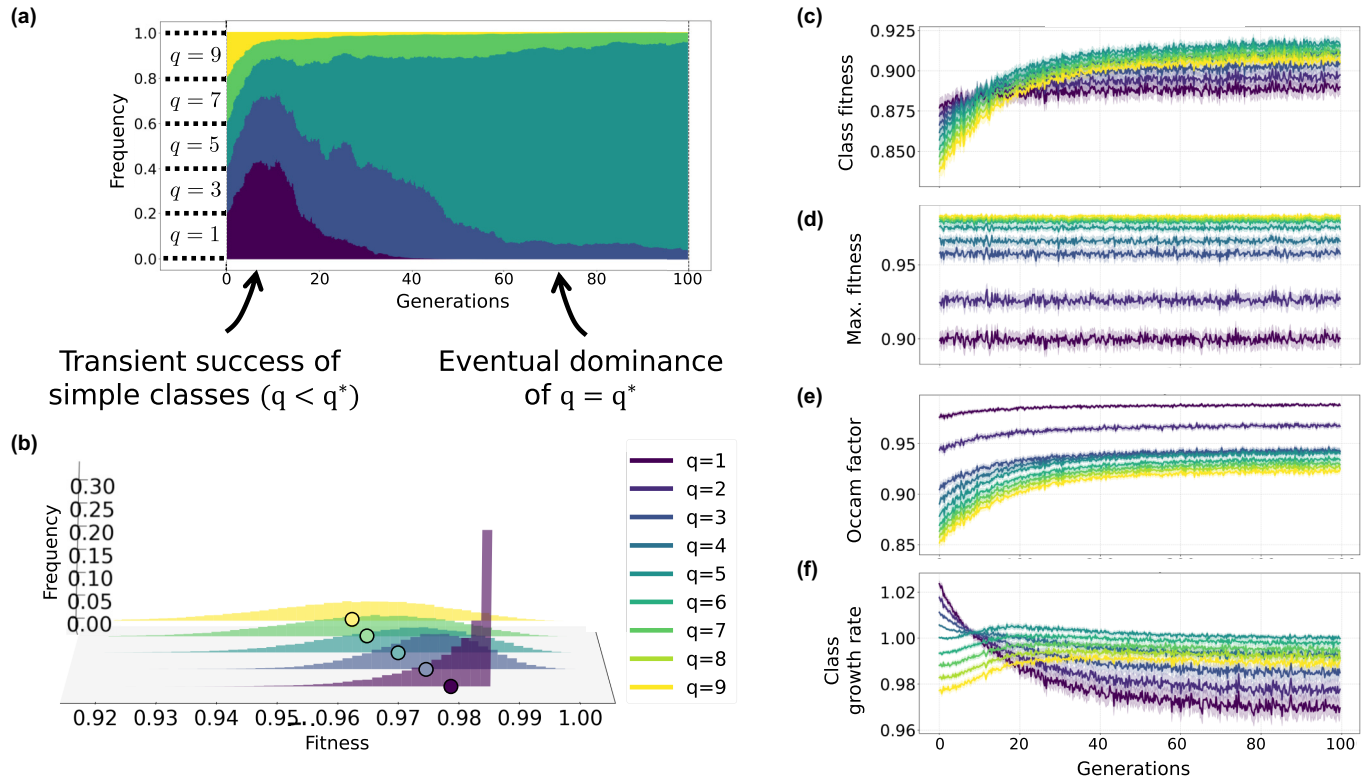


FIG. 3. Simple classes are transiently favored in the initial phase of selection. (a) Muller plot of class frequencies in a typical simulation realization. While the complexity of the member which eventually dominates selection is q^* (here $q^* = 5$), in the initial phase of selection there is a transient success (increase in frequency) of classes with $q < q^*$. (b) Fitness distribution of different classes with respect to a randomly chosen ground truth function. More complex classes have a wider fitness distribution with a higher maximal fitness value, but a lower mean (colored circles). (c)–(f) Dynamics (averaged over realizations) of the metrics shown in Fig. 2. (c) Class fitness changes over time, initially highest for simple classes before converging to favor the class with complexity equal to q^* . (d)–(f) Class fitness can be decomposed into the product of maximal class fitness (d) and its Occam factor (e), and normalized to give a class’s growth rate (f). Results in (b)–(f) are based on 100 simulation realizations with linear maps. Population sizes of 2000 per complexity class, $\sigma = 1$, $\gamma = 0.05$, and $T = 500$.

members in their inner composition and consequentially simple classes are selected at their expense. This effect is similar to the one discussed in [47,48], where the mean phenotype of a population in a changing environment lags behind the optimal phenotype. Using the language and tools of learning theory allows us to analyze this lag as a function of organism complexity, showing that lag increases with class complexity and hence that populations of simple organisms benefit in such conditions. A more general treatment of this effect is given in Appendix C, where we provide a general geometric argument for when it should arise, an analytical proof for the linear case, and numerical demonstrations in multipeak and nonlinear landscapes.

C. Rapidly changing environments select for lower complexity

As detailed above, the transient success of simple complexity classes reflects a period in which optimal members of each class in the given environment do not yet dominate their respective classes. We next asked if this transient would repeat if the environment, i.e., the ground truth function, changed midselection. So far, our model captured a certain degree of environmental change at each time step, via the environmental cue that follows changes in environmental

parameters. However, the underlying regularities in optimal response to these parameters, i.e., the ground truth function, was hitherto assumed constant. Changing the ground truth thus models a drastic environmental change or a local shift to a different environmental niche, one that could change the fitness landscape for a given set of environmental parameters. In our model, fitness is uncorrelated in two randomly chosen environments, so when the environment changes, each class effectively “starts over” in discovering its new best member (but see Appendix E for an analysis of temporal correlations between successive environments, as well as cue correlations in a fixed environment). Indeed, changing the environment in such a manner may lead to a second period of success for simple classes [Fig. 4(a)]. However, this success is transient as well, eventually dying out and giving way to a member of the complexity class that matches the environmental complexity.

Given this effect, the natural next question is what would happen if environments kept changing repeatedly, faster than the rate at which the transient dies out. Under these conditions, the success of simple classes is no longer transient, and the system may reach an equilibrium dominated by classes with lower complexity than environmental complexity [Fig. 4(b)]. In this regime, repeated environmental change

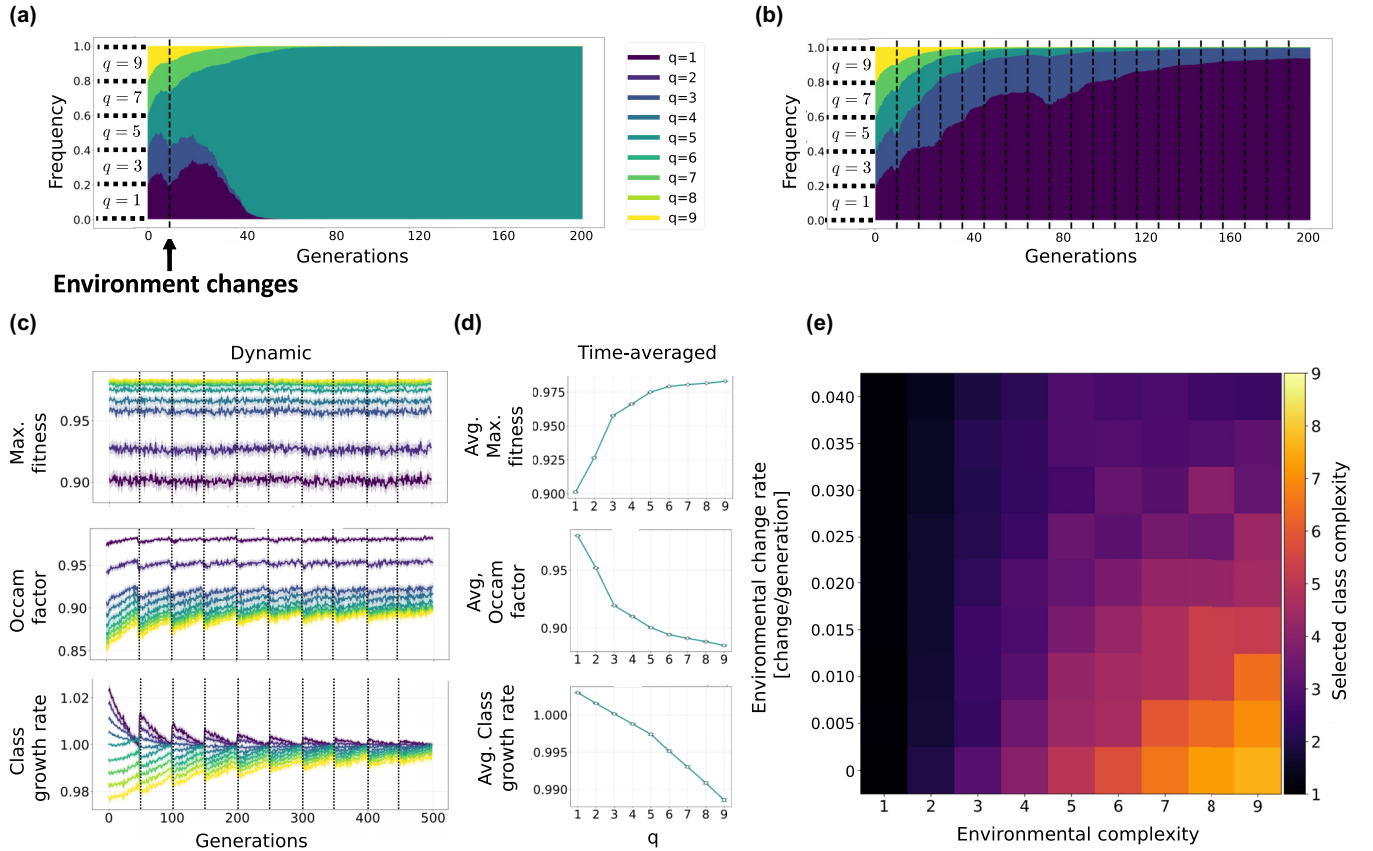


FIG. 4. Environmental change selects for reduced complexity. (a) Muller plot of class frequencies (as in Fig. 3), with the environment (the ground truth function ϕ^*) changed to a different randomly sampled one at $t = 50$. When the ground truth function changes (vertical dashed line), the transient success of simple classes repeats for a second time before subsiding. (b) When the environment changes at the same timescale as that of the transient, the transient becomes fixed, and simple classes come to dominate selection. (c) Dynamics of the same metrics as in Figs. 3(d)–3(f), only with environments changing every 50 time steps. (d) Time-average values of the metrics in (c). (e) Selected class (average class at the end of selection) as a function of both environmental complexity (x-axis) and environmental change rate (here, the number of different environments in a simulation of length $T = 1000$) (y-axis). Results in (c)–(e) are based on 100 simulation realizations with linear maps. Population sizes of 2000 per complexity class, $\sigma = 1$, $\gamma = 0.05$, and $T = 500$.

prevents the system from reaching the steady state observed in fixed environments, effectively converting transient dynamics into long-term outcomes.

Maximal fitness per time step monotonically increases with class complexity, and is unaffected when the environment changes, provided the new environment is of the same complexity as the previous one [Fig. 4(c), top]. However, a new environment implies new optimal members, and consequentially a drop in the Occam factor [Fig. 4(c), middle] until said optimal members rise in frequency. Since this effect is stronger in complex classes, when the environment changes sufficiently fast (relative to the timescale of the transient), the Occam factor of the complex classes never surpasses that of the simple classes, and the latter's growth rate remains highest [Fig. 4(c), bottom].

Quantifying this effect globally, by measuring the selected class for each value of q^* and for a range of environmental change rates, we find for all values of q^* a trend of decreased selected complexity with increasing environmental change rate [Fig. 4(e)].

To summarize, we find a consistent pattern of a transient “lag” period, where simple classes succeed at the expense

of more complex ones, even if the latter eventually come to dominate. We may interpret this lag as the time it takes complex classes to locate the correct hypothesis (or the optimal member, as it were) for a newly presented environmental challenge. When environments change at a rate comparable to this lag period, complex classes may always contain the optimal member per environment, but by the time they are able to promote this member in frequency and start growth, the environment changes again. Simple classes, on the other hand, maintain internal variability and are much less affected when environments change. Thus, the internal complexity of the environment, i.e., its structural richness, and the rate at which this structure changes, are both relevant to selection of complexity but have opposite effects.

IV. DISCUSSION

Recently, multiple lines of work in theoretical evolution have made use of mathematical isomorphisms which formalize the analogy between evolution and learning [33–35,37], allowing one to transfer theorems and results from one field into the other. Given that in learning theory,

complexity is well-defined and many of its effects understood, these isomorphisms present a hitherto untapped opportunity to transfer insights on complexity into evolutionary theory. A hypothesis' complexity in learning is defined via the number of tunable parameters it requires. We have adopted this definition, interpreting it as the number of tunable parameters in the environment-phenotype map. This definition instantiates a common theme considered in the literature on biological complexity, heterogeneity of possible states, and responses to external stimuli [25].

Using this mathematical framework, we showed how evolutionary dynamics in their simplest form contains implicit mechanisms that prevent both *overfitness* and *underfitness*. Namely, organisms with too many tunable parameters will tend to tune these parameters, via evolution, to match patterns in their current environment that are no more than temporary noise, not indicative of future challenges. Simpler organisms are more robust to this effect, but may be too simple for a given environment, unable to utilize environmental information that is useful as well as enduring. In fixed environments, this balance determines the long-term steady-state outcome of selection. The dynamics leading to this steady state, however, involve a transient adaptation phase in which simpler organisms may be temporarily favored. When environments change on timescales comparable to this transient, these short-term dynamics can influence long-term population composition. The optimal organismal complexity is found in balancing these two selective forces, and is dictated by the environmental complexity. The robustness of these qualitative findings to alternative fitness functions, noise distributions, and parameter choices is examined systematically in Appendix I.

Implicit regularization of complexity is well understood in computational learning [42,43]. However, the transient effects in the initial phase of selection we found are, to the best of our knowledge, not studied within this literature. Indeed, such transients bear little significance in learning of a single given task with sufficient training time (which is still the most common setting in computational learning, but perhaps not for long with the rise of online or continual learning challenges). In an evolutionary setting, the “learning task” (e.g., the correct behavior in a given ecological niche) often changes on a timescale commensurable with the selection timescale. It is therefore natural in this setting to investigate the effect of changing environments [47–49], which we find to be substantial, and to interact in nontrivial ways with population complexity.

The notion of “environmental complexity” as well as “environmental change rate” are both considered important in evolutionary dynamics in general and the evolution of complexity in particular, but they are mostly considered closely related (e.g., [50,51]; see [20] for a notable exception). However, a stable but diverse ecosystem with many potential resources may be environmentally complex but slowly changing, whereas a seasonally fluctuating environment with a small number of relevant factors per season may be simple yet rapidly changing. In evolutionary game-theoretic terms, a high-dimensional payoff structure may be static, while even low-dimensional games can change rapidly. These two axes are, in principle, empirically separable: environmental

complexity may correspond to the effective dimensionality of ecological challenges shaping fitness (e.g., the number of independent resources, stressors, or interacting species), while the environmental change rate corresponds to the timescale over which these selective forces turn over (e.g., seasonal variability or disturbance frequency). Our results suggest that they should be considered separately, as their effects may act in opposite directions.

Throughout this work, complexity was defined as the number of effective degrees of freedom in the environment-phenotype mapping implemented by an organism. This notion corresponds to *process* or *generative* complexity, akin to Kolmogorov complexity, rather than to emergent structural or developmental complexity. These latter types of complexity often arise in biological systems from an underlying modular structure, which in turn is governed by relatively few control parameters. From this perspective, hierarchy, feedback, and modularity are not additional degrees of freedom so much as organizing principles that constrain how a relatively small number of degrees of freedom are deployed. However, modularity is not necessarily always beneficial, and the fact that it has repeatedly evolved is indicative of the structure of the environment in which organisms evolve [52]. In the language of the learning-evolution analogy, biological “design principles” such as modularity and feedback may be said to act as *inductive biases* that shape dynamics, similarly to architecture choices in machine learning models that are chosen to reflect structure in the learning task. Our framework, therefore, addresses how evolution selects for the *expressive capacity* of organismal responses relative to environmental structure, and does not attempt to capture the full emergence of biological organization.

There are multiple factors that are relevant to the evolution of complexity but were not discussed here. We have focused on the selective value of complexity rather than its origins, and we have not addressed the emergence of novelty or open-ended evolutionary innovation. In the main text, therefore, we assumed that relevant variability already preexists in the population, and we did not model mutation or recombination explicitly (but see Appendixes G and H for invasion and mutation scenarios). In addition, our framework does not model frequency-dependent selection. This last point may prove especially important as it is a prerequisite for arms-races, often speculated as drivers of increased complexity [17,53]. We leave the incorporation of these effects for future work.

In our model, organisms could sense environmental parameters but not which environment, containing which regularities, they are currently in. In the presence of signals carrying such information, organisms that are able to sense and utilize them may be favored [54]. For example, a short-lived creature such as a fruit fly (*Drosophila melanogaster*) may have genetic variants that are differentially adapted to different seasons, and consequentially populations may experience annual selective sweeps favoring different genomes at different seasons [55]. In contrast, more complex organisms sense seasonal change and adapt accordingly. A natural extension of our framework would be to incorporate such effects.

We believe the framework introduced here allows for tackling some aspects of complexity evolution in a more

first-principles manner than was previously possible. In particular, it provides a framework with natural definitions of both organismal and environmental complexity, definitions that may be instantiated in different ways in different models, as well as translated into observables to be followed in natural and experimental systems.

ACKNOWLEDGMENTS

We wish to thank Nathalie Balaban and Roy Friedman for insightful discussions. M.N. is supported by the EMBO Young Investigator Programme

DATA AVAILABILITY

The data that support the findings of this article are openly available [56].

APPENDIX A: SIMULATION DETAILS

1. Function classes

All function classes considered consist of mappings from environmental cues $\vec{\xi}$ from an input space of dimension n to phenotypes $\phi(\vec{\xi})$ from an output space of dimension p . Parameters, for all functions described, are drawn for each population member independently from $\mathcal{N}(0, 1)$. Results in the main text refer to populations of linear maps. Simulations described in Figs. 7 and 8 were carried out over populations of size 2000 per complexity class for $T = 1000$ time steps. Simulations described in Fig. 11 were carried out over different class sizes (x -axis) and $T = 5000$ time steps. All simulations were performed with $\gamma = 0.05$ and $\sigma = 1$.

a. Linear maps

These are functions of the form $\phi : \vec{\xi} \mapsto A\vec{\xi}$, where A is an $n \times p$ matrix and $\xi \in \mathbb{R}^p$ ($n = p = 3$ in the simulations shown in the main text). Complexity $q \in [0, \dots, np]$ is defined as the number of nonzero entries in the matrix, added first to the main diagonal and then to the off-diagonals in a predetermined order—first the superdiagonal starting at $(1, 2)$, then the one starting at $(1, 3)$, etc. [so for the 3×3 case, $q = 3$ maps are the class of diagonal matrices, $q = 4$ maps have nonzeros in their diagonal and in their $(1, 2)$ entry, etc.].

b. Polynomials

Here we consider mappings ϕ defined as coordinate-wise polynomials parametrized by q coefficients $\{a_k\}_{k=1}^q$. The map is

$$\phi : \vec{\xi} \mapsto \begin{pmatrix} \sum_{k=1}^q a_k \xi_1^k \\ \sum_{k=1}^q a_k \xi_2^k \\ \vdots \\ \sum_{k=1}^q a_k \xi_p^k \end{pmatrix}.$$

Complexity q is thus the number of tunable coefficients in the polynomial. All parameters a_k are sampled i.i.d. from $\mathcal{N}(0, 1)$.

c. One-hidden-layer neural networks

These are defined as

$$\phi : \vec{\xi} \mapsto W^{(2)}\sigma(W^{(1)}\vec{\xi}),$$

where $W^{(1)} \in \mathbb{R}^{q \times n}$, $W^{(2)} \in \mathbb{R}^{p \times q}$, and $\sigma(x) = \frac{1}{1+e^{-x}}$.

Parameters $W_{i,j}$ for all functions are sampled i.i.d. from $\mathcal{N}(0, 1)$.

2. Simulations

Simulation dynamics consisted of applying the replicator equation [Eq. (1)] to populations of functions as described in the main text. All simulations were carried out over populations of size 2000 per complexity class, for $T = 1000$ time steps with parameters $\gamma = 0.05$ and $\sigma = 1$ except where explicitly mentioned otherwise. All Python code required to reproduce the results is available via Ref. [56].

APPENDIX B: DERIVATIONS

1. Class fitness

In a population evolving according to Eq. (1), the frequency of any member i in the next generation, denoted x'_i , is given by

$$x'_i = \frac{x_i f_i}{\sum_j x_j f_j}.$$

The frequency X_q of any class C_q is thus

$$X'_q = \sum_{i \in C_q} x'_i = \frac{\sum_{i \in C_q} x_i f_i}{\sum_j x_j f_j} = \frac{\sum_{i \in C_q} x_i f_i}{\sum_l \sum_{j \in C_l} x_j f_j}.$$

If we define

$$F_q = \frac{\sum_{i \in C_q} x_i f_i}{X_q} = \frac{\sum_{i \in C_q} x_i f_i}{\sum_{i \in C_q} x_i},$$

we recover Eq. (4),

$$X'_q = \frac{X_q F_q}{\sum_l X_l F_l} \propto X_q F_q.$$

Thus F_q plays an analogous role to individual member fitness f_i , but at the class level.

2. Occam factor

For any $i \in C_q$, let $\tilde{x}_i = \frac{x_i}{X_q}$ (i.e., normalized to class frequency). We have

$$f_i \frac{\tilde{x}_i}{\tilde{x}'_i} = f_i \frac{x_i}{x'_i} \frac{X'_q}{X_q} \stackrel{(1)}{=} \sum_j x_j f_j \frac{X'_q}{X_q} \stackrel{(2)}{=} \frac{\sum_{i \in C_q} x_i f_i}{\sum_{i \in C_q} x_i} = F_q.$$

(1) Substituting x'_i for $\frac{x_i f_i}{\sum_j x_j f_j}$.

(2) Substituting X'_q for $\frac{\sum_{i \in C_q} x_i f_i}{\sum_j x_j f_j}$.

Denoting $i_q^*(t)$ as the index of the member with highest fitness at time t in class q , we recover Eq. (5).

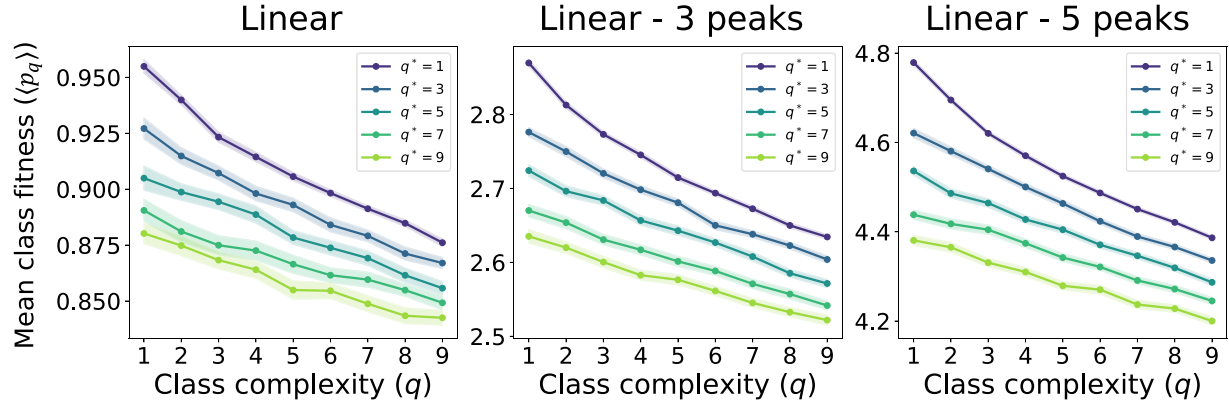


FIG. 5. Mean class fitness $\langle p_q \rangle$ (y-axis) as a function of class complexity q (x-axis) for linear and multiple-peak linear models. Each curve corresponds to a specific environmental complexity value. Error bands show ± 1.96 SEM over 100 independent seeds; selection strength $\gamma = 0.05$, noise $\sigma = 0.1$.

3. Convergence to globally optimal member

For any population evolving according to Eq. (1), we have

$$x_i(t) \propto x_i(0) \prod_{t=1}^T f_i(t)$$

and therefore

$$\log x_i(T) = \log [x_i(0)] + \sum_{t=1}^T \log [f_i(t)] - Z,$$

where Z is a normalizing constant. For sufficiently large T , by the law of large numbers we have for any i, j

$$\log x_i(T) - \log x_j(T) = T(\langle \log f_i \rangle - \langle \log f_j \rangle) + o(1)$$

and thus, denoting i^* as the index of the member with the largest mean log fitness, we have

$$x_{i^*}(t) \xrightarrow{t \rightarrow \infty} 1$$

and for the population's mean fitness F ,

$$F(t) \xrightarrow{t \rightarrow \infty} f_{i^*}$$

4. Fisher's fundamental theorem

For a class C_q evolving according to Eq. (1) with fixed fitness over time, we have

$$\begin{aligned} F_q(t+1) &= \sum_{i \in C_q} \tilde{x}_i(t+1) f_i(t+1) \\ &= \frac{\sum_{i \in C_q} \tilde{x}_i(t) f_i(t) f_i(t+1)}{\sum_{j \in C_q} \tilde{x}_j(t) f_j(t)} \\ &= \frac{\sum_{i \in C_q} \tilde{x}_i(t) f_i^2}{\sum_{j \in C_q} \tilde{x}_j(t) f_j} \\ &= \frac{\sum_{i \in C_q} \tilde{x}_i(t) f_i^2}{F_q(t)} \end{aligned}$$

and therefore

$$\Delta F_q(t) = F_q(t+1) - F_q(t) = \frac{\sum_{i \in C_q} \tilde{x}_i(t) f_i^2 - F_q^2(t)}{F_q(t)}.$$

The values f_i are assumed to be sampled from a distribution p_q . At time 0, when class frequencies x_i are uniform, we have

$$\langle \Delta F_q(0) \rangle_{p_q} = \frac{\langle f^2 \rangle_{p_q} - \langle f \rangle_{p_q}^2}{\langle f \rangle_{p_q}} = \frac{\text{Var}_{p_q}(f)}{\langle f \rangle_{p_q}}.$$

APPENDIX C: UNIVERSALITY OF THE INITIAL ADVANTAGE FOR SIMPLE CLASSES

The initial growth rate of a class is proportional to its weighted mean fitness at time 0,

$$\frac{X_q^{(1)}}{X_q^{(0)}} \propto F_q^{(0)} = \frac{\sum_{i \in C_q} x_i^{(0)} f_i^{(0)}}{X_q^{(0)}} = \sum_{i \in C_q} \tilde{x}_i^{(0)} f_i^{(0)}. \quad (\text{C1})$$

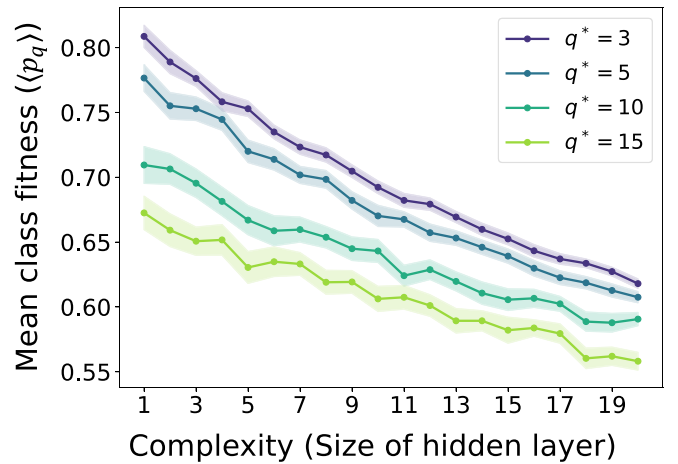


FIG. 6. Mean class fitness $\langle p_q \rangle$ as a function of complexity q for one-hidden-layer networks with sigmoid activations (hidden dimension q). Each curve corresponds to a specific environmental complexity value. Error bands show ± 1.96 SEM over 100 independent seeds; selection strength $\gamma = 0.05$, noise $\sigma = 0.1$.

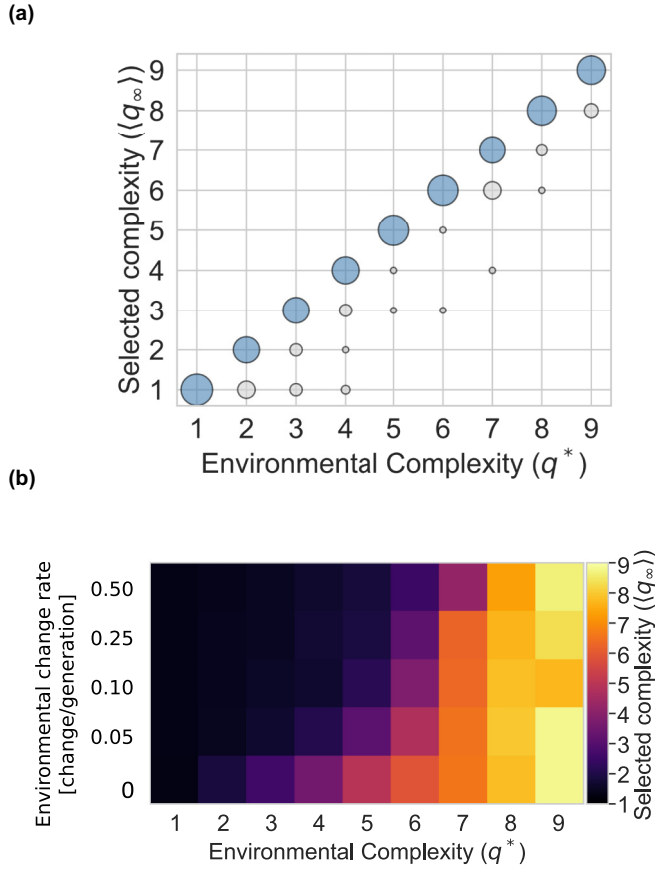


FIG. 7. Selected complexity in stable and changing environments for polynomials. (a) Circle sizes represent, for a given q^* (x-axis), the fraction out of 100 simulations in which time-averaged max fitness was found in a given complexity class (y-axis), with the highest fraction marked in blue. (b) Selected class (average class at the end of selection) as a function of both environmental complexity (x-axis) and environmental change rate (here, the number of different environments in a simulation of length $T = 1000$) (y-axis). Results are based on 100 simulation realizations. Population sizes of 500 per complexity class, $\sigma = 0.1$, $\gamma = 0.05$, and $T = 1000$.

If class composition is uniform (or equivalently if class composition is sampled from some arbitrary distribution that is independent of the ground truth), this is none other than the mean of the distribution from which class members sample their fitness,

$$\begin{aligned} \langle F_q^{(0)} \rangle &= \left\langle \sum_{i \in C_q} \tilde{x}_i^{(0)} f_i^{(0)} \right\rangle = \sum_{i \in C_q} \langle \tilde{x}_i^{(0)} \rangle \langle f_i^{(0)} \rangle \\ &= \sum_{i \in C_q} \langle \tilde{x}_i^{(0)} \rangle \langle p_q \rangle = \langle p_q \rangle. \end{aligned} \quad (C2)$$

Given a fitness function for an environment-phenotype function ϕ of the form used in this work,

$$f_\phi^{(t)} = d(\phi(\xi^{(t)}), \phi^*(\xi^{(t)})),$$

where d is strictly increasing, $\xi^{(t)}$ is the random input of generation t , and ϕ^* is the fixed ground truth mapping. The mean fitness value for a mapping of complexity q ($\langle p_q \rangle$) decreases with q whenever organisms of higher complexity produce

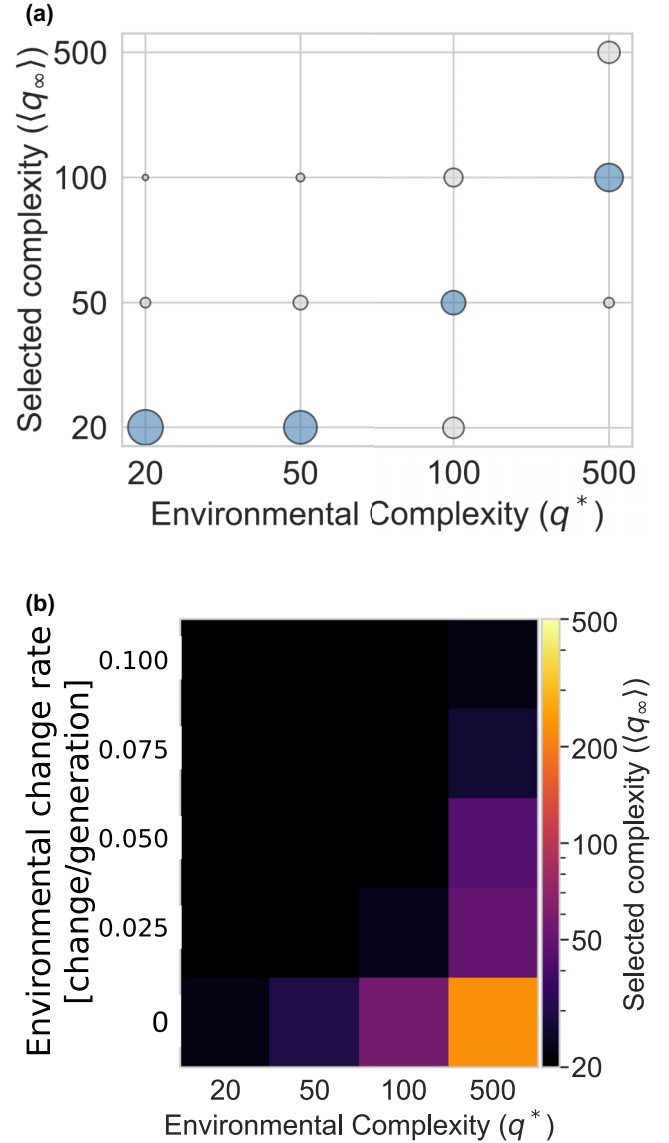


FIG. 8. Selected complexity in stable and changing environments for neural networks. (a) Circle sizes represent, for a given q^* (x-axis), the fraction out of 100 simulations in which time-averaged maximum fitness was found in a given complexity class (y-axis), with the highest fraction marked in blue. (b) Selected class (average class at the end of selection) as a function of both environmental complexity (x-axis) and environmental change rate (here, the number of different environments in a simulation of length $T = 1000$) (y-axis). Results are based on 100 simulation realizations. Population sizes of 2000 per complexity class, $\sigma = 1$, $\gamma = 0.01$, and $T = 1000$.

outputs whose typical distance to the target grows with q . This is a direct consequence of the *curse of dimensionality*: random organisms in a higher-dimensional parameter space produce more dispersed outputs, and therefore lie farther from the target on average.

For a concrete easily provable example of this general statement, consider the linear model studied in the main text with fitness function for mapping A given by $f = e^{-\gamma \|A\xi - A^*\xi\|}$, where A^* is the ground truth mapping, ξ is the environmental cue (with cues drawn i.i.d. each generation

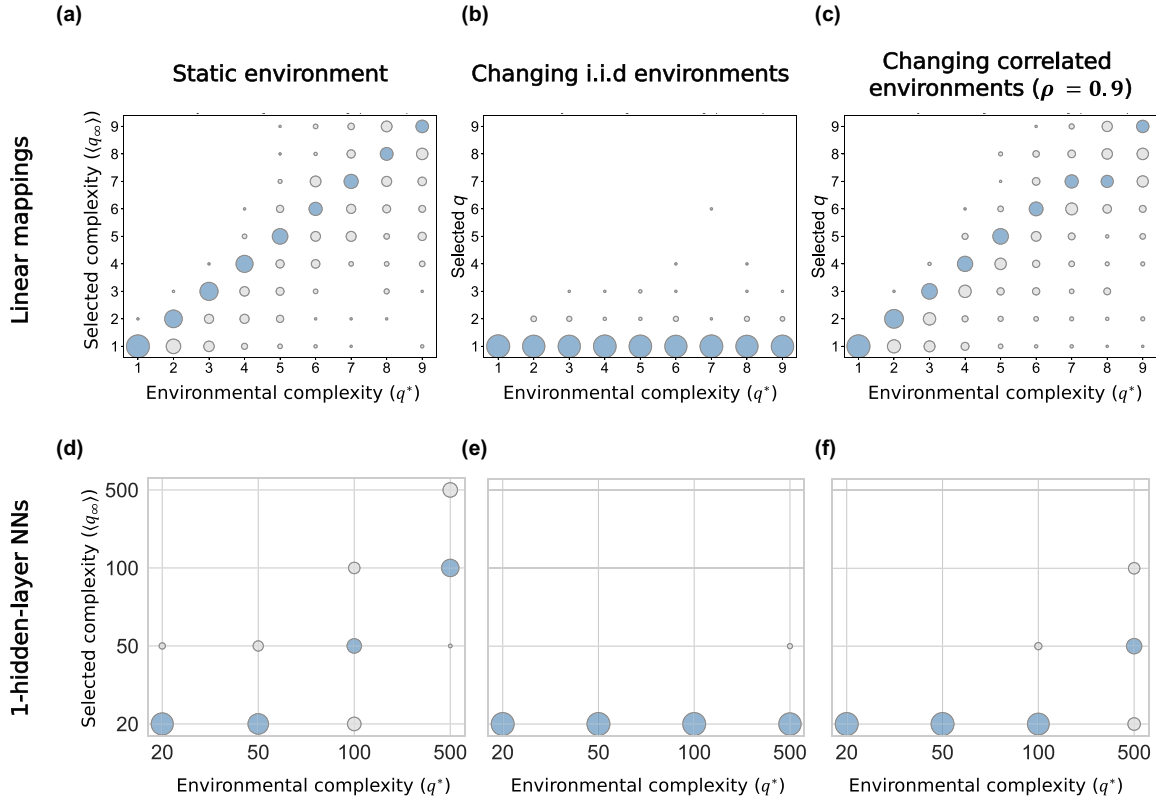


FIG. 9. Temporal correlation between successive environments attenuates the complexity-reducing effect of environmental change. Bubble charts show the fraction of realizations (circle size) in which the final selected complexity (y-axis) takes each value, as a function of environmental complexity q^* (x-axis), for a ground truth A^* changing at the indicated rate and each change following an AR(1) process $A_{t+}^* = \rho A_{t-}^* + \sqrt{1 - \rho^2} \eta$, where $\eta^{(t)}$ are i.i.d. draws from the same ground truth distribution from which $A^*(0)$ is drawn. (a) Selected class as a function of environmental complexity in a fixed environment (change rate = 0). As in Fig. 2, the selected complexity closely matches the environmental complexity. (b) Same analysis with environmental change rate 0.05 and independent redraws of the ground truth at each change event ($\rho = 0$). In this regime, simple classes are favored. (c) Same environmental change rate as in (b) but with strong temporal correlation between successive ground truth matrices A^* ($\rho = 0.99$). Under these conditions, the correspondence between selected complexity and environmental complexity is largely restored. (d)–(f) Same as (a)–(c), but with one-hidden-layer neural networks (NNs) instead of linear functions. All simulations use 200 realizations of linear maps or NNs as indicated and otherwise the same parameters as in the main text for these function classes.

from a zero-mean unit-variance distribution with uncorrelated components), and γ controls selection strength. Here, a class- q organism has q nonzero parameter entries in its matrix A drawn i.i.d. from $\mathcal{N}(0, \sigma^2)$. The mean fitness $\langle f \rangle$ can be written as

$$\langle f \rangle = \langle e^{-\gamma d_q} \rangle = 1 - \gamma \langle d_q \rangle + O(\gamma^2)$$

and so in the small γ (weak selection) regime, mean fitness decreases with increasing $\langle d_q \rangle$. We next show that for linear functions, $\langle d_q \rangle$ is monotonically increasing with complexity.

Proof. Let A be a random matrix and A^* be the ground truth matrix. Denote $\Delta := A - A^*$,

$$\begin{aligned} \langle d_q \rangle_\xi &= \langle \|A\xi - A^*\xi\|^2 \rangle_\xi = \langle \|\Delta\xi\|^2 \rangle_\xi \\ &= \langle (\Delta\xi)^\top (\Delta\xi) \rangle_\xi = \langle \xi^\top \Delta^\top \Delta \xi \rangle_\xi \\ &= \langle \text{tr}(\Delta^\top \Delta \xi \xi^\top) \rangle_\xi = \text{tr}(\Delta^\top \Delta \langle \xi \xi^\top \rangle_\xi). \end{aligned}$$

Since the inputs are zero-mean, uncorrelated, and with unit variance component-wise, we have $\langle \xi \xi^\top \rangle_\xi = I$. Denoting the

set of nonzero entries of A as $\text{supp}(A)$, we have

$$\begin{aligned} \langle d_q \rangle_{\xi, A} &= \langle \|A\xi - A^*\xi\|^2 \rangle_{\xi, A} \\ &= \langle \|A - A^*\|_F^2 \rangle_A \\ &= \sum_{(i,j) \in \text{supp}(A)} \langle (A_{ij} - A_{ij}^*)^2 \rangle_A + \sum_{(i,j) \notin \text{supp}(A)} (A_{ij}^*)^2 \\ &= \sum_{(i,j) \in \text{supp}(A)} ((A_{ij}^*)^2 + \sigma^2) + \sum_{(i,j) \notin \text{supp}(A)} (A_{ij}^*)^2 \\ &= \|A^*\|_F^2 + q\sigma^2. \end{aligned}$$

Combining this with $\langle p_q \rangle = \langle e^{-\gamma d_q} \rangle \leq e^{-\gamma \langle d_q \rangle}$, we have $\langle p_q \rangle$ strictly decreasing with q .

Crucially, this argument for mean fitness decreasing with complexity applies to *any* function class for which the expected output distance to the target grows with the number of free parameters. Below we verify this numerically for two classes with fundamentally different fitness landscape geometries.

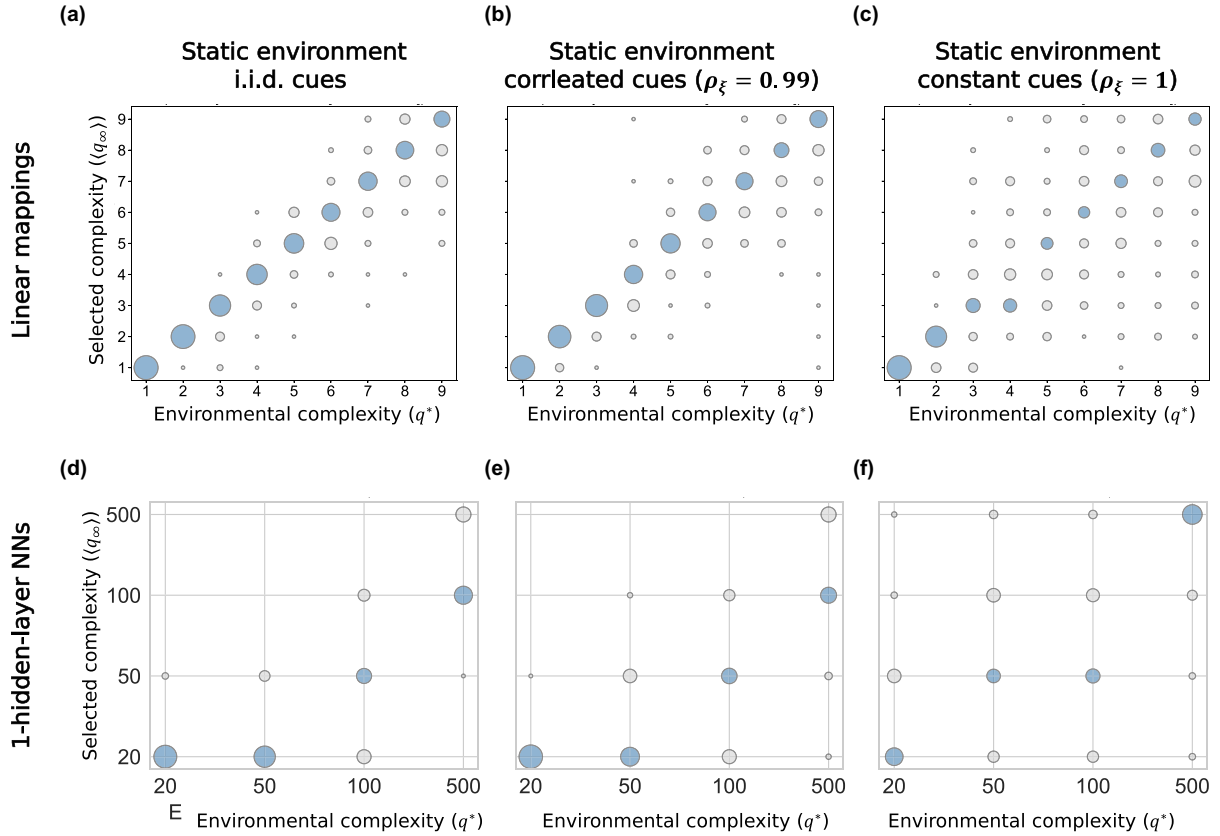


FIG. 10. Correlated cues in a fixed environment reduce selection accuracy. Bubble charts show the fraction of realizations (circle size) in which the final selected complexity (y-axis) takes each value, as a function of environmental complexity q^* (x-axis), for a fixed ground truth A^* and temporally correlated cue sequences $\{\xi^{(t)}\}$. Cues follow an AR(1) process $\xi^{(t+1)} = \rho_\xi \xi^{(t)} + \sqrt{1 - \rho_\xi^2} \zeta^{(t)}$, where $\zeta^{(t)}$ are i.i.d. draws from the same cue distribution as in the main text. (a) $\rho_\xi = 0$ (i.i.d. cues; standard setting). (b) $\rho_\xi = 0.99$ (strongly correlated cues). (c) $\rho_\xi = 1$ (identical cue at every time step). Increasing cue correlation mildly decreases the concentration of outcomes on $q_\infty = q^*$, but does not induce a shift toward higher selected complexity. Even with constant cues, stochasticity from the fitness noise term is sufficient to prevent systematic dominance of high-complexity classes. (d)–(f) Same as (a)–(c), but with one-hidden-layer neural networks (NNs) instead of linear functions. All simulations use 200 realizations of linear maps or NNs as indicated and otherwise the same parameters as in the main text for these function classes.

1. Numerical validation across fitness landscapes

We test whether $\langle p_q \rangle$ decreases with q for three qualitatively distinct function classes. For each class and ground truth complexity value q^* , we sample 2000 functions at random, evaluate their fitness against the fixed ground truth over 200 random input signals, and average; the procedure is repeated for 200 independent random samples of the ground truth function to obtain error estimates. Results are shown in Fig. 5.

a. Linear maps (main-text model)

Here we consider $n \times n$ matrices ($n = 3$) with q nonzero entries drawn from $\mathcal{N}(0, 1)$. Here $\langle p_q \rangle$ decreases monotonically with q for all q^* (Fig. 5, left), consistent with the analytic argument above.

b. Multipeak fitness landscape

Here we relax the single-optimum assumption by using fitness

$$f = \sum_{j=1}^J e^{-\gamma \|A\xi - A_j^* \xi + \sigma \eta_j\|}, \quad (\text{C3})$$

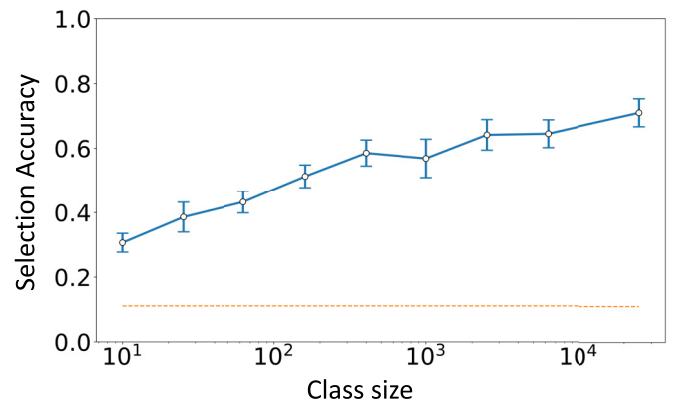


FIG. 11. The effect of class size on the strength of implicit regularization. Simulations carried out over increasing population sizes (x-axis) of linear functions result in increasingly larger fractions of realizations, which end up with the dominant class at the end of selection equal to q^* . Results from 100 simulation realizations per value of q^* in the range $[1, 9]$, with parameters $T = 5000$, $\gamma = 0.05$, and $\sigma = 1$.

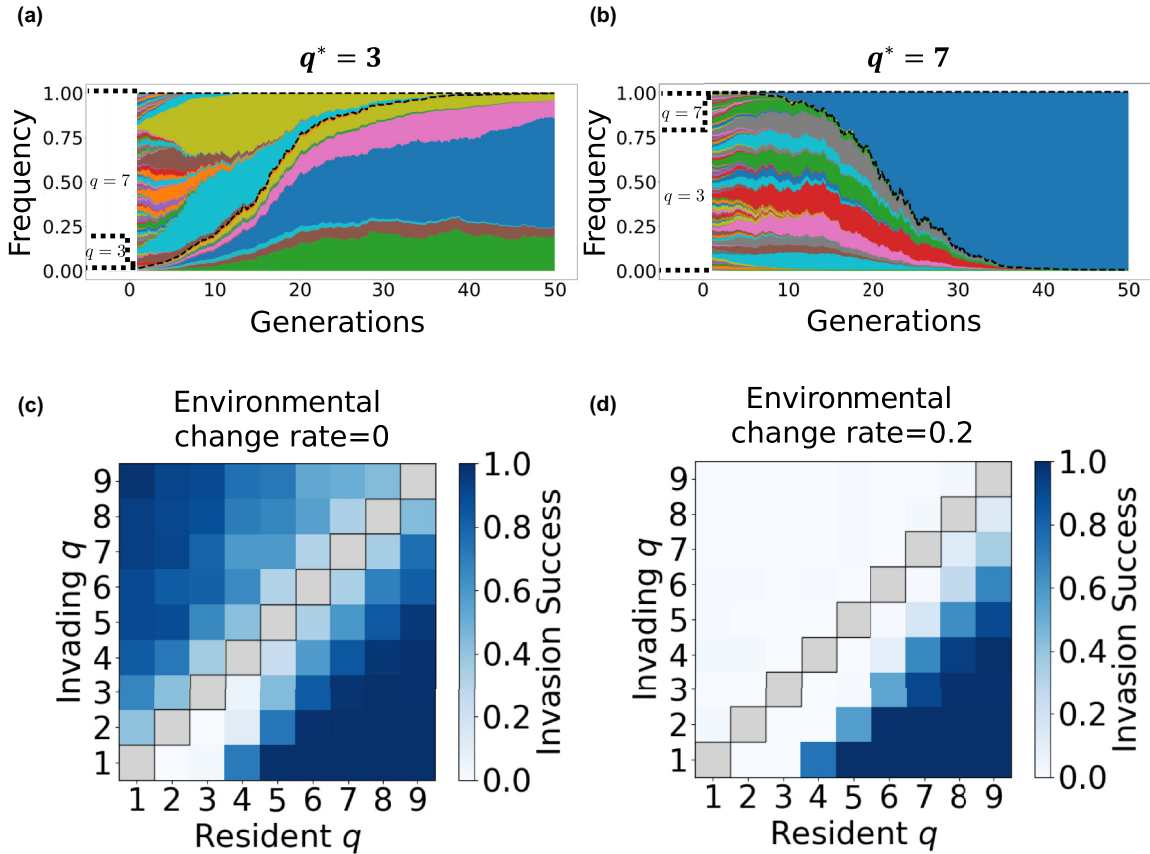


FIG. 12. Invasion simulations. (a), (b) Resident populations whose complexity mismatches environmental complexity are typically successfully invaded by a small (1%) population, which matches the environmental complexity. (c) Invasion success [fraction out of 200 invasion simulations, similar to (a),(b), in which the invading population dominates at the end of selection] for all pairs of resident and invading complexity, where the environmental complexity, q^* , is always equal to the invading complexity. (d) Same as (c) but with changing environments at a rate of 0.2 changes per generation. Results in (c), (d) from 200 simulation realizations per resident-invading pair. All results in (a)–(d) with population sizes of 2000 per complexity class, $T = 1000$, $\gamma = 0.05$, and $\sigma = 1$.

where $A_1^*, \dots, A_J^*$ are J independently drawn ground truth matrices of complexity q^* . An organism is rewarded for being close to *any* of the J optima (an *OR* structure). Despite the multipeak landscape, $\langle p_q \rangle$ remains a decreasing function of q (Fig. 5, middle, right); increasing J shifts all curves upward (more peaks to hit) but preserves the ordering across q .

c. One-hidden-layer neural networks

Here we consider organisms that are one-hidden-layer networks with sigmoid hidden dimension q , input/output dimension $n = 3$, and all weights and biases drawn from $\mathcal{N}(0, 1)$. The ground truth is a fixed random network with hidden dimension q^* . This model has *multiple optima in parameter space* (weight-space symmetries mean many parameter configurations realize the same function), yet $\langle p_q \rangle$ decreases with q across all q^* values (Fig. 6).

APPENDIX D: IMPLICIT REGULARIZATION IN DIVERSE FUNCTION CLASSES

The results in the main text focused on selection of linear functions. However, the phenomena described are universal (qualitatively), across different types of functions. We demonstrate this using populations of polynomials and

one-hidden-layer neural networks (see Methods for implementation details).

Polynomials were chosen as an illustrative case as they are a classic example of the effect of overfitting [57]. Here, we see the same behavior as with the linear functions. Namely, selected class corresponds closely to the environmental complexity [Fig. 7(a)], and increasing environmental change rate decreases selected complexity [Fig. 7(b)].

Neural networks were chosen for their nature as universal function approximators [58], and for their use as models of gene regulatory networks [37]. While these functions do not have a one-to-one correspondence between environmental complexity and selected complexity as linear functions, they still show a trend of increasing selected complexity with increasing environmental complexity [Fig. 8(a)]. Changing environments has an effect of decreasing selected complexity [Fig. 8(b)], as in the case of linear functions.

APPENDIX E: CORRELATED ENVIRONMENTS

In the main text, environmental change was implemented as independent redraws of the ground truth mapping, i.e., each change event produces a new ground truth matrix A^* whose fitness landscape is effectively uncorrelated with

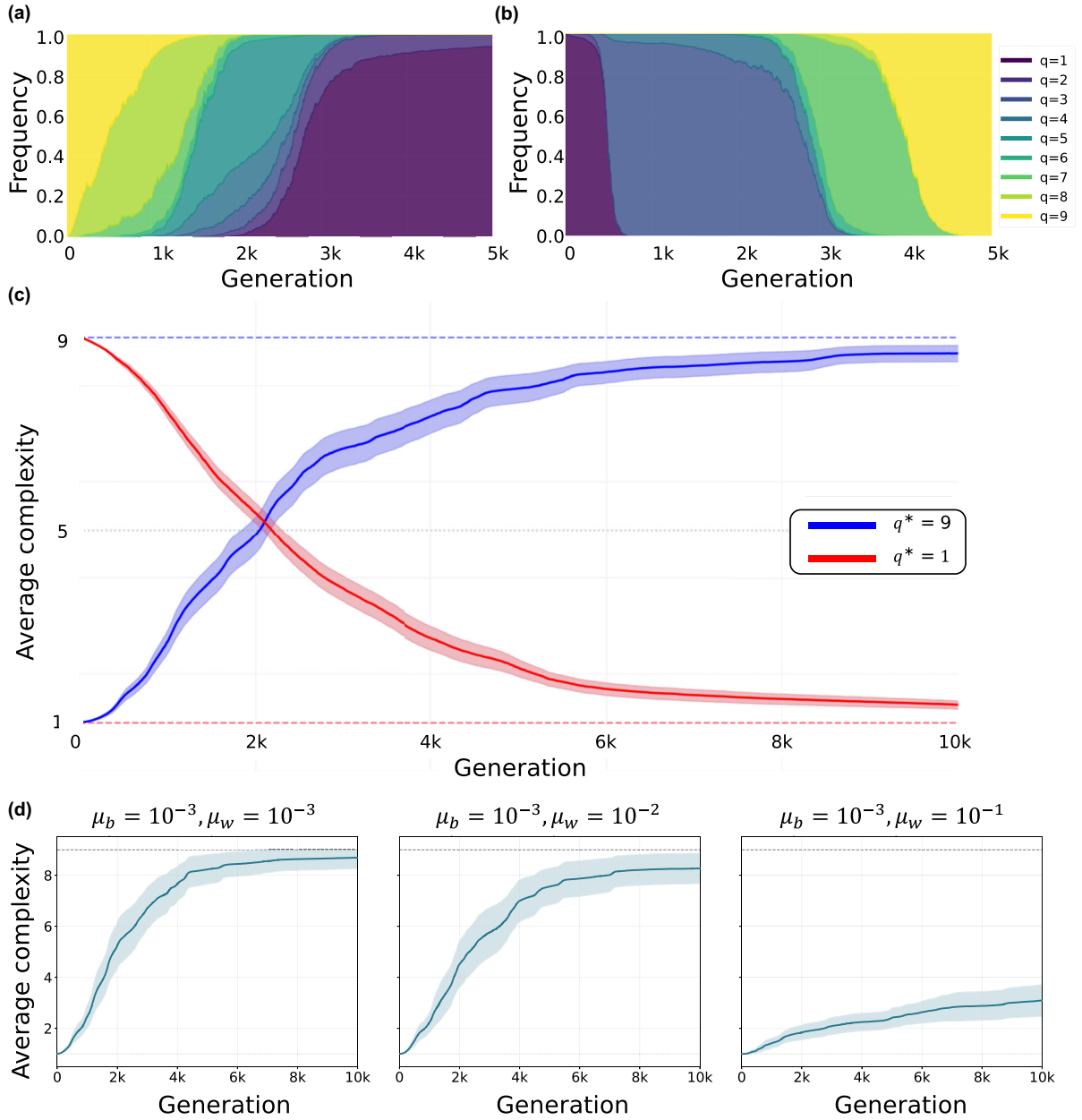


FIG. 13. Selection with mutation drives stepwise evolution toward the environmental complexity. (a) Typical realization in a simple environment ($q^* = 1$), starting from a population entirely concentrated in the maximal complexity class ($q^0 = 9$). Between-class mutations and selection drive a progressive shift through intermediate classes until low-complexity classes dominate. (b) Typical realization in a complex environment ($q^* = 9$), starting from a population entirely concentrated in the minimal complexity class ($q^0 = 1$). Stepwise between-class mutations generate a gradual increase in complexity until high-complexity classes dominate. (c) Mean population complexity as a function of time for the two scenarios in (a), (b), averaged over 200 simulation realizations. Dashed horizontal lines indicate the environmental complexity in each case. Shaded bands show ± 1.96 SEM. (d) Effect of within-class mutation rate on adaptation toward a complex environment ($q^* = 9$, $q^0 = 1$), for fixed between-class mutation rate $\mu_b = 10^{-3}$. For $\mu_w \leq \mu_b$, the mean population complexity still approaches the environmental complexity, whereas for $\mu_w \gg \mu_b$ the population remains at substantially lower complexity. Shaded bands show ± 1.96 SEM over 200 simulation realizations. Results are based on simulations with linear maps, $T = 10\,000$, class size = 200, $M = 20\,000$, $\mu_b = 10^{-3}$, $\gamma = 0.05$, $\xi = 1.0$, and $n = 3$. In (a)–(c), $\mu_w = 0$; in (d), μ_w is varied as indicated.

the previous one. This is a maximally disruptive model of change. To capture gradual drift, we instead introduce temporal correlation between successive ground truth matrices. Specifically, when an environment switch occurs, we update the ground truth by an order-1 autoregressive

rule [AR(1)]

$$A_{t^+}^* = \rho A_{t^-}^* + \sqrt{1 - \rho^2} \eta, \quad (\text{E1})$$

where t^- and t^+ denote the time steps immediately before and after the switch, and η is an innovation matrix drawn from the

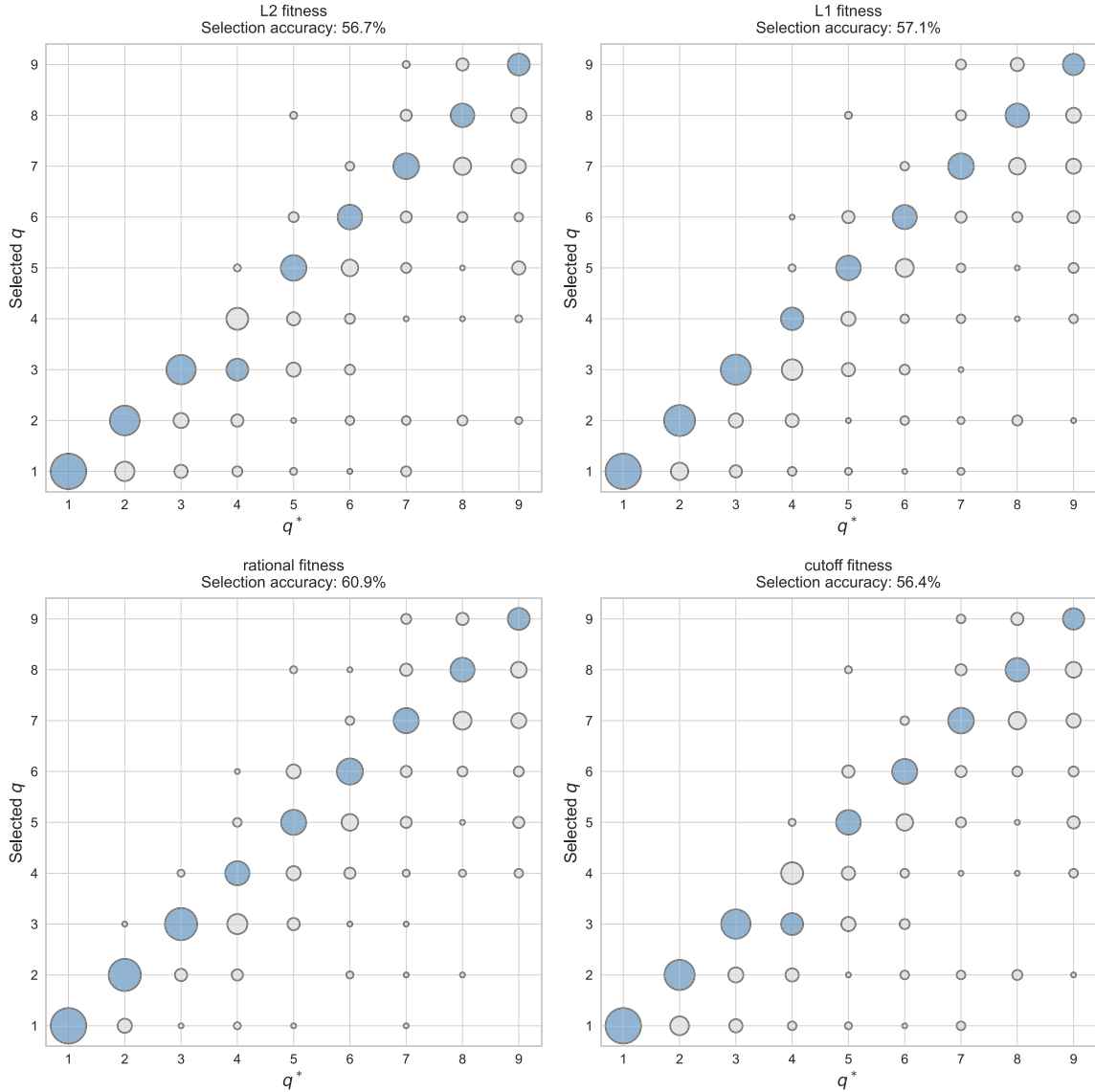


FIG. 14. Selected complexity in stable environments for different fitness functions. Bubble charts show selected complexity (y-axis) as a function of environmental complexity (x-axis) based on the same experiment as in Fig. 2(b) with linear functions and the same parameters but using the indicated fitness function per plot (see Table I for details).

same distribution as the initial ground truth A_0^* (with the same complexity q^* , i.e., the same sparsity pattern and i.i.d. nonzero entries). Thus, $\rho = 0$ recovers the independent-redraw setting used in the main text, while $\rho \rightarrow 1$ yields small perturbations of the current environment. The main text discussed that while in a fixed environment selection matches environmental complexity [Fig. 9(a)], frequent uncorrelated switches bias selection toward low complexity [Fig. 9(b)]. However, correlation between successive environments attenuates this effect. At a change-rate that selects for simple classes irrespective of environmental complexity [Fig. 9(b)], adding a $\rho = 0.99$ correlation between successive environment largely restores the q^* -matching trend [Fig. 9(c)]. Mechanistically, uncorrelated switches repeatedly reshuffle which members are optimal within each class, forcing large within-class “restarts” that favor simple classes; strong correlation preserves partial ordering of fitness across members, allowing previously fit

types (especially in the q^* class) to remain competitive after a switch and be amplified rather than rediscovered.

We next applied the same idea to the *cue process* in a fixed environment. Here A^* is held constant throughout the simulation, but the cue vectors $\{\xi^{(t)}\}$ are drawn from a temporally correlated process with autocorrelation ρ_ξ :

$$\xi^{(t+1)} = \rho_\xi \xi^{(t)} + \sqrt{1 - \rho_\xi^2} \zeta^{(t)}, \quad (\text{E2})$$

where $\zeta^{(t)}$ are i.i.d. draws from the same cue distribution used in the main text (so $\rho_\xi = 0$ corresponds to i.i.d. cues, and $\rho_\xi = 1$ corresponds to presenting the same cue at every time step). Increasing cue correlation mildly reduces selection accuracy (Fig. 10): outcomes are less concentrated on the diagonal $q_\infty = q^*$ as ρ_ξ increases, including in the extreme $\rho_\xi = 1$ case. Even when cues are constant, fitness evaluations remain stochastic due to the additive noise term in the

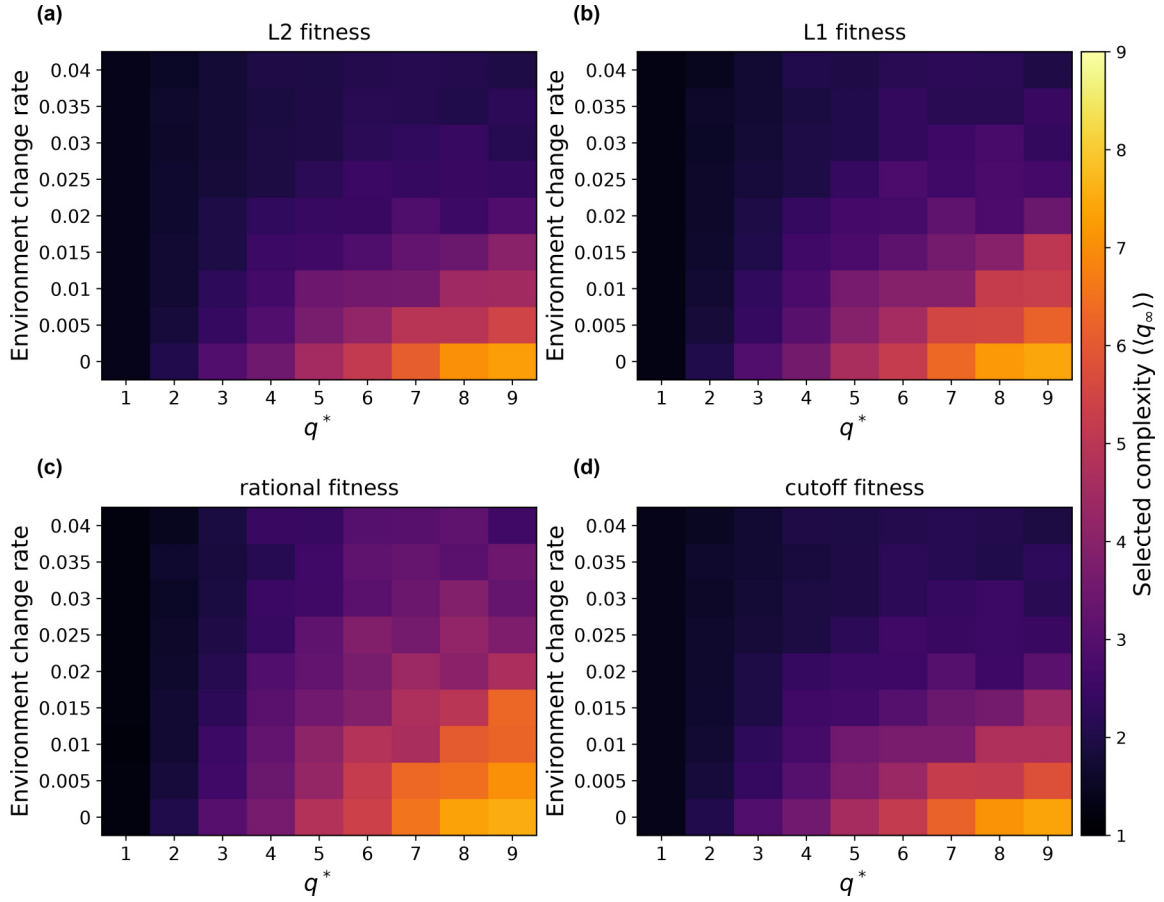


FIG. 15. Selected complexity in changing environments for different fitness functions. Color indicates average selected complexity for different environmental complexity values (x-axis) and environmental change rates (y-axis) based on the same experiment as in Fig. 4(e) with linear functions and the same parameters but using the indicated fitness function per plot (see Table I for details).

selection equation, so high-complexity classes can still overfit transient noise realizations rather than gaining a systematic advantage from repeated exposure to the same cue.

APPENDIX F: EFFECT OF POPULATION SIZE

Function classes with more parameters are capable of more complex mappings, but they are also, by definition, larger.

Most results in learning theory comparing function classes of different dimensionality assume that the entire continuum of possible parametrizations in each class is considered. In our case, however, we compare populations that are finite-sized random samples from each class.

If the optimal mapping ϕ^* contains d_{opt} parameters and populations are of fixed size n , then for any dimension $d > d_{\text{opt}}$ the probability of the population containing a member

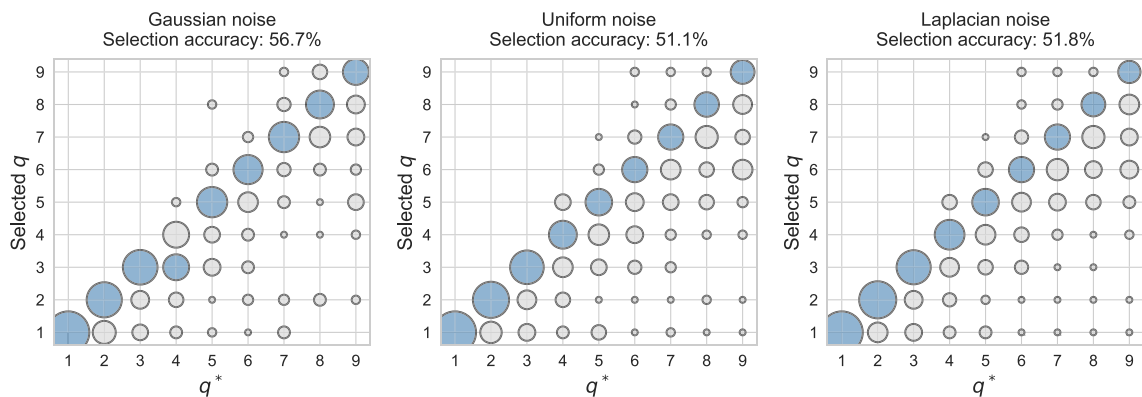


FIG. 16. Selected complexity in stable environments for different noise distributions. Bubble charts showing selected complexity (y-axis) as a function of environmental complexity (x-axis) based on the same experiment as in Fig. 2(b) with linear functions and the same parameters but using the indicated noise model per plot (see Table II for details).

within some close distance to ϕ^* decreases with d but increases with n .

To illustrate this effect, we measured the correspondence between environmental complexity and selected complexity in linear functions, as described in the main text, but for varying population sizes. For each given population size, we ran 900 simulation realizations, 100 for each q^* in the range [1,9], and we recorded the fraction of simulations where the selected complexity matched q^* (Fig. 11). As expected, we observed a trend of increasing “precision” of the selection process for increasing population size.

APPENDIX G: EFFECT OF INITIAL CONDITIONS

All results in the main text refer to uniform initial population composition, both within and between complexity classes. To demonstrate the robustness of the main findings to this assumption, we consider *invasion experiments*, where a resident population of the “wrong” complexity (one that does not match environmental complexity) is seeded at t_0 with a small (1%) subpopulation whose complexity is q^* (equal to the underlying environmental complexity). Typical realizations of this setting result in the invading population dominating [Figs. 12(a)–12(c)]. In the case of rapidly changing environments, however, simple invading populations are able to invade while complex invaders cannot [Fig. 12(d)], because when the environment changes faster than the transient discussed in the main text (Fig. 4 and accompanying text), a complex invading population, even if it contains the optimal type for any given environment, is unable to take hold before the environment changes.

APPENDIX H: MUTATIONS

The main text focused on pure selection acting on preexisting variation. A natural extension is then to ask to what extent the main findings persist when novelty is continuously generated throughout the dynamics. To this end, we extended the model to include *mutations*.

For this analysis, we focused on the linear setting, where environment-phenotype maps as well as the ground truth function are represented by matrices with q nonzero entries

(see Appendix A). At each time step t , a parent is chosen with probability proportional to its current frequency in the population and to a *mutation rate* μ . We split mutations into *within-class mutations*, which change one of the nonzero entries and occur at rate μ_w , and *between-class mutations*, which either add a nonzero entry or zero an existing one, thereby changing the complexity class of the mutant compared to its parent. These latter mutations occur at a rate μ_b . For a parent with frequency x_i , its mutant offspring starts off with frequency $\frac{x_i}{M}$, where M represents the number of individuals in the population, and this fraction is removed from its parent’s population to ensure the population sums to unity. We start by considering only between-class mutations ($\mu_w = 0$).

Unlike the invasion experiments above, here the population is not seeded with members of the environmentally matched class. Nevertheless, we observe the same qualitative trend as in the mutation-free setting. In a simple environment ($q^* = 1$), a population initialized at the maximal complexity class ($q^0 = 9$) evolves downward through a sequence of intermediate classes until low-complexity types dominate [Fig. 13(a)]. Conversely, in a complex environment ($q^* = 9$), a population initialized at the minimal complexity class ($q^0 = 1$) evolves upward, again through successive intermediate classes, until high-complexity types dominate [Fig. 13(b)].

Averaging over realizations shows that the mean population complexity moves monotonically toward the environmental complexity from both directions [Fig. 13(c)]. Convergence is slower than in the main-text setting, since adaptation must now proceed through local moves in complexity space rather than by immediately selecting among all preexisting classes. In addition, because mutation continues to introduce neighboring classes even after adaptation, the long-time state is not an exact fixation at q^* but rather a mutation-selection balance centered near it. Thus, ongoing mutation does not abolish the complexity-matching tendency described in the main text; instead, it provides a mechanism by which populations can *discover* the appropriate complexity when it is not initially present.

We next asked how this picture changes when mutations within a complexity class are also allowed. Unlike

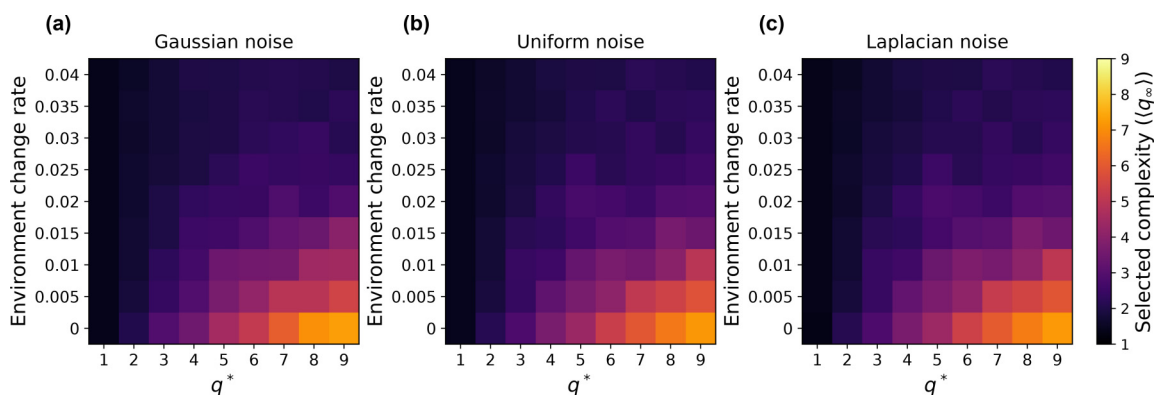


FIG. 17. Selected complexity in changing environments for different noise distributions. Color indicates average selected complexity for different environmental complexity values (x-axis) and environmental change rates (y-axis) based on the same experiment as in Fig. 4(e) with linear functions and the same parameters but using the indicated noise model per plot (see Table II for details).

between-class mutations, which move populations through complexity space, within-class mutations resample parameters while keeping complexity fixed and therefore act primarily to reshuffle the identity of candidate types within each class. For modest within-class mutation rates ($\mu_w \leq \mu_b$), the qualitative behavior remains unchanged: populations still evolve toward the environmental complexity, albeit somewhat more slowly. However, when within-class mutation becomes large compared to the between-class rate ($\mu_w > \mu_b$), adaptation toward high-complexity environments is strongly hindered [Fig. 13(d)]. In this regime, fit members within complex classes are continuously replaced before selection can amplify them, preventing those classes from stabilizing around well-adapted types. Thus, between-class mutation enables exploration of complexity, whereas excessive within-class mutation erodes the selective advantage of complex classes by continually disrupting their internal optimization.

APPENDIX I: SENSITIVITY ANALYSIS

1. Overview

This Appendix presents a comprehensive sensitivity analysis, testing robustness across alternative modeling choices and parameter variations for two of the main results in the paper:

- (1) In stable environments, selected complexity matches environmental complexity ($q_\infty \approx q^*$) [Fig. 2(b)].
- (2) Rapidly changing environments select for simpler organisms (lower q_∞) [Fig. 4(e)].

2. Alternative modeling choices

a. Fitness functions

We test four fitness function formulations (Table I), all relating fitness as a decreasing functions of prediction error $\phi(\xi) - \phi^*(\xi)$, where ϕ is the mapping whose fitness is measured, ϕ^* is the ground truth mapping, and ξ is an environmental cue. In all fitness functions, we use a parameter γ to control selection strength and a noise term $\epsilon \sim \mathcal{N}(0, 1)$.

All fitness functions produce results that are qualitatively similar to the ones reported in the main text. Namely, populations evolve to match environmental complexity (Fig. 14), and increasing environmental change rate decreases selected complexity (Fig. 15).

TABLE I. Alternative fitness function formulations tested in the sensitivity analysis.

Fitness function	Formula
L2 (baseline)	$f = \exp(-\gamma \ \phi(\xi) - \phi^*(\xi) + \epsilon\ _2)$
L1	$f = \exp(-\gamma \ \phi(\xi) - \phi^*(\xi) + \epsilon\ _1)$
Rational	$f = 1/(1 + \gamma \ \phi(\xi) - \phi^*(\xi) + \epsilon\ ^2)$
Cutoff	$f = \max(0, 1 - \gamma \ \phi(\xi) - \phi^*(\xi) + \epsilon\)$

TABLE II. Alternative noise distributions tested in the sensitivity analysis.

Distribution	Formula
Gaussian (baseline)	$\epsilon \sim \mathcal{N}(0, \sigma^2)$
Uniform	$\epsilon \sim U(-\sqrt{3}\sigma, \sqrt{3}\sigma)$
Laplacian	$\epsilon \sim \text{Laplace}(0, \sigma/\sqrt{2})$

b. Noise distributions

In the main text, we used a fitness function of the form $f = \exp(-\gamma \|\phi(\xi) - \phi^*(\xi) + \epsilon\|_2)$, where ϕ is the mapping whose fitness is measured, ϕ^* is the ground truth mapping, ξ is an environmental cue, and γ controls selection strength. The noise term ϵ was assumed standard normal. We test two other noise models, both parametrized to have equivalent variance (Table II).

All noise distributions produce results that are qualitatively similar to the ones reported in the main text. Namely, populations evolve to match environmental complexity (Fig. 16), and increasing environmental change rate decreases selected complexity (Fig. 17), demonstrating a robustness of our main findings to noise statistics.

3. Parameter sensitivity

In addition to testing different fitness model and noise distributions, we also tested the robustness of the implicit regularization effect (selected complexity evolves to match environmental complexity) to different parameter choices. We performed a parameter sweep for the four main parameters of the model (Table III), systematically varying each parameter while holding others at baseline values. For each point in this sweep, we measured a summary statistic capturing the degree to which selection converges on a complexity class matching environmental complexity. Namely, we define the *selection accuracy* as the fraction of realizations where the selected complexity class matches environmental complexity ($q_\infty = q^*$). This is the sum of all of the bubble sizes for bubbles sitting on the diagonal in Fig. 2(b). The resulting selection accuracies across parameter sweeps are shown in Fig. 18.

TABLE III. Parameter ranges tested in the sensitivity analysis.

Parameter	Symbol	Baseline	Range tested
Selection strength	γ	0.05	0.01–0.2
Noise level	σ	1.0	0.1–5.0
Population per class	N	2000	100–5000
Generations	T	1000	100–2000

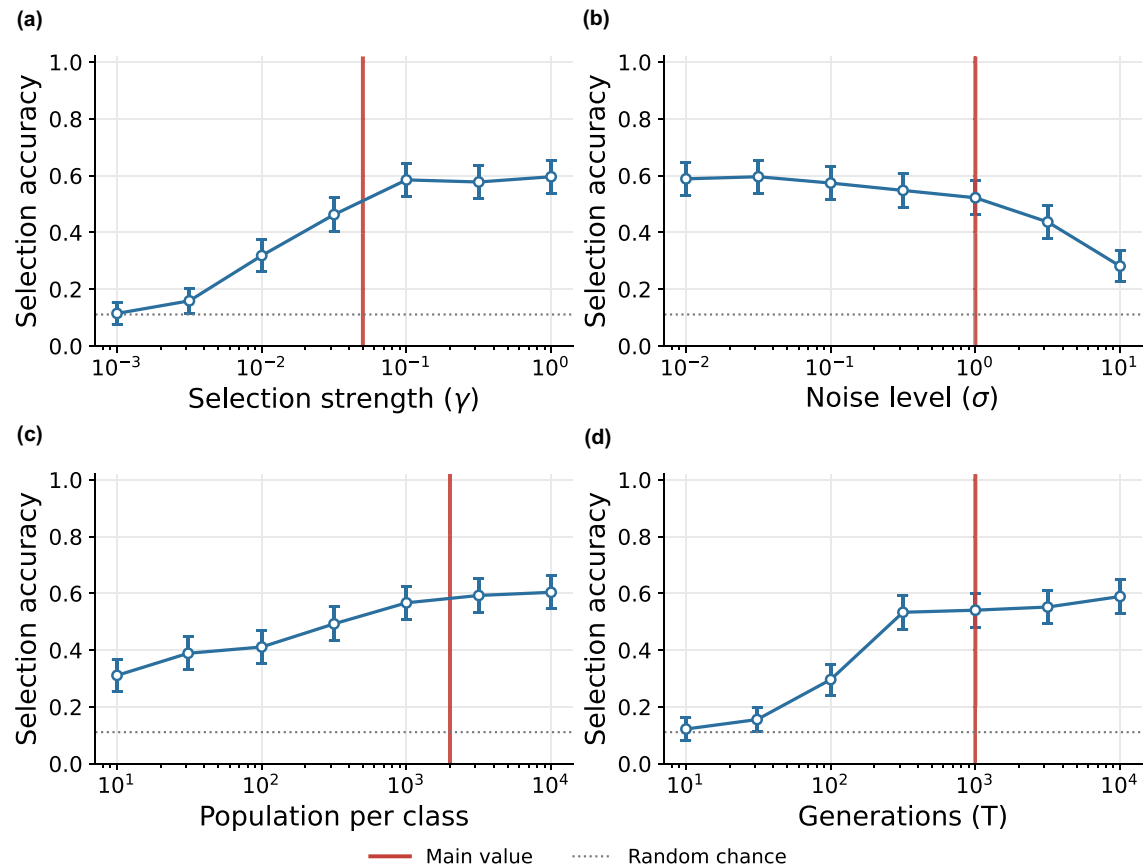


FIG. 18. Complexity matching for different parameter choices. For each parameter range of the four parameters in Table III (x-axes), we perform a sweep while holding the other three parameter at their baseline values (the values used in the main text, indicated for each parameter by a vertical red marker). For each parameter combination, we perform the same experiment as in Fig. 2(b) with linear functions and the same parameters except the one being changed. We measure selection accuracy (y-axes) defined as the fraction of realizations where the selected complexity class matches environmental complexity.

- [1] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press, New York, 2011), nachdr. edition.
- [2] D. W. McShea, Complexity and evolution: What everybody knows, *Biol. Philos.* **6**, 303 (1991).
- [3] Y. I. Wolf, M. I. Katsnelson, and E. V. Koonin, Physical foundations of biological complexity, *Proc. Natl. Acad. Sci. USA* **115**, E8678 (2018).
- [4] E. Szathmáry and J. M. Smith, The major evolutionary transitions, *Nature (London)* **374**, 227 (1995).
- [5] C. Adami, C. Ofria, and T. C. Collier, Evolution of biological complexity, *Proc. Natl. Acad. Sci. USA* **97**, 4463 (2000).
- [6] C. Darwin, *On the Origin of Species by Means of Natural Selection* (Murray, London, 1859), or the Preservation of Favored Races in the Struggle for Life.
- [7] H. Spencer, *The Principles of Biology* (Outlook Verlag, Frankfurt am Main, 2020), Vol. 1.
- [8] R. E. Lenski, C. Ofria, R. T. Pennock, and C. Adami, The evolutionary origin of complex features, *Nature (London)* **423**, 139 (2003).
- [9] S. B. Carroll, Chance and necessity: The evolution of morphological complexity and diversity, *Nature (London)* **409**, 1102 (2001).
- [10] P. J. Keeling and C. H. Slamovits, Simplicity and complexity of microsporidian genomes, *Eukaryotic Cell* **3**, 1363 (2004).
- [11] W. Veit, S. J. L. Gascoigne, and R. Salguero-Gómez, Evolution, complexity, and life history theory, *Biol. Theory* **20**, 212 (2025).
- [12] A. Wagner, Energy constraints on the evolution of gene expression, *Mol. Biol. Evol.* **22**, 1365 (2005).
- [13] M. Lynch, The bioenergetic cost of building a metazoan, *Proc. Natl. Acad. Sci. USA* **121**, e2414742121 (2024).
- [14] H. J. Muller, Some genetic aspects of sex, *Am. Nat.* **66**, 118 (1932).
- [15] N. A. Moran, Accelerated evolution and Muller's ratchet in endosymbiotic bacteria, *Proc. Natl. Acad. Sci. USA* **93**, 2873 (1996).
- [16] H. Maughan, J. Masel, C. W. Birky Jr, and W. L. Nicholson, The roles of mutation accumulation and selection in loss of sporulation in experimental populations of *Bacillus subtilis*, *Genetics* **177**, 937 (2007).
- [17] F. Heylighen, The growth of structural and functional complexity during evolution, *Evol. Complex.* **8**, 17 (1999).
- [18] H. A. Orr, Adaptation and the cost of complexity, *Evolution* **54**, 13 (2000).

- [19] G. P. Wagner, J. P. Kenney-Hunt, M. Pavlicev, J. R. Peck, D. Waxman, and J. M. Cheverud, Pleiotropic scaling of gene effects and the cost of complexity, *Nature (London)* **452**, 470 (2008).
- [20] C. Adami, What is complexity? *BioEssays* **24**, 1085 (2002).
- [21] R. Hinegardner and J. Engelberg, Biological complexity, *J. Theor. Biol.* **104**, 7 (1983).
- [22] P. D. Grünwald, *The Minimum Description Length Principle, Adaptive Computation and Machine Learning* (MIT Press, Cambridge, MA, 2007).
- [23] H. A. Simon, The architecture of complexity, *Proc. Am. Philos. Soc.* **106**, 467 (1962).
- [24] K. Dingle, C. Q. Camargo, and A. A. Louis, Input–output maps are strongly biased towards simple outputs, *Nat. Commun.* **9**, 761 (2018).
- [25] P. Godfrey-Smith, *Complexity and the Function of Mind in Nature* (Cambridge University Press, Cambridge, 1998).
- [26] A. P. Bird, Gene number, noise reduction and biological complexity, *Trends Genet.* **11**, 94 (1995).
- [27] R. B. Brem and L. Kruglyak, The landscape of genetic complexity across 5,700 gene expression traits in yeast, *Proc. Natl. Acad. Sci. USA* **102**, 1572 (2005).
- [28] M. H. V. Van Regenmortel, Biological complexity emerges from the ashes of genetic reductionism, in *HIV/AIDS: Immunochemistry, Reductionism and Vaccine Design: A Review of 20 Years of Research* (Springer, Cham, 2019), pp. 79–86.
- [29] D. W. McShea, Perspective metazoan complexity and evolution: Is there a trend? *Evolution* **50**, 477 (1996).
- [30] J. W. Valentine, A. G. Collins, and C. P. Meyer, Morphological complexity increase in metazoans, *Paleobiology* **20**, 131 (1994).
- [31] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (The Belknap Press of Harvard University Press, Cambridge, MA, 2006).
- [32] M. Harper, The replicator equation as an inference dynamic, *arXiv:0911.1763*.
- [33] R. A. Watson and E. Szathmáry, How can evolution learn, *Trends Ecol. Evol.* **31**, 147 (2016).
- [34] K. Kouvaris, J. Clune, L. Kounios, M. Brede, and R. A. Watson, How evolution learns to generalise: Using the principles of learning theory to understand the evolution of developmental organisation, *PLoS Comput. Biol.* **13**, e1005358 (2017).
- [35] D. Czégel, H. Giffar, J. B. Tenenbaum, and E. Szathmáry, Bayes and Darwin: How replicator populations implement Bayesian computations, *BioEssays* **44**, 2100255 (2022).
- [36] J. O. Campbell, Universal Darwinism as a process of Bayesian inference, *Front. Syst. Neurosci.* **10**, 49 (2016).
- [37] R. A. Watson, G. P. Wagner, M. Pavlicev, D. M. Weinreich, and R. Mills, The evolution of phenotypic correlations and developmental memory, *Evolution* **68**, 1124 (2014).
- [38] H. Akaike, A new look at the statistical model identification, *IEEE Trans. Autom. Control* **19**, 716 (1974).
- [39] *Model Selection and Multimodel Inference*, edited by K. P. Burnham and D. R. Anderson (Springer, New York, 2004).
- [40] V. N. Vapnik, *Nature of Statistical Learning Theory* (Springer, New York, New York, 1995).
- [41] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning* (Springer, New York, 2006), Vol. 4.
- [42] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, Cambridge, 2003).
- [43] D. H. Wolpert, A rigorous investigation of evidence and Occam factors in Bayesian reasoning, Technical Report No. 92-03-013, Sante Fe Institute, Santa Fe, NM, 1992.
- [44] B. Malnic, J. Hirono, T. Sato, and L. B. Buck, Combinatorial receptor codes for odors, *Cell* **96**, 713 (1999).
- [45] B. Xue, P. Sartori, and S. Leibler, Environment-to-phenotype mapping and adaptation strategies in varying environments, *Proc. Natl. Acad. Sci. USA* **116**, 13847 (2019).
- [46] R. A. Fisher, *The Genetical Theory of Natural Selection: A Complete Variorum Edition* (Oxford University Press, Oxford, 1999).
- [47] R. Bürger and M. Lynch, Evolution and extinction in a changing environment: A quantitative-genetic analysis, *Evolution* **49**, 151 (1995).
- [48] L.-M. Chevin, R. Lande, and G. M. Mace, Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory, *PLoS Biol.* **8**, e1000357 (2010).
- [49] V. Mustonen and M. Lässig, From fitness landscapes to seascapes: Non-equilibrium dynamics of selection and adaptation, *Trends Genet.* **25**, 111 (2009).
- [50] S. Dridi and L. Lehmann, Environmental complexity favors the evolution of learning, *Behav. Ecol.* **27**, 842 (2016).
- [51] P. Godfrey-Smith, Environmental complexity and the evolution of cognition, in *The Evolution of Intelligence* (Psychology Press, London, 2013), pp. 223–249.
- [52] N. Kashtan and U. Alon, Spontaneous evolution of modularity and network motifs, *Proc. Natl. Acad. Sci. USA* **102**, 13773 (2005).
- [53] R. Dawkins and J. R. Krebs, Arms races between and within species, *Proc. R. Soc. London, Ser. B* **205**, 489 (1979).
- [54] A. Bergman and M. W. Feldman, On the evolution of learning: Representation of a stochastic environment, *Theor. Pop. Biol.* **48**, 251 (1995).
- [55] A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov, Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*, *PLoS Genet.* **10**, e1004775 (2014).
- [56] H. Rapoport, Code and data for “Fitness and overfitness: Implicit regularization in evolutionary dynamics,” <https://github.com/nitzanlab/Overfitness> (2026).
- [57] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, New York, 2009), Vol. 2.
- [58] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Contr. Sign. Syst.* **2**, 303 (1989).