

The Boundedness Illusion: Asymptotic projections from early evolution underestimate evolutionary potential

Michael J Wiser^{1,2,3}, Emily L Dolson^{1,2,3}, Anya Vostinar⁴, Richard E Lenski^{1,2,3}, and Charles Ofria^{1,2,3}

¹Michigan State University, East Lansing, MI 48824

²BEACON Center, Michigan State University

³Ecology, Evolutionary Biology, and Behavior Program, Michigan State University

⁴Grinnell College, Grinnell, IA 50112

mwiser@msu.edu

Abstract

Open-ended evolution researchers seek to create systems that continually produce new evolutionary outcomes, attempting to reflect the power and diversity of evolution in nature. The specific metrics used (novelty, complexity, diversity, etc) vary by researcher, but the holy grail would be a system where any of these can accumulate indefinitely. Of course, one challenge that we face in reaching this goal is even recognizing if we have succeeded. To determine the evolutionary potential of a system, we must conduct finite experiments; based on their results we can predict how we would expect evolution to progress were it to continue. Here we examine how such predictions might be made and how accurate they might be. We focus on predicting fitness; this metric is often easy to calculate, and correlated with increases in traits like novelty and complexity. For each run in a simple digital evolution experiment, we find the best fit to measured values of fitness, and demonstrate that projecting this fit out usually predicts that fitness will be constrained by an asymptote. Upon extending the experiments, however, we see that fitness often far exceeds this asymptote, belying the boundedness that it implies. Extending past a premature end point allows us to see beyond this “boundedness illusion”.

Introduction

Evolution has produced an astounding degree of diversity in biological life. Current evidence suggests that this has been happening for approximately 3.8 billion years on Earth (Mojzsis et al., 1996; Rosing, 1999), and yet evolutionary innovation is far from over. New strains of respiratory diseases circulate yearly (Grenfell et al., 2004; Nelson et al., 2007). Populations of predators continuously race against prey in co-evolutionary cycles, where changes in one population alter selection pressures on the other, and those changes in the second population reflect back as new selection pressures on the first (Abrams, 2000). Even a long-term laboratory study of *E. coli* in a simple glucose-limited environment provides no evidence of evolution reaching an endpoint (Wiser et al., 2013). Natural systems are thus characterized as undergoing open-ended evolution, showing continual generation of change, novelty, complexity, and diversity (Taylor et al., 2016).

However, computational evolutionary systems often seem to show rapid change early followed by stagnation (Lampinen and Zelinka, 2000; Piotrowski, 2014). After a sufficient period of little or no change, researchers conclude that the system has reached an endpoint. Unfortunately, there are some potential pitfalls to this conclusion. For one, not all curves that decelerate are bounded. Power law relationships, which are common in both biology (Clauset et al., 2009; May, 2006) and physics (Adams et al., 1993; Newman, 2005), show rapid changes early on, followed by decelerating changes. But power law relationships with positive exponents are not asymptotic; there is no upper limit to the function the way there is with, say, a rectangular hyperbola. Secondly, and perhaps more importantly, experiments are, by nature, finite; they run for a specified period, and stop. As a field, we often use these experiments to make predictions about what will happen over much longer, sometimes indefinite, time scales (Sajjad et al., 2016), but these predictions are, by their very nature, extrapolations. Portions of a curve that look like stagnation when viewed at one scale can be part of a steep increase when viewed over a larger interval. Further, certain features of an evolutionary trajectory – for example, a rare event of large magnitude in the midst of more common events of relatively small magnitude, – may cause a trajectory to *appear* bounded, even when the general pattern is *unbounded*. We term this discrepancy the “boundedness illusion”.

Background

Open-ended Evolution

Is there a fundamental difference between evolutionary dynamics in populations of DNA-based organisms and evolutionary dynamics in today’s artificial life systems? If so, any disparity would be important to understand when drawing inferences from computational systems. Moreover, figuring out the cause of such a difference would provide insight into the conditions that are necessary or sufficient to produce evolution akin to what we observe in nature. These topics are the focus of research on open-ended evolution (Taylor et al., 2016).

In order to address these questions, we must first identify the dynamics that we expect an open-ended evolutionary system to exhibit. Previous work has identified several major categories of relevant dynamics (Taylor et al., 2016). “Unbounded growth in fitness” is not one of these categories. It is, however, interwoven with all of them. Over the time frames and population sizes commonly used in artificial life experiments, unbounded growth in fitness is sufficient to continuously produce change and novelty, as it requires at least an occasional introduction of new, fitter genotypes. Moreover, most ways of filtering for *meaningful* evolutionary dynamics rely in some way on fitness. Previously, we suggested an approach in which genomes are simplified down to sites that affect fitness (*i.e.* informative sites), and populations are filtered to include only those genotypes that survive for sufficiently long to be meaningful (Taylor et al., 2016). Unbounded growth in fitness should facilitate the evolution of new genotypes that pass these filters and are thus able to contribute to metrics of open-endedness.

Other approaches to quantifying open-ended evolution are also closely tied to fitness and boundedness. For example, Bedau et al.’s Evolutionary Activity Statistics classify systems into groups based on the long-term behavior of three measurements: diversity, novel evolutionary activity, and average evolutionary activity per taxonomic group (Bedau et al., 1998). In order for a system to be classified among the most open-ended, novel evolutionary activity must be positive and the other two metrics must be unbounded. While these metrics do not explicitly measure fitness, unbounded growth in fitness guarantees that novelty will be positive. In most cases, it will also result in mean evolutionary activity being unbounded. Thus, unbounded fitness growth dynamics have clear implications for a system’s open-endedness under this framework as well. Because of this relationship, we should be aware that the boundedness illusion may also apply to the metrics used in evolutionary activity statistics.

Boundedness

Many researchers see fitness in existing computational systems as inherently bounded (Taylor et al., 2016). In one sense, there must be a global fitness optimum within a computational system since there are a finite number of possible organisms. As such, every individual that could exist has an explicit fitness value for a given computational environment, with the highest such value being a fundamental bound. We could, in principle, rank all of the possible individuals. By necessity, one of these individuals will be the most fit, whether individually or tied with one or more others. As such, we would expect that as the population reached this global optimum, fitness gains would eventually stop.

However, the eventual theoretical bound on a population’s fitness may not translate into a practical bound. The potential search space in non-trivial computational evolutionary systems is vast. If the time it would take a system to reach

its upper boundary of fitness, on average, is orders of magnitude longer than the experiment, that boundary is likely to be irrelevant over experimental time scales. Shorter time scales – even those which are very long by experimental standards – may show unbounded increases in fitness, regardless of any theoretical upper bound in the system. Similarly, while the whole genomic search space of individuals may be technically ergodic – any state is theoretically reachable from any other state – how long it would take to transition from one particular state to another may make the space practically non-ergodic, even if it theoretically is so.

It is also important to note that certain features may impose theoretical boundaries in ways that do not have a practical effect. For example, a finite genome length inherently places an upper boundary on the complexity of an organism. However, if we repeat the experiment but allow the size of the genome to change and see no effect on the complexity of organisms evolved, this theoretical boundary is not exerting pressure on the evolutionary dynamics we observe. That doesn’t mean that this boundary could never have a practical effect, but the fact that a theoretical boundary *could* matter does not mean that it always *does*.

Fitness

Selection can act upon many different features to improve the performance of individuals in a population. Ultimately, these selected features influence Darwinian fitness, which is a measure of an individual’s genetic contribution to future generations. Indeed, biologists measure fitness in terms of contribution to future gene pools. This fitness can be measured directly through competitions over generations (Wiser and Lenski, 2015) or through traits that are components of fitness but easier to measure than integrated fitness in some systems. Examples of such fitness components include maximum growth rate in a population (Gerstein and Otto, 2011), number of young fledged (Velmalu et al., 2015), or seeds set (Remold, 2002). In any population subject to selection pressures stronger than drift, the expectation is that, on average, fitness will increase from one generation to the next (Orr, 2009). How fitness changes over time is tightly linked with whether the evolution in a system is open ended, or is itself bounded by an upper value. By focusing our analysis on fitness, we expect to address a fundamental driver of other aspects of open-ended evolution as well.

Adaptive Landscapes

To make sense of the vast genetic search space available in multi-objective optimization, researchers often turn to the concept of adaptive landscapes. Introduced by Sewall Wright (Wright, 1932), adaptive landscapes are a heuristic tool for conceptualizing the relationship between traits and fitness Fig. 1. Every point along the landscape represents a combination of trait values. When researchers assign a fitness value to each one of those points, a multidimensional

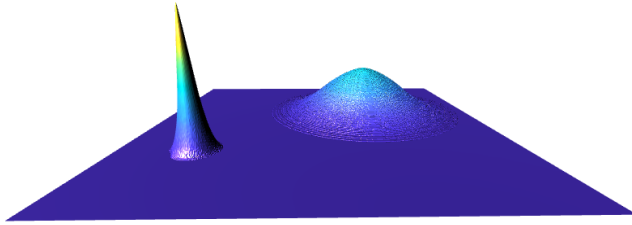


Figure 1: Example of a fitness landscape. This landscape consists of a narrow, high peak, and a broader, shorter peak. Which peak a population will settle on is partly a function of the mutation rate (Wilke et al., 2001)

landscape emerges, with peaks or plateaus of high fitness, and valleys or basins of low fitness. Of course, true fitness landscapes are unlikely to be only 3-dimensional; it is exceedingly rare for only two traits to influence fitness (though see (Ganco and Hoetker, 2009) about the use of NK-landscapes in research). Nevertheless, discussions of the adaptive landscape are pervasive in the literature in both evolutionary computation (Vassilev and Miller, 2000; Yu and Miller, 2001; Islam et al., 2012) and evolutionary biology (Page and Nowak, 2002; Mahler et al., 2013; Martin and Wainwright, 2013; de Visser and Krug, 2014), with a great deal of attention paid to how populations can move from a local optimum to a better optimum despite a region of poor fitness in between (Covert et al., 2013; Chou et al., 2014; Kvitek and Sherlock, 2011).

Wright elaborated different landscapes based on either phenotype or genotype, with the ones based on genotype more relevant to computational work. Because there are many sites in a genome, and each site has one or more possible mutational changes, genotypic landscapes are inherently highly multidimensional. The landscapes associated with non-trivial problems will themselves be complex, and thus have far more than three dimensions. For reasons directly analogous to how improvements in fitness may be theoretically bounded but practically unbounded, so too may a fitness landscape be treated as infinite even when it is theoretically finite.

Unlike in biological systems, it is often plausible in a digital system to quickly gather information about the local adaptive landscape by systematically measuring the impact of all possible one- or two-step mutations from a particular organism (Covert et al., 2013; Lenski et al., 2003). Yet even this local knowledge of the landscape does not necessarily provide information about the landscape as a whole; different regions of the landscape may have different properties. From an artificial life perspective, then, it is less the case that we can use the adaptive landscape to make predictions about

how evolution will proceed than that we can use information about how evolution has proceeded – and, in particular, from different evolutionary runs starting at the same point – to infer the shape of the adaptive landscape.

Natural systems rarely, if ever, exist in perfectly static environments. Changes in the environment may be abiotic, such as temperature, or biotic, such as changes in predator, prey, or competing populations. Changes may even be intrinsic to a population, such as when the selective value of a trait depends on that trait's frequency. All of these changes lead to dynamic landscapes, which are beyond the scope of this paper.

Methods

Study System

Avida is a digital evolution software platform (Ofria and Wilke, 2004; Ofria et al., 2009). In this software, organisms are individual programs, written in an assembly-like Turing-complete language. These organisms reproduce themselves by execution of their instructions. This reproduction, however, is not perfect; the user defines rates of mutations, such as instructions being replaced by other instructions, new instructions being added, or existing instructions being deleted. These mutations produce variation within the population of organisms. The user also creates an environment for these organisms, defining whether certain behaviors are rewarded with additional CPU cycles, and what the value of this reward is. Organisms have associated fitness values, which correspond to their expected rate of reproduction. With mutations as a source of variation, heredity due to the self-replication of organisms, and selection imposed by the environment, Avida represents an instance of evolution, rather than a mere simulation of it (Pennock, 2007).

Experimental Design

We evolved ten populations of digital organisms in an environment called logic-77, in which organisms were rewarded for performing any of 77 distinct 1-, 2-, or 3-input logic tasks (excluding the simplest 1-input task, Echo, where organisms output the same number they received as input). Organisms could perform any combination of these tasks, and be rewarded for repeating any individual task up to 10 times. We chose the logic-77 environment for two main reasons. First, the logic-77 environment is relatively complex, and we therefore expect there to be many different adaptive pathways available. Secondly, in earlier work we have found evolution in the logic-77 environment to show substantial variation across replicate runs (Wiser, 2015), which is an important element for our questions. In the first phase, we allowed these initial ten populations to evolve for 200,000 generations. In the second phase, we extracted the most common genotype in each population, and allowed it to evolve for another 200,000 generations to examine long-term dynamics.

For this second phase of evolution, we ran ten replicate instances of Avida from each of the ten replicates of the first phase to assess the variability in potential outcomes. In both phases, we recorded data every 20 generations

Statistical Methods

We fit our data to two distinct mathematical models of fitness change over time. One of the models is a rectangular hyperbola, of the form

$$w = \frac{a * t}{t + b} + 1 \quad (1)$$

The other model is a power law, of the form

$$w = (b * t + 1)^a \quad (2)$$

In both cases, w is relative fitness (measured fitness divided by ancestral fitness), a and b are model parameters, and t is time, measured here in generations. We fit these models with the `nls()` command in R version 3.5.1 (R Core Team, 2018). Before fitting models, we $\log(\text{base } 2)$ -transformed the fitness data. We compared our model fits by the difference in BIC value; note that because both of our models have the same number of parameters, and are fit on the same underlying data, the difference in BIC value is formally equivalent to the difference in AIC value. We used these differences in BIC values to compare model fits, as outlined in (Raftery, 1995).

Results and Discussion

All evolutionary runs start from the same initial ancestor in phase 1. From this starting point, differences accrue, as different runs will experience different random events (birth locations, deaths, mutations, etc.). Sometimes, this variation has no effect on the phenotype of either the organism or the population; at other times, it does. Fig. 2 shows that ten replicate trials evolving from the same ancestor reach ten different fitness values after 200,000 generations, and they do so through different trajectories.

For each replicate, we considered two different models to explain the fitness trajectory. One of them is a bounded model (Equation 1); in it, fitness will approach, though never quite reach, an upper limit. In this case, as t approaches infinity, the equation approaches a . The other model (Equation 2) is unbounded; in it, the rate of fitness increase decelerates, but fitness does not have an upper limit. Both of these models have two parameters, so are of the same complexity. The specific models considered are taken from a similar analysis of bacterial data (Wiser et al., 2013). The model fits for this dataset are presented in Table 1. We are not claiming that these specific models are the best fit of all possible models, but merely that they represent simple examples of a fundamental difference between a model that is bounded and a model that is not.

Seed	BIC - Unbounded	BIC - Bounded	Difference
1001	62532	54223	8310
1002	64908	47684	17224
1003	57051	42340	14711
1004	74087	67613	6474
1005	73477	67538	5939
1006	69025	64012	5013
1007	64835	49042	15792
1008	48405	36917	11488
1009	60901	57751	3150
1010	57096	46967	10129

Table 1: Model comparison between the bounded and unbounded models for 200,000 generations of evolution from the original ancestor. The Difference column is the BIC value for the unbounded model minus the BIC value for the bounded model. Positive values indicate a better fit for the bounded model, while negative values indicate a better fit for the unbounded model. BIC differences >10 are considered very strong support (Raftery, 1995). Note that values reported are rounded to the nearest integer, leading to some rounding differences in the last column

In all ten cases, the data are much better fit by the bounded model than by the unbounded model. At first glance, this result would seem to imply that fitness is approaching a maximum value that it cannot exceed. However, that logically cannot be the case. Each organism in each of the populations is descended from the same original ancestor through an unbroken chain of viable organisms. As such, there is a mutational path from any organism in one population to any organism in another population that passes entirely through organisms capable of survival and reproduction. This means that, in theory, any organism has the potential to mutate into any other organism in this experiment in a stepwise fashion. The probability of many of these paths will be extremely low, as they go through extended sections of poor fitness relative to the existing population, but they are still potential, viable paths. As such, any population which appears as if it has an asymptote to fitness lower than the highest population fitness observed must necessarily be capable of reaching the same fitness as that highest observed population, and thus cannot truly be bounded by that asymptote.

We also directly tested whether predictions of upper bounds of fitness were accurate by taking advantage of a convenient aspect of digital evolution. We took organisms from the end of the initial 200,000 generations and used them as the ancestors for a subsequent 200,000 generations of evolution (Fig. 3). For each of the ten intermediate ancestors, we conducted ten replicate evolutionary runs, for a total of 100 new runs. If the predictions of an asymptote from the first phase were accurate, we would expect none of the sub-

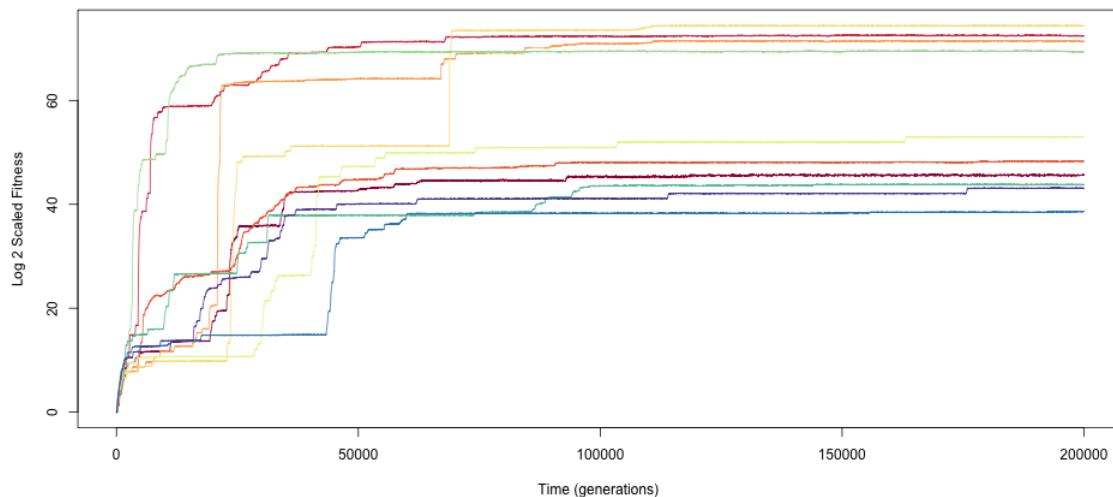


Figure 2: Evolution from the original ancestor. Each line represents one independent evolutionary replicate.

sequent runs to exceed the corresponding asymptote. This is not the case. Table 2 shows the predicted asymptote for each initial run, and the highest fitness attained in the subsequent evolution. In each case, the bounded model predicted lower long-term fitness than the unbounded model did. In 17 of 100 cases, the predicted maximum fitness was exceeded by merely doubling the number of generations.

1 st Seed	Pred. Asymptote	Highest Fit	# Exceed
1001	52.45	46.39	0
1002	75.73	74.80	0
1003	53.21	59.04	3
1004	82.56	75.26	0
1005	97.02	74.47	0
1006	68.55	70.37	2
1007	72.04	74.52	2
1008	47.22	56.23	4
1009	48.46	72.96	6
1010	47.98	45.18	0

Table 2: Comparisons of predicted asymptotes with highest realized fitnesses. All values have been log(base 2) transformed. Highest Fit denotes the highest fitness measured in the subsequent 200,000 generations from an ancestor corresponding to an individual from the end of the first 200,000 generations of evolution. # Exceed indicates how many of the ten replicates started from that ancestor exceeded this predicted asymptote in the second 200,000 generations of evolution.

The results stemming from these intermediate ancestors fall into several major categories. Broadly speaking, the cat-

egories are cases where 1) the replicates consistently demonstrate lower fitness than bounded model predicts; 2) the empirical data as a whole are more consistent with the bounded model, but individual replicates are highly variable; 3) a substantial fraction of the replicates exceed even the unbounded model's predictions. The third category is the least common of these – only one of the ten initial populations falls into this category – but the data are evenly split between categories 1 (five replicates) and 2 (four replicates).

For replicates in category 1, such as in Fig. 4A, even the bounded model from the first 200,000 generations overestimates the fitness trajectory in the second 200,000 generations. In cases like this, the actual populations exhibit fewer substantial increases in fitness in this second phase, or these increases are smaller than expected, or both. The lower or less frequent increases can sometimes be explained directly by the increases themselves being small or rare in the second phase. At other times, the increases in the second phase are substantial, but particularly large or early gains in the first phase led to an *expectation* of large or frequent (or both) increases, beyond what the subsequent data exhibit. Particularly in the latter cases, later portions of the evolutionary trajectory appear to better fit the bounded model than the unbounded model, even though the fitness of populations is increasing notably. These increases are, however, less than predicted from the large, early gains, and thus can give the illusion of boundedness.

In other cases, such as in Fig. 4B, the set of the subsequent runs as a whole are better fit by the bounded model, but many individual runs are not. Instead, some of the subsequent runs achieve notably higher fitness than the bounded model predicts, while others underperform compared to

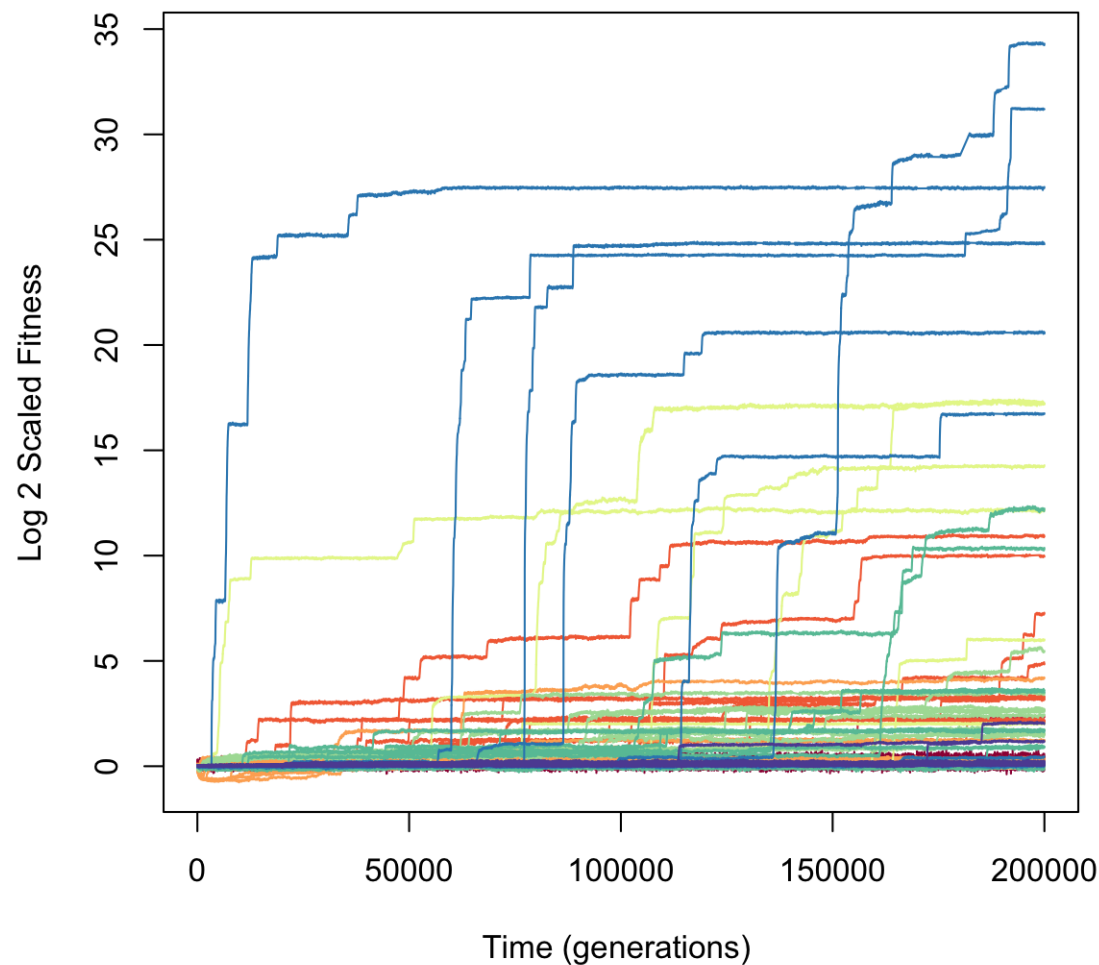


Figure 3: Evolution from intermediate ancestors. Each line represents one evolutionary replicate. Different colors represent different intermediate ancestors. The colors match those from Fig. 2; for example, the run from 2 that produced the red line there is used as the ancestor for all the lines with the same shade of red in this figure. Note that in this figure, each ancestor is scaled to its own starting point. Because both this and Fig. 2 have Log 2 y axes, this scaling moves each trajectory down vertically to start at 0 in this figure; the magnitude of increase is unaffected

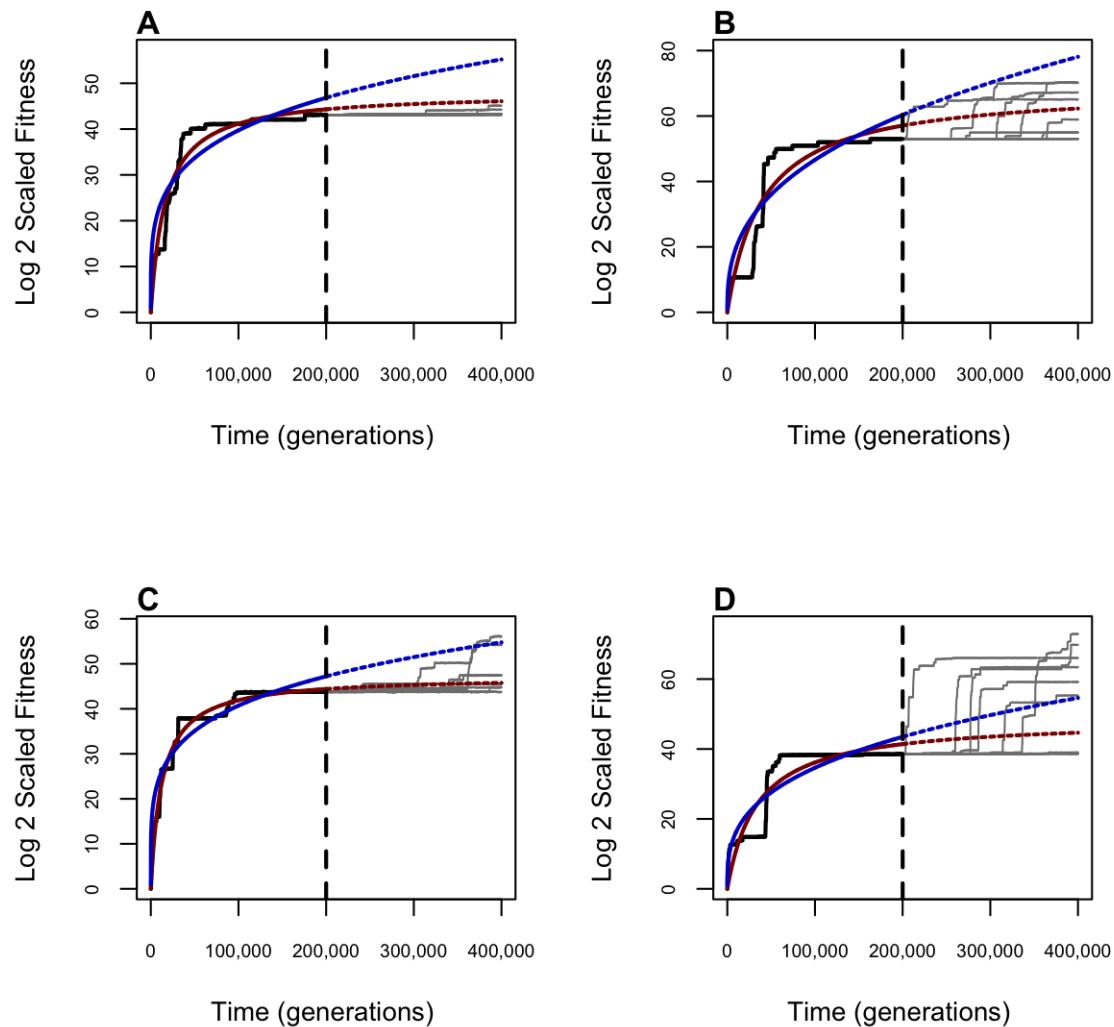


Figure 4: Example comparisons of model projections from the first evolutionary phase to empirical data. For each case, the bounded model is shown in dark red; the unbounded model is shown in medium blue. The models are fit through the first 200,000 generation (solid lines), and then projected for the next 200,000 generations (dashed lines). The solid black line shows the empirical data for the first phase. The thin gray lines show the ten replicates of the second phase, each started from the end of the first phase. Note that the different panels have different y-axes, as the different initial populations reach notably different fitness levels.

model expectations. Indeed, in some cases, such as in Fig. 4C, some of the individual replicates actually better fit with the unbounded model, even occasionally exceeding this model's predictions. This category, then, reflects circumstances where the average across all replicates from this intermediate ancestor appears to be better described by a bounded model, but individual replicates show very different results.

In the final category, shown in Fig. 4D, not only do many replicates from the intermediate ancestor exceed the theoretical asymptote from the bounded model, they even exceed the predictions of the particular unbounded model used. This case shown is a good example of how happenstance can play a large role in the accuracy of long-term predictions. The initial run leading to the intermediate ancestor had only modest changes in fitness between 80,000 and 200,000 generations, leading to the appearance of a population having reached a fitness plateau. Yet in only a short time after the start of evolution from the intermediate ancestor, one of the replicates experienced a rapid growth in fitness, rising from a $\log_2$ fitness of slightly less than 40, to one greater than 60 in a few, rapid increases. Had this rapid increase happened slightly before the end of the initial phase, rather than slightly after the start of the second phase, it would have had a profound impact on both the predictions made and the subsequent analysis.

Across the whole set of experiments, many of the subsequent evolutionary runs – 83 of 100 – remain below the theoretical asymptote predicted by the initial 200,000 generations of evolution. At first glance, this result might seem to support that these asymptote predictions are largely correct. That, however, would be a mistake. Firstly, these predictions are of the maximum fitness the relevant populations could attain; an example of exceeding this limit is sufficient to prove that the theoretical limit is incorrect. Secondly, these asymptotes are predictions of the maximum fitness these populations would ever be able to attain, given infinite time, if the pattern of fitness change in the first 200,000 generations continued. Instead, we are testing these only against an additional 200,000 generations. This is a large number of generations, to be sure, but it is a trivial number compared to the history of microbes on Earth. Thirdly, we have examples of populations exceeding these theoretical maximum from five of the ten intermediate ancestors examined. This result is not just a case of a small percentage of predicted limits being routinely exceeded, but instead a case where fully half of the predicted limits were exceeded at least once – in a set of only ten replicates each running, as mentioned, for just twice the initial period.

What features of these fitness trajectories may drive the boundedness illusion? One striking feature is the large step-wise increases in fitness available in this environment. Any time a new task evolves, this has the potential to immediately double fitness. Moreover, because tasks can be rewarded up

to ten times per organism, a task performed inside a loop could be rewarded up to 2^{10} -fold. These large increases in fitness will occur instantly for the individual experiencing the mutation, and rapidly for the population, as the large selective coefficient will lead to a rapid selective sweep. These very large rewards will be far more visible in the fitness trajectory than the more frequent but smaller improvements that come from more efficient replication, which can lead to the *appearance* of fitness being constant, even while it is increasing, albeit much more slowly. For example, in Fig. 4D – the most extreme case of subsequent evolution exceeding the theoretical asymptote of the bounded model – one of the ten replicate extensions underwent a series of rapid increases in fitness almost immediately, resulting in a more than 2^{20} gain in fitness in this second phase of evolution. Yet even so, it was eventually superseded by several other replicates from this same intermediate ancestor that took slower paths toward even higher fitness regions of the landscape. For any given replicate, there are points at which it seems that the population has plateaued – this ancestral population spent more than 100,000 generations at a nearly constant fitness – but these times of apparent stasis do not signify a lack of future change.

Further, the scale of other changes can obscure real evolutionary improvements. In Fig. 4D, four replicates started from the intermediate ancestor look as if they have flat trajectories in the second phase. In reality, these four range from one that declines slightly ($\downarrow 2\%$), to one that increases $\downarrow 35\%$ over the second 200,000 generations. Changes of this magnitude are important in biological systems – populations in the aforementioned *E. coli* study increased in fitness roughly 60-70% over 50,000 generations of adaptation (Wiser et al., 2013) – but are invisible in this graph because they are dwarfed by the relatively rare but exceptionally large effects of evolving new functions that are highly rewarded.

We expect that what we have termed the “boundedness illusion” exists in almost any open-ended evolving system. Even in a system that can always generate more novelty or complexity, there will likely be long periods without obvious signs of novelty generation or increases in complexity. As such, sampling within a particular time frame can make the novelty or complexity of a system appear bounded even when it is not. As a simple example, if one were to measure the complexity of life on Earth, but only using the hundreds of millions of years in which unicellular life existed and multicellular life did not, the complexity of life would appear to be strictly limited. Yet that limit would be illusory, as later changes opened vast new regions of the search space to exploration, leading to massive growth in complexity and entirely novel niches being created.

Conclusion

Contrary to the views of many in the Artificial Life community, we present both logical argument and empirical data that imply Artificial Life systems such as Avida can exhibit certain open-ended dynamics. Features such as rare, large effect mutations masking the impact of more frequent, smaller effect mutations can lead to the appearance of boundedness that is not borne out by subsequent evolution.

For systems that lack feedback cycles and frequency-dependence, a global optimal genotype must exist. That optimum, however, may be so far from an arbitrary starting position that an upper bound does not have an appreciable impact on evolutionary dynamics within the system. Deceleration in the rate of fitness gain can give a strong impression of fitness being bounded by an upper limit, lending statistical support to models which have an asymptotic limit. However, subsequent evolution frequently exceeds these predicted limits, thereby calling into question the utility of such models. Further, the variation in predicted limits across different replicates starting from the same point logically implies that, at most, only one of these limits could be real. Instead, we should be careful to provide evidence of effective limits to evolution, rather than assume they are present without adequately demonstrating so. Even better, we suggest researchers can extend their experiments – either in time or space – and thus see beyond the “boundedness illusion”.

Acknowledgments

We thank Caroline Turner, Alita Burmeister, and Michael Blazanin for helpful discussion and feedback on this manuscript. This research was supported in part by the BEACON Center for the Study of Evolution in Action (NSF Cooperative Agreement No. DBI-0939454), NSF Award No. DEB-1655715, and Michigan State University through the computational resources provided by the Institute for Cyber-Enabled Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Abrams, P. A. (2000). The evolution of predator-prey interactions: Theory and evidence. *Annual Review of Ecology and Systematics*, 31(1):79–105.

Adams, F. C., Bond, J. R., Freese, K., Frieman, J. A., and Olinto, A. V. (1993). Natural inflation: Particle physics models, power-law spectra for large-scale structure, and constraints from the Cosmic Background Explorer. *Physical Review D*, 47(2):426–455.

Bedau, M. A., Snyder, E., and Packard, N. H. (1998). A classification of long-term evolutionary dynamics. In *Artificial life VI*, pages 228–237. Cambridge: MIT Press.

Chou, H.-H., Delaney, N. F., Draghi, J. A., and Marx, C. J. (2014). Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLOS Genetics*, 10(2):e1004149.

Clauset, A., Shalizi, C., and Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

Covert, A. W., Lenski, R. E., Wilke, C. O., and Ofria, C. (2013). Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *Proceedings of the National Academy of Sciences*, 110(34):E3171–E3178.

de Visser, J. A. G. and Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, 15:480.

Ganco, M. and Hoetker, G. (2009). NK modeling methodology in the strategy literature: Bounded search on a rugged landscape. In *Research Methodology in Strategy and Management*, volume 5 of *Research Methodology in Strategy and Management*, pages 237–268. Emerald Group Publishing Limited.

Gerstein, A. C. and Otto, S. P. (2011). Cryptic fitness advantage: Diploids invade haploid populations despite lacking any apparent advantage as measured by standard fitness assays. *PLOS ONE*, 6(12):e26599.

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., and Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303(5656):327.

Islam, S. M., Das, S., Ghosh, S., Roy, S., and Suganthan, P. N. (2012). An adaptive differential evolution algorithm with novel mutation and crossover strategies for global numerical optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2):482–500.

Kvitek, D. J. and Sherlock, G. (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLOS Genetics*, 7(4):e1002056.

Lampinen, J. and Zelinka, I. (2000). On stagnation of the differential evolution algorithm. In Osmara, P., editor, *Proceedings of MENDEL 2000*, pages 76–83, Brno, Czech Republic.

Lenski, R. E., Ofria, C., Pennock, R. T., and Adami, C. (2003). The evolutionary origin of complex features. *Nature*, 423(6936):139–144.

Mahler, D. L., Ingram, T., Revell, L. J., and Losos, J. B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*, 341(6143):292.

Martin, C. H. and Wainwright, P. C. (2013). Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. *Science*, 339(6116):208.

May, R. M. (2006). Network structure and the biology of populations. *Twenty years of TREE - part 2*, 21(7):394–399.

Mojzsis, S. J., Arrhenius, G., McKeegan, K. D., Harrison, T. M., Nutman, A. P., and Friend, C. R. L. (1996). Evidence for life on Earth before 3,800 million years ago. *Nature*, 384:55.

- Nelson, M. I., Simonsen, L., Viboud, C., Miller, M. A., and Holmes, E. C. (2007). Phylogenetic analysis reveals the global migration of seasonal Influenza A viruses. *PLOS Pathogens*, 3(9):e131.
- Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323–351.
- Ofria, C., Bryson, D. M., and Wilke, C. O. (2009). Avida. In *Artificial Life Models in Software*, pages 3–35. Springer London.
- Ofria, C. and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial Life*, 10(2):191–229.
- Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, 10:531.
- Page, K. M. and Nowak, M. A. (2002). Unifying evolutionary dynamics. *Journal of Theoretical Biology*, 219(1):93–98.
- Pennock, R. T. (2007). Models, simulations, instantiations, and evidence: the case of digital evolution. *Journal of Experimental & Theoretical Artificial Intelligence*, 19(1):29–42.
- Piotrowski, A. P. (2014). Differential evolution algorithms applied to neural network training suffer from stagnation. *Applied Soft Computing*, 21:382–406.
- R Core Team (2018). R: A language and environment for statistical computing.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, 25:111–164.
- Remold, S. K. (2002). Unapparent virus infection and host fitness in three weedy grass species. *Journal of Ecology*, 90(6):967–977.
- Rosing, M. T. (1999). ¹³C-Depleted carbon microparticles in >3700-Ma sea-floor sedimentary rocks from west Greenland. *Science*, 283(5402):674.
- Sajjad, M., Singh, K., Paik, E., and Ahn, C.-W. (2016). A data-driven approach for agent-based modeling: simulating the dynamics of family formation. *Journal of Artificial Societies and Social Simulation*, 19(1):9.
- Taylor, T., Bedau, M., Ackley, D., Channon, A., Banzhaf, W., Beslon, G., Dolson, E., Froese, T., Hickinbotham, S., Ikegami, T., McMullin, B., Packard, N., Rasmussen, S., Virgo, N., Agmon, E., Clark, E., McGregor, S., Ofria, C., Roppella, G., Spector, L., Stanley, K. O., Stanton, A., Timperley, C., Vostinar, A., and Wiser, M. (2016). Open-Ended Evolution: Perspectives from the OEE workshop in York. *Artificial Life*, 22(3):408–423.
- Vassilev, V. K. and Miller, J. F. (2000). The advantages of landscape neutrality in digital circuit evolution. In Miller, J., Thompson, A., Thomson, P., and Fogarty, T. C., editors, *Evolvable Systems: From Biology to Hardware*, pages 252–263. Springer Berlin Heidelberg.
- Velmal, W., Helle, S., Ahola, M. P., Klaassen, M., Lehtikainen, E., Rainio, K., Sirki, P. M., and Laaksonen, T. (2015). Natural selection for earlier male arrival to breeding grounds through direct and indirect effects in a migratory songbird. *Ecology and Evolution*, 5(6):1205–1213.
- Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E., and Adami, C. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412(6844):331–333.
- Wiser, M. J. (2015). *An analysis of fitness in long-term asexual evolution experiments*. PhD thesis, Michigan State University.
- Wiser, M. J. and Lenski, R. E. (2015). A comparison of methods to measure fitness in *Escherichia coli*. *PLOS ONE*, 10(5):e0126210.
- Wiser, M. J., Ribeck, N., and Lenski, R. E. (2013). Long-term dynamics of adaptation in asexual populations. *Science*, 342(6164):1364–1367.
- Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the 6th International Congress of Genetics*, pages 356–366.
- Yu, T. and Miller, J. (2001). Neutrality and the evolvability of Boolean function landscape. In Miller, J., Tomassini, M., Lanzi, P. L., Ryan, C., Tettamanzi, A. G. B., and Langdon, W. B., editors, *Genetic Programming*, pages 204–217. Springer Berlin Heidelberg.