

AN EMPIRICAL ANALYSIS OF HISTORICAL CONTINGENCY IN EVOLUTION

By

Austin James Ferguson

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Computer Science - Doctor of Philosophy  
Ecology, Evolution, and Behavior - Dual Major

2024

## ABSTRACT

While evolution has created a stunning diversity of complex traits in nature, isolating the details for *how* a particular trait evolved remains challenging. Specifically, what were the critical events in evolutionary history that made that trait more or less likely to arise? We must consider historical contingency, where even small changes, such as an apparently neutral mutation, can have substantial influence on long-term evolutionary outcomes. Evolutionary biologists have long been interested in the role that historical contingency plays in evolution, but testing hypotheses of its effects has traditionally been difficult and time consuming, if it is even possible at all.

Here I leverage the speed and power of digital evolution to experimentally test the role of historical contingency in evolution. I start by observing how the evolution of phenotypic plasticity stabilizes future evolutionary dynamics. Next, I employ analytic replay experiments to empirically test which mutations in a population's history increased the likelihood that associative learning evolves, first as case studies and then using more statistically powerful experimental approaches. I demonstrate that single mutations can drastically increase the odds of learning appearing, shifting it from a rare possibility to a near inevitability, and I find that these "potentiating" mutations exist in all studied lineages. Finally, I use potentiating mutations to develop an intuitive view into how adaptive momentum increases evolutionary exploration in populations experiencing disequilibrium.

We are only beginning to scratch the surface of how historical contingency influences evolution, but digital evolution systems can expedite this process by testing these hypotheses and further refining these techniques for use in natural organisms. This work, and those like it, are pivotal in understanding how populations previously evolved, how their accumulated history currently affects them, and how they might evolve far into the future.

Copyright by  
AUSTIN JAMES FERGUSON  
2024

In memory of my grandmother, Mammaw Debbie.  
Oh, how I wish you were here to celebrate.

## ACKNOWLEDGEMENTS

When writing a dissertation about historical contingency, it is impossible to avoid thinking about the events in my own life, both big and small, that have brought me to this point. Among all of the decisions, failures, and chance events that potentiated this path, the most important factors are the plethora of amazing people whose paths have crossed my own. Here I wish to thank them. Fully encapsulating my gratitude within this text is impossible, but I will try my best.

To those who seemingly never tire: Barbara Bloemers, Vinny Mattison, Dr. Elise Zipkin, Dr. Sandeep Kulkarni, and Dr. Eric Torng, you each deserve sainthood for your demonstrations of patience and tenacity in pushing for us graduate students.

To those most likely to read this dissertation: Dr. Chris Adami, Dr. Wolfgang Banzhaf, and Dr. Rob Pennock, your patience and guidance have helped shape me into a scientist.

To those who led the way: Dr. Emily Dolson, Dr. Josh Nahum, Dr. Alex Lalejini, Alexa Lalejini, Dr. Matthew Moreno, Dr. Jose Hernandez, Dr. Anselmo Pontes, Dr. Anya Vostinar, Dr. Mike Wiser, Dr. Vinny Ragusa, Dr. Jory Schossau, Santiago Rodriguez Papa, Keron Greene, and Dr. Nkrumah Grant, you welcomed me with open arms and offered friendship, guidance, and much needed reassurance every step of the way.

To those who kept me sane: Dr. Acacia Ackles and Kate Skocelas, without the two of you I would not have made it through the isolation of a global pandemic.

To those who carried me through the finish: Sydney Leither, Max Foreback, Shakiba Shahbandegan, and Karen Suzue, the countless nights of laughter have made these last couple years some of my best. Your unwavering support and wisdom beyond your years keep me in grateful awe.

To the one who is uncategorizable: Cliff Bohm, who always has an idea to debate, a paper to collaborate on, and a warm dinner for good company.

To that good company: Chris Reeves, Kira Chan, Lucky Voutour, Ariel Cooper, and the rest of the crew, your comradery was always a highlight of my week.

To those who made Monday mornings tolerable: the EEB coworking crew, there is no one I would rather work in silence around. You are truly the most supportive group of people I have ever met.

To those who let me learn how to teach: my undergraduate mentees Daniel Junghans, Nick Lloyd, Caroline Gormerly, Dylan Rainbow, Lanea Rohan, Aria Killebrew-Bruehl, Valdine Peggy

Tchinda, and dozen of students across classes and workshops, thank you for your patience and in teaching me more than I taught you.

To those who are always there: Dr. Ronnie Emmons, Logan Boggs, Dariel Marine, Dr. Gabe Howard, and Jason Montavon, your companionship and endless laughs are truly appreciated.

To the one who pushed me: Andrew Polanco, you single-handedly convinced me to aim for what I really wanted and refused to let me falter.

To those who demonstrated the role: my instructors and colleagues at Shawnee, it was your passion for educating and your kind hearts that led me to believe I could maybe pull this off.

To the friends who have truly become family: Alex Bellomy, who continuously raises the bar for what it means to be someone's best friend; and Ashlyn Bellomy, who has demonstrated that true friends do not require a lifetime of history.

To the one who took a chance on me: Dr. Charles Ofria, my advisor who foolishly accepted a clueless new graduate student six years ago. Through a wild journey, you have shaped me into a scientist and educator, never showing the slightest hesitation to help me or just anyone around you. While I am sad to see this journey end, I am so happy to call you a friend at the end of it.

To those that gave me unwavering support: Martha, Jim, and Marty Ferguson and Jeff Adkins, you have never forgotten to say "I'm proud of you" when I needed it most.

To the one taken too soon: Debbie Adkins, you taught me the most important lessons: not to take life too seriously and to always love with all your heart.

To the one who changed my life: Ethan Ferguson, your big heart and determination continuously fill me with pride. You may be the younger brother, but I am the one learning from you.

To the ones who made it possible: Donnie and Alisha Ferguson, you have always pushed to give me the opportunities you did not have. You taught me the true value in "we are here to support you". You also taught me how to work hard, laugh harder, and make the most of life. Everything I have done and will do in this life are thanks to the sacrifices you have made.

To those mentioned and those not: from the bottom of my heart, thank you.

## TABLE OF CONTENTS

Chapter 1	Introduction . . . . .	1
Chapter 2	Adaptive phenotypic plasticity stabilizes evolution in fluctuating environments . . . . .	12
Chapter 3	Potentiating mutations facilitate the evolution of associative learning in digital organisms . . . . .	40
Chapter 4	A large-scale digital evolution study of historical contingencies in associative learning . . . . .	56
Chapter 5	Predicting the unpredictable: Using replay experiments to disentangle how evolutionary outcomes are altered by adaptive momentum . . . . .	77
Chapter 6	Conclusions and outlook . . . . .	96
	BIBLIOGRAPHY . . . . .	108
	APPENDIX . . . . .	118

# Chapter 1

## Introduction

I feel obliged to start this dissertation with the standard observation: there is seemingly limitless diversity to the organisms that have evolved in nature. From microbes to megafauna, evolutionary processes have created a stunning array of traits and behaviors in organisms. Were each of these features mere evolutionary flukes, or was their eventual evolution inevitable? Are there at least general evolutionary trends that we can abstract from these examples? Questions like these, of course, are grand challenges of evolutionary biology. Here, I focus on one such question: how do the events of the past influence future evolution?

When looking at evolution in nature, we often encounter beneficial traits that were uniquely evolved in one type of organism, while organisms in the broader taxonomic unit found different survival strategies. For example, most birds and many insects exhibit self-powered flight, but only one branch of mammals does: bats (Gunnell and Simmons, 2005). Similarly, while many snakes have both venom and the means to deliver it via a bite, venomous bites are only found in one clade of lizards (Fry et al., 2006). Why are these adaptations so rare, and what conditions led to their evolution in only a single clade each? In nature, there is only so much we can do; unfortunately we cannot travel back in time and “replay the tape of life” (as evoked by Gould (1990)) to observe if flight consistently evolved in bats thanks to that mammalian clade’s history, or if it was an uncommon stroke of luck.

While replaying the tape of life *as we know it* is impossible, experimental evolution studies allow us to ask similar questions. As I describe below, evolutionary biologists have successfully employed experimental approaches to better understand the role that accumulated history has played in the evolution of microbial traits. In this work I continue this endeavor; I use *in silico* evolution experiments to study the role of history in evolving traits, both simple and complex, in digital organisms. I first ask how the evolution of one trait, a phenotypically plastic response to the environment, affects downstream evolutionary dynamics. I then pivot, switching from observations of how a known trait altered evolution, to retroactively identifying mutations that drastically shifted the long-term fate of the population. Finally, I demonstrate how we can leverage these techniques to better understand fundamental evolutionary dynamics. Though still in their infancy, these techniques and ideas show amazing potential for understanding the nuanced role that history plays in evolution.

## 1.1 Analyzing evolution: selection, chance, and history

In thinking about the evolution of a particular trait, I adopt a mindset modified from Travisano et al. (1995) and consider the contributions of three different factors: selection, chance, and history. New or modified traits that provide a net benefit to the survival and reproduction of an organism (i.e., adaptations) are more likely to increase in frequency, thus illustrating the role that *selection* plays in shaping evolutionary outcomes. Here I split from Travisano et al. (1995), focusing on selection as a process, which results in these adaptations becoming more prevalent in the population. Of course, selection can only act upon genetic sequences that are available in the population. The stochastic nature of random mutations means that some genetic sequences will arise, while others will never even appear for selection to consider in the first place. The random appearance of sequences – or disappearance as misfortune can remove otherwise fit genotypes (i.e., genetic drift) – highlights some of the roles of *chance* in evolution. The influence of chance is constrained, though, by the starting genotypes that mutations act upon; traits can appear or be modified only if such changes are available in the local genetic neighborhood. This distribution of genetic sequences in the population at the point under investigation in an evolutionary study can be described as the product of its *history*. It is the interplay of these three aspects that produce the complex dynamics that we observe in evolving populations.

Of course, what we call history is just a matter of temporal perspective. From the vantage point of a population in a given state, all of the dynamics, including selection and chance, that brought the population to that state are now consolidated under the label of history. Looking forward, however, both selection and chance will be at play, with different mutations or combinations of mutations occurring at different probabilities, and the resulting combinations having differing survival potential. As time advances, these changes are again relegated to history, while new outlooks now exist for the population based on its new composition. As such, for any study where we examine the balance among these three factors, we need to be clear about the starting point from which our perspective will be based.

In this work I focus on the role of history in evolution, which is to say that I investigate how the balance of chance and selection shifts over the course of the population's history. Since evolution is a stochastic process, we must always discuss possible evolutionary outcomes in terms

of probabilities. From one point along a lineage, the evolution of a given trait may be incredibly unlikely, dependent on a rare double mutation or a deleterious mutation that is not immediately purified. From a later point along that lineage, however, that same trait's evolution may shift to being a near certainty, with strong selective pressure for that trait, or its preceding parts, more in control of the population's fate. Is there some way for us to predict these shifts in influence between selection and chance? Can we distinguish between a population that is being driven to a specific outcome from one that is simply adrift? And how do these shifts occur? Does chance give way to selection in small increments, or can an individual mutation dramatically alter the balance?

## 1.2 The role of history and historical contingency in evolution

While the ideas of evolution and natural selection have been around for well over one hundred and fifty years (Darwin, 1859), evolutionary biologists continue to argue about, test, and expand upon the different factors that contribute to evolution. While selection was the initial frontrunner, researchers argued for the importance of chance (Kimura, 1968; King and Jukes, 1969; Mayr, 1983) and later history (Gould, 1990; Gould and Lewontin, 1979) in evolution.

It may appear obvious that history plays an important role in evolution, as the set of genetic sequences that could feasibly appear in the population relies on what sequences currently exist. Here, however, I mainly focus on the idea of "historical contingency" – the idea that small, often initially inconsequential changes can have a drastic effect on what ultimately evolves. As an idealized example, consider a set of three mutations, A, B, and C, that together give rise to a highly beneficial trait. All three mutations are equally beneficial in isolation, while AB is slightly more beneficial but AC and BC are both detrimental to fitness. In this scenario, populations that have fixed either A or B in isolation have a beneficial pathway to the combined trait, ABC. If instead a population has fixed C by itself, the ABC trait becomes much harder to evolve, as both intermediate steps are deleterious; a double mutation is then needed to reach the trait without losing fitness. Even in this simple example, the initial fixing of C has no penalty when it occurs, but it shifts the possibilities of what is likely to evolve in the future. For a thorough review of the ideas and complications of historical contingency, as well as empirical investigations into its role in evolution, see (Blount et al., 2018).

While work has been done to study the role of historical contingency in the evolution of natural populations (e.g., (Keller and Taylor, 2008; Losos et al., 1998; Oke et al., 2019)), here I base my work

on empirical studies of historical contingency in experimental evolution. Early work in *Escherichia coli* produced two drastically different results. Researchers found no influence of the initial value of fitness reflected in the populations' final evolved fitness, but in the same experiment, they found that the final evolved cell size of a population was highly contingent in the initial cell size of that replicate (Travisano et al., 1995). Building off this framework, Flores-Moya et al. (2012) found evidence that history plays a key role in the evolution of growth rate and toxin cell quota in algae. Santos-Lopez et al. (2021) found that even shallow differences in history can influence the evolution of antimicrobial resistance. Recently, Smith et al. (2022) have shown that, while history does play a role in the evolution of *E. coli*, the interactions between selection, chance, and history can heavily depend on traits under investigation and the environment being studied. By leveraging clever experimental evolution studies, these researchers have shown that it is possible to disentangle the influence of selection, chance, and history on evolution in a particular system.

Further work has investigated historical contingency in evolution without explicitly quantifying the contributions of these three basal factors. Many studies can be categorized under this label, but here I highlight a few representatives. Multiple teams have investigated how historical contingency often creates genotypic variance while selection is able to derive similar phenotypic outcomes (Bedhomme et al., 2013; Simões et al., 2017). In antibiotic resistance, historical contingency can alter both evolvability and the particular genetic pathways evolution takes (Card et al., 2019, 2021). While historical contingency can be caused by any genetic change, recent work has highlighted contingency caused by cryptic genetic variation (Zheng et al., 2019) and synonymous mutations (McGrath et al., 2024).

There is also a long and rich history of using digital systems for studying historical contingency's role in evolution (Taylor and Hallam, 1998). We can even consider studies that were not initially thought of as investigations of historical contingency, such as how to choose the starting population in evolutionary computation (Burke et al., 1998; Schmidt and Lipson, 2009). Specifically analyzing contingency, the seminal work of Travisano et al. (1995) on the roles of selection, chance, and history over time was also replicated and expanded in the digital evolution system Avida, with far more statistical power thanks to the perfect record keeping of digital systems (Wagenaar and Adami, 2004). Braught and Dean (2007) again expanded this work by using neural networks and demonstrated an interaction between selection, chance, history, and the Baldwin effect; they found

that learning can influence the impact of the three factors. By comparing normal evolutionary replicates to those where deleterious mutations were automatically reverted, Covert III et al. (2013) found that initially-deleterious mutations can increase the probability of complex traits evolving. Yedid et al. (2008) used a controlled extinction event and found that the pre-extinction presence of a complex trait factored into the re-evolution of the trait after the extinction event. More recently, Bundy et al. (2021) leveraged the speed of digital evolution to test how the depth of history affects future evolution, dealing with generation counts well beyond what is currently feasible in microbial systems. These studies show that not only are these techniques viable in digital systems, but that digital systems can *expand upon them* to conduct research that would otherwise be impossible.

### 1.3 Genetic potentiation

While this dissertation investigates the role of history in evolution, most chapters narrow that focus to the “potentiation” of a focal trait. The term “potentiation” has been used in many different ways, but here I define potentiation as the likelihood that the focal trait evolves from a given initial genotype or population (adapted from Blount et al. (2008)). If a change (e.g., a mutation or an environmental shift) increases the likelihood that the focal trait evolves, that is a *potentiating* change, as indicated by the increase in the trait’s potentiation value. Here my emphasis is on genetic potentiation, where changes in potentiation are due solely to accumulated genetic changes in the organisms, as opposed to environmental or anatomical potentiation (Heyduk et al., 2019). Critically, the idea of potentiation is to separate the *actualizing* mutations that confer the focal trait from the *potentiating* mutations that made the evolution of that trait more likely. Indeed, a potentiating mutation may have no (or even negative) consequences on fitness in the short term, until additional mutations build upon it.

Early studies of potentiation often follow a variation of the question: If mutation X arises, how does that affect the evolution of trait Y? As an example, we can consider the work by Gifford et al. (2018). The authors identify mutations in the transcription factor *ampR* as present among strains of *Pseudomonas* capable of quickly evolving resistance to the antibiotic ceftazidime. Importantly, the authors show that mutations to *ampR* alone do not constitute antibiotic resistance; they only prime the evolution of antibiotic resistance. That is, the mutations in *ampR* *potentiated* the evolution of antibiotic resistance.

While it is easy to conflate the notion of potentiating mutations with positive (or synergistic)

epistatic interactions, it should be noted that the two are technically distinct ideas. Here we consider epistasis as originally outlined by Fisher (1919), where epistatic interactions occur when the appearance of two genes create a trait value that differs from the predicted combination of their effects in isolation (typically with an additive or multiplicative expectation). Epistatic interactions can be potentiating, and in fact, many examples of potentiating mutations in the literature demonstrate epistasis between the potentiating mutation and the focal trait (Douglas et al., 2017; Gifford et al., 2018). This relationship, however, does not need to be direct; we can imagine scenarios where the potentiating mutation is a “gateway” to the desired region of the fitness landscape, or where the potentiating mutation epistatically interacts with a third mutation that then increases the fitness of the actualizing step. To consider a mutation potentiating, it only needs to improve the likelihood that the focal trait evolves.

How do we find these potentiating mutations? A common approach is to isolate lineages that have evolved the focal trait from an outgroup that have not (either via experimental evolution or collection from the wild) and then analyze genomes from both groups. Mutations commonly found in the actualized group but rarely in the outgroup are likely to be vital to the evolution of the focal trait, either potentiating or actualizing mutations, and further experiments can disentangle the relationship of the mutations (Cummins et al., 2021; Gifford et al., 2018).

Here, however, I employ *replay experiments*, an empirical technique to retroactively identify potentiating mutations within a given evolved lineage. In the pioneering work on replay experiments, Blount et al. (2008) empirically tested whether the novel citrate metabolism in one of Richard Lenski’s Long-Term Evolution Experiment (LTEE) populations (Lenski et al., 1991) was due to a fluke mutation or the accumulation of a potentiated genetic background. To do so, they revived frozen samples from the LTEE to found multiple “replay” populations from various points along the lineage that originally evolved to metabolize citrate. Conceptually, if the fraction of replay populations that evolved citrate metabolism increased between generation  $N$  and generation  $N + K$ , we know that one or more potentiating mutations occurred in those  $K$  generations. Indeed, this result is what they found; the metabolization of citrate was more likely to evolve from samples further along the lineage, providing support that genetic potentiation was a key factor.

Essentially, this framework is applying the analysis of Travisano et al. (1995) along steps of a lineage. By starting evolutionary replicates from various points in the population’s history, we are

experimentally manipulating the shared history among replicates. Differences in what evolves in these replicates can then identify changes that occurred in the population's history that shifted the contributions of selection and chance in the evolution of the focal trait. Increases in potentiation indicate an increased contribution of selection, as the trait is now more likely to evolve. Such an increase could happen if the population has moved to a position in genetic space where a more adaptive (or less maladaptive) pathway to the focal trait now exists. Decreases in potentiation, on the other hand, indicate a stronger reliance on chance and could, for example, be the result of convergence to a local optimum in the fitness landscape. The selective pressure of this optimum could leave the population reliant on fluke mutations or genetic drift to escape and potentially find the focal trait. Ultimately, a difference in potentiation between two points on a lineage indicates that the genetic changes between them, which are considered history in the context of the later point, are important in whether the focal trait ultimately evolves. Identifying a substantial change in potentiation from one time point to another opens the possibility of examining what mutations fall in this window, their immediate effect on the organism, and if their appearance in the lineage was due to selection or chance. Additionally, the size of the replay window is arbitrary; we can skip thousands of generations between checks for potentiating events or, in the extreme case, test every single generation.

Since the initial study (Blount et al., 2008), researchers have conducted similar experiments (now called analytic replay experiments, (Blount et al., 2018)) in multiple systems and looking at various traits. These include clade extinction and evolvability in *E. coli* (Turner et al., 2015; Woods et al., 2011), novel receptor usage in Phage  $\lambda$  (Gupta et al., 2022; Meyer et al., 2012), colistin resistance in *Pseudomonas aeruginosa* (Jochumsen et al., 2016), and, recently, the epistatic interactions in yeast (Vignogna et al., 2021). Across these systems, the researchers showed that the accumulated genetic background is profoundly important in the eventual evolution of the focal trait. While these techniques are relatively new, they offer valuable insight into how the interplay of selection, chance, and history can influence subsequent evolution. These studies look backward, empirically testing what changes led to the final evolved behaviors, but the resulting potentiation dynamics are deeply intertwined with concepts such as predictability in evolution. As such, I expect research in the near future to begin weaving these findings into the broader tapestry of evolutionary dynamics.

## 1.4 Why digital evolution?

Researchers addressing these questions about the role of history in evolution initially started where many evolutionary biology questions started: in the fossil record. Indeed, some of the most influential thoughts on the role of history come from fossils in the Burgess Shale (Gould, 1990). More recent techniques allow researchers to supplement these ideas with genomic analysis and phylogenetic reconstruction, powerful tools for determining the history of life on earth. Looking to history, however, has many limitations: we are provided with only a single instance of evolution, the data we do have is incomplete, and we are not able to go back in time to conduct controlled experiments. In recent decades there has been a surge in experimental approaches to studying evolution (Kawecki et al., 2012). If we evolve populations in controlled laboratory conditions, we are able to evolve many populations in parallel, observe almost everything that occurs, and build a range of experimental conditions – thus addressing each of the problems above. Evolutionary biologists have leveraged experimental evolution to empirically test a multitude of hypotheses, such as possible selection pressures for the transition to multicellularity (Ratcliff et al., 2012) and the feedback cycle between phage- and antibiotic- resistance and its clinical implications (Chan et al., 2016; Kortright et al., 2019). I adopt this experimental mindset throughout this dissertation in an attempt to understand the role of historical contingency in evolution. I will use the possibilities revealed from counterfactual experiments to understand “life as it could have been” in our study systems. These techniques will help me separate the flukes from the inevitabilities in the dynamics that shaped the course of evolution as it was originally realized in these studies.

However, even the control and speed of *in vitro* experimental evolution is sometimes not enough. Specifically, we must be able to isolate all of the individual mutations along a lineage and replay evolution using each step as a new starting point. Furthermore, for each of these starting points, we need to be able to conduct enough replays to generate statistically powerful conclusions about evolutionary outcomes. While experimental evolution techniques are powerful (and getting more powerful with time), performing experiments at this scale would still take decades of dedicated work in the laboratory.

To overcome these hurdles, I leverage the power of computational evolution systems. These systems maintain populations of self-replicating computer programs that mutate and evolve in the

same way as natural organisms. In fact, compelling arguments have been made that digital evolution is not a simulation of evolution, but rather an instantiation of it (Pennock, 2007). These digital evolution systems are not a departure from experimental evolution, but rather a choice of experimental system. These computational systems give many benefits over *in vitro* approaches, including greater speed, automatic high resolution data collection, and the ability to start an experiment with the exact conditions of our choosing. Of course, digital evolution also has its drawbacks. For example, there is a much lower limit to the complexity of the organisms and what we have been able to evolve *in silico*, and open-ended evolution remains elusive (Taylor et al., 2016). Further, due to technological constraints, digital evolution is typically limited to scales of at most tens of thousands of organisms, while natural populations can be vastly larger (Moreno et al., 2024).

As a digital evolution researcher, I am often asked how my computational models tell us anything about how life evolved in the natural world. While other computational researchers may ask such questions, I do not claim to *directly* explain life *as it was or will be*. Instead, my questions are theoretical; studying evolution as a process using *life as it could be*. My goal is to conduct experiments that are infeasible or impossible to conduct in laboratory studies of evolution to further refine our understanding of basic evolutionary dynamics. For example, we can show evidence of the minimal set of requirements for the evolution of a particular trait or behavior (Lalejini and Ofria, 2016; Pontes et al., 2020). We can also boil questions down to their most basic scenarios, removing all possible confounding factors in a way that is not possible in natural organisms and allowing for strong claims about the impacts of the remaining dynamics (Bohm et al., 2024; Wilke et al., 2001). Finally, in the case of history in evolution, we can show examples of how these dynamics unfold, though we cannot claim that is how these dynamics will *always* unfold. By improving our theoretical understanding and fine-tuning experimental methodologies, we enable wet-lab or natural-system evolutionary biologists to better understand their study system, and, more generally, the evolution of life on earth.

## 1.5 Thesis Statement

The core thesis of this dissertation is two-part: 1) advancement of evolutionary theory requires us to understand how past events result in particular evolutionary outcomes, and 2) computational systems make large-scale replay experiments feasible and allow us to empirically test the role that history played in evolved populations. This rapid experimentation in computational systems allows

us to both improve our theoretical understanding of the role of historical contingency in evolution and to refine the methodologies of these experiments, which can then be leveraged in future wet-lab studies.

## 1.6 Contributions

Overall, this dissertation advances our knowledge of the role that history plays in evolution. This is not done by inventing new methodologies nor discovering new phenomena. Instead, the main contributions of this work lie in the intersections; by applying existing techniques to new areas. By leveraging the power of digital evolution, I am able to experimentally test ideas of historical contingency at scales that were previously impossible.

Moving to specifics, in **Chapter 2** I ask a simple question: What happens *after* phenotypic plasticity evolves? Here, phenotypic plasticity, where different phenotypes can arise from a single genotype due to environmental variation, is studied because it is a precursor to learning (a theme in later chapters, and what initially brought me to study digital evolution). However, this question is often difficult; while we can readily find species capable of phenotypic plasticity, what would we compare them to? I thus use digital tools to evolve both plastic and non-plastic populations, and then test how their evolutionary dynamics change as a consequence. Across all experiments, I find that the evolution of phenotypic plasticity stabilizes future evolution in a cyclic environment that consistently alternates between two states. Plastic populations experience less evolutionary change, maintain more novel beneficial traits, and accumulate fewer deleterious traits than their non-plastic counterparts. Indeed, evidence suggests that the evolutionary dynamics of plastic populations more closely align to those of populations evolving in a static environment than non-plastic populations in the same environment.

For the rest of this dissertation, my focus shifts from effect to cause; from identifying the consequences that result from a known change in history to retroactively identifying the changes that shifted the overall evolutionary fate. This starts with **Chapter 3**, where I conduct case studies to find and characterize “potentiating mutations” in the evolution of associative learning. By conducting analytic replay experiments (Blount et al., 2018), I identify a single parent-to-offspring reproduction event that substantially increased the likelihood that learning ultimately evolved in each of four case study lineages. This initial foray demonstrates that even single mutations can drastically alter the fate of an evolving population.

While the results of Chapter 3 enhance our understanding of potentiation and potentiating mutations, the low count of case studies prevents us from being able to make any statistical claims. Thus we are left with the question: do *most* of the populations that evolve learning in Avida do so via large potentiating mutations, or were those four replicates unlikely? To answer this question, I replay 50 additional learning replicates in **Chapter 4**. I find that the four original lineages were not flukes. All 50 lineages not only see a significant increase in potentiation across 50 lineages steps, they all see a significant increase in a *single* step. Additionally, we see that most of these potentiating lineage steps either occur with a single mutation, or only one mutation had a significant effect on potentiation. This work is by far the largest study of potentiation to date, and it allows for comparison against both previous studies and those in the future.

Finally, **Chapter 5** takes a different approach. I leverage analytic replay experiments again, but instead of studying potentiation, I use it as a tool for viewing a newly identified evolutionary dynamic: adaptive momentum (Bohm et al., 2024), which broadly states that populations in disequilibrium see more evolutionary exploration. While adaptive momentum has been shown in the aggregate, here I demonstrate the effect within the history of an evolving population by analyzing the potentiation of crossing a fitness valley. I show that when a spatial population is experiencing a selective sweep (an example of disequilibrium), mutations near the leading edge of the sweep have a drastic effect on the overall evolution of the population. This contrasts with populations in equilibrium, which effectively rely on pure chance to cross the fitness valley. Further, I demonstrate an example where the structure of the population matters, as the potentiation of the population diminishes if we shuffle the order of the organisms.

## Chapter 2

# Adaptive phenotypic plasticity stabilizes evolution in fluctuating environments

Authors: Alexander Lalejini, Austin J. Ferguson, Nkrumah A. Grant, and Charles Ofria

This chapter is adapted from (Lalejini et al., 2021), which was peer-reviewed and published in *Frontiers in Ecology and Evolution*.

In this work we use digital evolution to empirically test how the evolution of phenotypic plasticity alters future evolutionary dynamics. In all three cases we tested, we find that phenotypic plasticity stabilizes these dynamics. Specifically, compared to non-plastic populations, plasticity 1) slows evolutionary change, 2) maintains the rate of discovering novel beneficial traits while decreasing the probability they are lost, and 3) reduces the accumulation of deleterious traits. In the language of the later chapters, we can say that phenotypic plasticity in this system *potentiates* the evolution of the novel beneficial traits and decreases the probability of accumulating deleterious traits.

## 2.1 Introduction

Natural organisms employ a wide range of evolved strategies for coping with environmental change, such as periodic migration (Winger et al., 2019), bet-hedging (Beaumont et al., 2009), adaptive tracking (Barrett and Schluter, 2008), and phenotypic plasticity (Ghalambor et al., 2007). The particular mechanisms that evolve in response to fluctuating environments will also shift the course of subsequent evolution (Schaum and Collins, 2014; Wennersten and Forsman, 2012). As such, if we are to understand or predict evolutionary outcomes, we must be able to identify which mechanisms are most likely to evolve and what constraints and opportunities they impart on subsequent evolution.

In this work, we focus on phenotypic plasticity, which can be defined as the capacity for a single genotype to alter phenotypic expression in response to a change in its environment (West-Eberhard, 2003). Phenotypic plasticity is controlled by genes whose expression is coupled to one or more environmental signals, which may be either biotic or abiotic. For example, the sex ratio of the crustacean *Gammarus duebeni* is modulated by changes in photoperiod and temperature (Dunn et al., 2005), and the reproductive output of some invertebrate species is heightened when infected with parasites to compensate for offspring loss (Chadwick and Little, 2005).

Evolutionary biologists have long been interested in how evolutionary change is influenced by phenotypic plasticity because of its role in generating phenotypic variance (Gibert et al., 2019). The effects of phenotypic plasticity on adaptive evolution have been disputed, as few studies have been able to observe both the initial patterns of plasticity and the subsequent divergence of traits in natural populations (Forsman, 2015; Ghalambor et al., 2015, 2007; Hendry, 2016; Wund, 2012). In changing environments, adaptive phenotypic plasticity provides a mechanism for organisms to regulate trait expression within their lifetime, which can stabilize populations through those changes (Gibert et al., 2019). In this context, the stabilizing effect of adaptive plasticity has been hypothesized to constrain the rate of adaptive evolution (Ancel, 2000; Gupta and Lewontin, 1982; Huey et al., 2003; Paenke et al., 2007; Price et al., 2003). That is, directional selection may be weak if environmentally-induced phenotypes are close to the optimum; as such, adaptively plastic populations may evolve slowly (relative to non-plastic populations) unless there is a substantial fitness cost to plasticity.

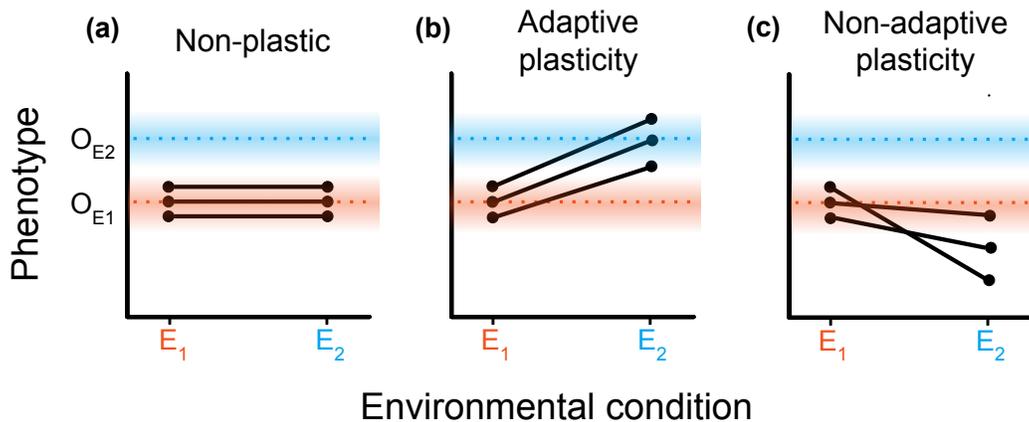


Figure 2.1: Hypothetical reaction norms for populations comprising genotypes placed in different environments. A reaction norm describes phenotypic change (or lack thereof) induced by environmental variation (West-Eberhard, 2008). In all panels, two environmental conditions (denoted  $E_1$  and  $E_2$ ) are shown on the x-axis. The y-axis indicates the phenotype expressed in each environment with  $O_{E_1}$  and  $O_{E_2}$  designating the optimal phenotype for  $E_1$  and  $E_2$ , respectively. Each pair of points connected by a solid black line denotes a genotype, with the points themselves representing its hypothetical phenotypes in each environment. We present three scenarios for how populations could respond to a change from  $E_1$  to  $E_2$ . (a) A non-plastic population where phenotypes do not change with environmental shifts. (b) An adaptively plastic population where phenotypes dynamically adjust to the new optimum. (c) A population exhibiting non-adaptive plasticity where environmental change induces phenotypes further away from the optimum.

Phenotypic plasticity allows for the accumulation of genetic variation in genomic regions that

are unexpressed under current environmental conditions. Such cryptic (“hidden”) genetic variation can serve as a source of diversity in the population, upon which selection can act when the environment changes (Levis and Pfennig, 2016; Schlichting, 2008). It remains unclear to what extent and under what circumstances this cryptic variation caches adaptive potential or merely accumulates deleterious alleles (Gibson and Dworkin, 2004; Paaby and Rockman, 2014; Zheng et al., 2019).

The “genes as followers” hypothesis (also known as the “plasticity first” hypothesis) predicts that phenotypic plasticity may facilitate adaptive evolutionary change by producing variants with enhanced fitness under stressful or novel conditions (Levis and Pfennig, 2016; Schwander and Leimar, 2011; West-Eberhard, 2003). Environmentally-induced trait changes can be refined through selection over time (*i.e.*, genetic accommodation). Further, selection may drive plastic phenotypes to lose their environmental dependence over time in a process known as genetic assimilation (Crispo, 2007; Levis and Pfennig, 2016; Pigliucci, 2006; Schlichting and Wund, 2014; West-Eberhard, 2005). In this way, environmentally-induced phenotypic changes can precede an evolutionary response.

Phenotypic plasticity may also “rescue” populations from extinction under changing environmental conditions by buffering populations against novel stressors. This buffer promotes stability and persistence and grants populations time to further adapt to rapidly changing environmental conditions (Chevin and Lande, 2010; West-Eberhard, 2003).

Disparate predictions about how phenotypic plasticity may shift the course of subsequent evolution are not necessarily mutually exclusive. Genetic and environmental contexts determine if, and to what extent, phenotypic plasticity promotes or constrains subsequent evolution. Figure 2.1 overviews how we might expect different forms of phenotypic plasticity to result in different evolutionary responses after an environmental change. In Figure 2.1a, we would expect non-plastic populations to experience strong directional selection toward the new optimum ( $O_{E2}$ ) after the environment changes. We would expect an adaptively plastic population (Figure 2.1b) to remain relatively stable after the environment changes, as plasticity shifts organisms’ phenotypes to the new optimum. In Figure 2.1c, we would expect the non-adaptively plastic population to experience strong directional selection on their response to the new environmental conditions; indeed, such maladaptive plasticity may even put the population at risk of extinction in the absence of beneficial mutations.

Experimental studies investigating the relationship between phenotypic plasticity and evolu-

tionary outcomes can be challenging to conduct in natural systems. Such experiments would require the ability to irreversibly toggle plasticity followed by long periods of evolution during which detailed phenotypic data would need to be collected. Digital evolution experiments have emerged as a powerful research framework from which evolution can be studied. In digital evolution, self-replicating computer programs (digital organisms) compete for resources, mutate, and evolve following Darwinian dynamics (Wilke and Adami, 2002). Digital evolution studies balance the speed and transparency of mathematical and computational simulations with the open-ended realism of laboratory experiments. Modern computers allow us to observe many generations of digital evolution at tractable time scales; thousands of generations can take mere minutes as opposed to months, years, or millennia. Digital evolution systems also allow for perfect, non-invasive data tracking. Such transparency permits the tracking of complete evolutionary histories within an experiment, which circumvents the historical problem of drawing evolutionary inferences using incomplete records (from frozen samples or fossils) and extant genetic sequences. Additionally, digital evolution systems allow for experimental manipulations and analyses that go beyond what is possible in wet-lab experiments. Such analyses have included exhaustive knockouts of every site in a genome to identify the functionality of each (Lenski et al., 2003), comprehensive characterization of local mutational landscapes (Canino-Koning et al., 2019; Lenski et al., 1999), and the real-time reversion of all deleterious mutations as they occur to isolate their long-term effects on evolutionary outcomes (Covert III et al., 2013). Furthermore, digital evolution studies allow us to directly toggle the possibility for adaptive plastic responses to evolve, which enables us to empirically test hypotheses that were previously relegated to theoretical analyses.

In this study, we conduct digital evolution experiments to investigate how the evolution of adaptive phenotypic plasticity shifts the course of evolution in a cyclically changing environment. We use the Avida Digital Evolution Platform (Ofria et al., 2009). Avida is an open-source system that has been used to conduct a wide range of well-regarded studies on evolutionary dynamics, including the origins of complex features (Lenski et al., 2003), the survival of the flattest effect (Wilke et al., 2001), and the origins of reproductive division of labor (Goldsby et al., 2014). Our experiments build directly on previous studies in Avida that characterized the *de novo* evolution of adaptive phenotypic plasticity (Clune et al., 2007; Lalejini and Ofria, 2016) as well as previous work investigating the evolutionary consequences of fluctuating environments for populations of

non-plastic digital organisms (Canino-Koning et al., 2019; Li and Wilke, 2004). Of particular relevance, Clune et al. (2007) and Lalejini and Ofria (2016) experimentally demonstrated that adaptive phenotypic plasticity can evolve given the following four conditions (as described by Ghalambor et al. 2010): (1) populations experience temporal environmental variation, (2) these environments are differentiable by reliable cues, (3) each environment favors different phenotypic traits, and (4) no single phenotype exhibits high fitness across all environments. We build on this previous work, but we shift our focus from the evolutionary causes of adaptive phenotypic plasticity to investigate its evolutionary consequences in a fluctuating environment. Specifically, we examine the effects of adaptive plasticity on subsequent genomic and phenotypic change, the capacity to evolve and then maintain novel traits, and the accumulation of deleterious alleles.

Each of our experiments are divided into two phases: in phase one, we precondition sets of founder organisms with differing plastic or non-plastic adaptations; in phase two, we examine the subsequent evolution of populations founded with organisms from phase one under specific environmental conditions (Figure 2.2). First, we examine the evolutionary histories of phase two populations to test whether adaptive plasticity constrained subsequent genomic and phenotypic changes. Next, we evaluate how adaptive plasticity influences how well populations produced by each type of founder can evolve and retain novel adaptive traits. Finally, we examine lineages to determine whether adaptive plasticity facilitated the accumulation of cryptic genetic variation that would prove deleterious when the environment changed.

We found that the evolution of adaptive plasticity reduced subsequent rates of evolutionary change in a cyclic environment. The non-plastic populations underwent more frequent selective sweeps and accumulated many more genetic changes over time, as non-plastic populations relied on genetic variation from *de novo* mutations to continuously readapt to environmental changes. The evolution of adaptive phenotypic plasticity buffered populations against environmental fluctuations, whereas repeated selective sweeps in non-plastic populations drove the accumulation of deleterious mutations and the loss of secondary beneficial traits. As such, adaptively plastic populations were better able to retain novel traits than their non-plastic counterparts. In general, the evolution of adaptive phenotypic plasticity shifted evolutionary dynamics to be more similar to that of populations evolving in a static environment than to non-plastic populations evolving in an identical fluctuating environment.

## 2.2 Materials and Methods

### 2.2.1 The Avida Digital Evolution Platform

Avida is a study system wherein self-replicating computer programs (digital organisms) compete for space on a finite toroidal grid (Ofria et al., 2009). Each digital organism is defined by a linear sequence of program instructions (its genome) and a set of virtual hardware components used to interpret and express those instructions. Genomes are expressed sequentially except when the execution of one instruction (*e.g.*, a “jump” instruction) deterministically changes which instruction should be executed next. Genomes are built using an instruction set that is both robust (*i.e.*, any ordering of instructions is syntactically valid, though not necessarily meaningful) and Turing Complete (*i.e.*, able to represent any computable function, though not necessarily in an efficient manner). The instruction set includes operations for basic computations, flow control (*e.g.*, conditional logic and looping), input, output, and self-replication.

Organisms in Avida reproduce asexually by copying their genome instruction-by-instruction and then dividing. However, copy operations are imperfect and can result in single-instruction substitution mutations in an offspring’s genome. For this work, we configured copy operations to err at a rate of one expected mutation for every 400 instructions copied (*i.e.*, a per-instruction error rate of 0.0025). We held individual genomes at a fixed length of 100 instructions (the default genome size in Avida); that is, we did not include insertion and deletion mutations. We used fixed-length genomes to control for treatment-specific conditions resulting in the evolution of substantially different genome sizes (Lalejini and Ferguson, 2021b)<sup>1</sup>, which could, on its own, drive differences in evolutionary outcomes among experimental treatments. When an organism divides in Avida, its offspring is placed in a random location on the toroidal grid, replacing any previous occupant. For this work, we used the default 60 by 60 grid size, which limits the maximum population size to 3600 organisms. As such, improvements to the speed of self-replication are advantageous in the competition for space.

During evolution, organism replication rates improve in two ways: by improving genome efficiency (*e.g.*, using a more compact encoding) or by accelerating the rate at which the genome is expressed (their “metabolic rate”). An organism’s metabolic rate determines the speed at which it

---

<sup>1</sup>We repeated our experiments without genome size restrictions and observed qualitatively similar results (see supplemental material, Lalejini and Ferguson 2021b).

executes instructions in its genome. Initially, an organism’s metabolic rate is proportional to the length of its genome, but that rate is adjusted as it completes designated functions, such as performing Boolean logic functions (Ofria et al., 2009). In this way, we can reward or punish particular phenotypic traits (*i.e.*, Boolean logic functions).

### Phenotypic plasticity in Avida

In this work, we measure a digital organism’s phenotype as the set of Boolean logic functions that it performs (*i.e.* expresses) in a given environment. Sensory instructions in the Avida instruction set allow organisms to detect how performing a particular function would affect their metabolic rate (see supplemental material for more details, Lalejini and Ferguson 2021b). We define a phenotypically plastic organism as one that uses sensory information to alter which functions it performs based on the environment.

Phenotypic plasticity in Avida can be adaptive or non-adaptive for a given set of environments. Adaptive plasticity shifts net function expression closer to the optimum for the given environments. Non-adaptive plasticity changes function expression in either a neutral or deleterious way. In this work, optimal plasticity toggles functions to always perfectly match the set of rewarded functions for the given set of environments.

#### 2.2.2 Experimental design

We conducted three independent experiments using Avida to investigate how the evolution of adaptive plasticity influences evolutionary outcomes in fluctuating environments. For each experiment, we compared the evolutionary outcomes of populations evolved under three treatments (Figure 2.2): (1) a **PLASTIC** treatment where the environment fluctuates, and digital organisms can use sensory instructions to differentiate between environmental states; (2) a **NON-PLASTIC** treatment with identical environment fluctuations, but where sensory instructions are disabled; and (3) a **STATIC** control where organisms evolve in a constant environment.

Each experiment was divided into two phases that each lasted for 200,000 updates<sup>2</sup> of evolution (Figure 2.2), which is equivalent to approximately 30,000 to 40,000 generations. In phase one of each experiment, we evolved plastic, non-plastic, and control organisms for use in phase two. In phase two, we founded new populations with these evolved organisms and examined their subsequent

---

<sup>2</sup>One update in Avida is the amount of time required for the average organism to execute 30 instructions. See (Ofria et al., 2009) for more details.

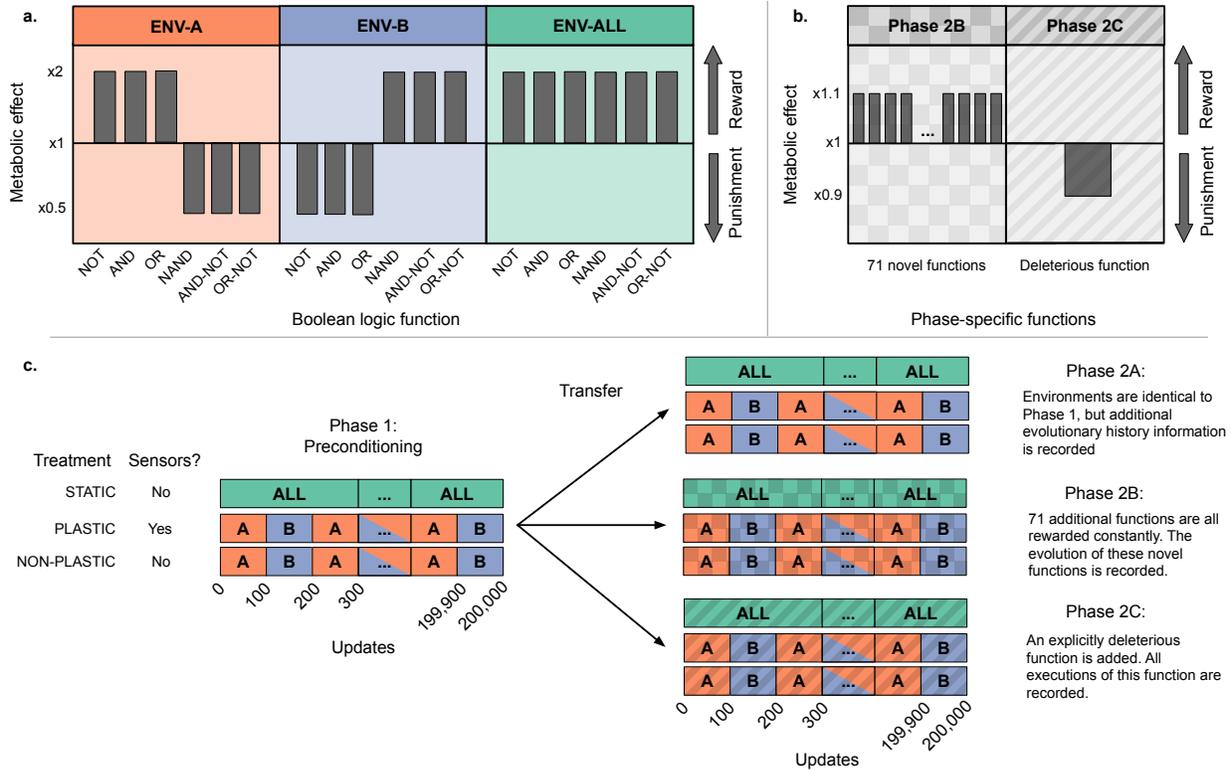


Figure 2.2: The three plots in panel (a) show the environmental conditions used in every experiment and whether they reward or punish each base function. The two plots in (b) show the additional functions added in phases 2B and 2C. Panel (c) shows treatment differences and experimental phases. Treatments are listed on the left, with each treatment specifying its environmental configuration and whether sensors are functional. We conducted three independent two-phase experiments, each described on the right. Phases 2B and 2C are textured to match their function definitions in panel (b). Phase one is repeated for *each* experiment with 100 replicate populations per treatment per experiment. For each replicate at the end of phase one, we used an organism of the most abundant genotype to found the second phase population. All STATIC and NON-PLASTIC populations move on to phase two, but PLASTIC populations only continue to the second phase if their most abundant genotype exhibits optimal plasticity. Metrics are recorded only in phase two.

evolution under given combinations of treatment and experimental conditions. During phase two, we tracked and saved each population’s evolutionary history as well as saving the full final population. Phase one was for preconditioning only; all comparisons between treatments were performed on phase two data.

## Environments

We constructed three experimental environmental conditions, abbreviated hereafter as “ENV-A”, “ENV-B”, and “ENV-ALL”. In ENV-A, organisms are rewarded for expressing the NOT, AND, and OR Boolean logic functions, and organisms are punished for expressing the NAND, AND-

NOT, and OR-NOT functions. ENV-B is the reverse of ENV-A; that is, in ENV-B, organisms are rewarded for expressing the NAND, AND-NOT, and OR-NOT functions and are punished for expressing the NOT, AND, and OR functions. In ENV-ALL, organisms are rewarded for expressing each of the NOT, AND, OR, NAND, AND-NOT, and OR-NOT functions. In all environmental conditions (ENV-A, ENV-B, and ENV-ALL), a rewarded function expressed by an organism doubles their metabolic rate, allowing them to execute twice as many instructions in the same amount of time. A punished function halves an organism’s metabolic rate. Each Boolean logic function is a non-trivial trait to evolve, as they each require the coordination of multiple genetic instructions to express (Lenski et al., 2003).

In both the PLASTIC and NON-PLASTIC treatments, the environment cycles between equal-length periods of ENV-A and ENV-B. Each of these periods persist for 100 updates (approximately 15 to 20 generations). Thus, populations experience a total of 1,000 full periods of ENV-A interlaced with 1,000 full periods of ENV-B during each experimental phase. Previous work has shown that this rate of change reliably allows for the evolution of adaptive phenotypic plasticity in *Avida* (Clune et al., 2007; Lalejini and Ofria, 2016).

Organisms in the PLASTIC treatments differentiate between ENV-A and ENV-B by executing one of six sensory instructions, each associated with a particular Boolean logic function; these sensory instructions detect whether their associated function is currently rewarded or punished. By using sensory information in combination with execution flow-control instructions, organisms can conditionally perform different functions depending on the current environmental conditions.

### **Experiment Phase 1 – Environment preconditioning**

For each treatment, we founded 100 independent populations from a common ancestral strain capable only of self-replication. At the end of phase one, we identified the most abundant (*i.e.*, dominant) genotype and sampled an organism with that genotype from each replicate population to found a new population for phase two.

For the PLASTIC treatment, we needed to ensure that our observations during the second phase of each experiment reflected the evolutionary consequences of adaptive plasticity. To do so, measured the plasticity of each PLASTIC-treatment population’s dominant genotype by independently testing that genotype in each of ENV-A and ENV-B and recording the phenotype expressed in each environment. We discarded PLASTIC-treatment phase one populations if the

dominant genotype did not exhibit optimal plasticity (as defined in Section 2.2.1), which ensured that PLASTIC-treatment phase two populations were founded with an optimally plastic organism.

### **Experiment Phase 2A – Evolutionary change rate**

We conducted experiment phase 2A to test for differences in evolutionary change—the accumulation of genetic and phenotypic changes—among populations evolving under each of our three treatment conditions (PLASTIC, NON-PLASTIC, and STATIC). Phase 2A continued exactly as phase one, except we tracked the rates of evolutionary change in each of the PLASTIC-, NON-PLASTIC-, and STATIC-treatment populations. Specifically, we quantified evolutionary change using four metrics (each described in Table 2.1): (1) coalescence event count, (2) mutation count, (3) phenotypic volatility, and (4) mutational robustness.

While environmental conditions during phases one and 2A are identical, these phases are distinct in how populations are founded: each phase one population is founded with a common ancestor capable only of self-replication, whereas each phase two population is founded with an organism that was evolved in its treatment-specific conditions. Thus, during phase two, phylogenies are rooted with an ancestor that is well-adapted to its treatment conditions, which, in turn, ensures that our observations can exclude dynamics associated with initial adaptation.

### **Experiment Phase 2B – Novel function evolution**

We conducted experiment phase 2B to quantify the extent to which organisms evolved under PLASTIC-, NON-PLASTIC-, and STATIC-treatment conditions were able to acquire and retain novel functions. Phase 2B extended the conditions of phase one by adding 71 novel Boolean logic functions (Ofria et al., 2009), which were always rewarded in all treatments. The original six phase one functions (NOT, NAND, AND, OR-NOT, OR, and AND-NOT; hereafter called “base” functions) continued to be rewarded or punished according to the particular treatment conditions. An organism’s metabolic rate was increased by 10% for each novel function that it expressed (limited to one reward per function). This reward provided a selective pressure to evolve these functions, but their benefits did not overwhelm the 100% metabolic rate increase conferred by rewarded base functions. As such, populations in the PLASTIC and NON-PLASTIC treatments could not easily escape environmental fluctuations by abandoning the fluctuating base functions.

During this experiment, we used three metrics to quantify novel function acquisition and retention in evolving populations (each described in Table 2.1): (1) final novel function count, (2) novel

function discovery, and (3) novel function loss.

### **Experiment Phase 2C – Deleterious instruction accumulation**

We conducted experiment phase 2C to quantify the extent to which organisms evolved under PLASTIC-, NON-PLASTIC-, and STATIC-treatment conditions acquired deleterious instructions via mutation. Phase 2C extended the instruction set of phase one with a **deleterious** instruction. When an organism executes a **deleterious** instruction, it performs a “deleterious” function, which reduces the organism’s metabolic rate (and thus reproductive success) but does not otherwise alter the organism’s behavior. We imposed a 10% penalty each time an organism performed the deleterious function, making the **deleterious** instruction explicitly harmful to execute. We did not limit the number of times that an organism could perform the deleterious function, and as such, organisms could perform the deleterious function as many times as they executed the **deleterious** instruction.

We tracked the number of times each organism along the dominant lineage performed the deleterious function. Specifically, we used two metrics (each described in Table 2.1): (1) final deleterious function count and (2) deleterious function acquisition count.

### **2.2.3 Experimental analyses**

For each of our experiments, we tracked and analyzed the phylogenetic histories of evolving populations during phase two and generated a set of summary statistics (Table 2.1). For each phylogenetic history, we counted the number of times that the most recent common ancestor for the population shifted and used this value as the number of coalescence events. Next, at the final time point, we identified the most abundant genotype in the evolved population, and chose a *representative organism* from that genotype for further analysis. We used the lineage from the founding organism to the representative organism to summarize the evolutionary pathway of a population. This lineage represents the majority of the evolutionary history from the given population as long as the entire population traces back to the lineage in recent history. We manually inspected evolved phylogenies and found no evidence that any of our experimental treatments supported long-term coexistence. As such, analyses of the representative lineages reflect the majority of evolutionary history for a given population.

Some of our metrics (Table 2.1) required us to measure genotype-by-environment interactions. Importantly, in the fluctuating environments, we needed to differentiate phenotypic changes that

were caused by mutations from those that were caused by environmental changes. To accomplish this, we produced organisms with the given focal genotype, measured their phenotype in each environmental condition, and aggregated the resulting phenotypes to create a *phenotypic profile*. Although organisms with different genotypes may express the same set of functions across environmental conditions, their phenotypic profiles may not necessarily be the same. For example, an organism that expresses NOT in ENV-A and NAND in ENV-B has a distinct phenotypic profile from one that expresses NAND in ENV-A and NOT in ENV-B. Because phenotypic profiles encapsulate function expression across all relevant environmental conditions (ENV-A and ENV-B), a change in phenotypic profile from parent to offspring indicates a mutationally-induced phenotypic change.

While most analyses employed here are retrospective metrics applied to lineages, digital evolution allows precise manipulations on individual organisms and genomes. Mutational robustness uses this technique when looking at the possible mutations on a representative genotype. Genomes in Avida are linear sequences of instructions, and as such, possible mutations can be simulated by substituting other instructions at the desired site. Indeed, the mutational robustness of a genotype examines all one-step mutations (*i.e.*, each mutation where exactly one instruction is substituted). This allows us to disentangle whether the frequency of mutationally-induced phenotypic changes observed along a lineage is a consequence of evolved genetic architectures versus the result of population dynamics that actively skewed the distribution of organisms along the lineage; that is, are genomes organized such that mutations are more likely to induce phenotypic changes, or are organisms with phenotypes different from their parents more likely to be successful and thus appear along the representative lineage?

#### **2.2.4 Statistical analyses**

Across all of our experiments, we differentiated between sample distributions using non-parametric statistical tests. For each major analysis, we first performed a Kruskal-Wallis test (Kruskal and Wallis, 1952) to determine if there were significant differences in results from the PLASTIC, NON-PLASTIC, and STATIC treatments (significance level  $\alpha = 0.05$ ). If so, we applied a Wilcoxon rank-sum test (Wilcoxon, 1992) to distinguish between pairs of treatments. We applied Bonferroni corrections for multiple comparisons (Rice, 1989) where appropriate.

### 2.2.5 Software availability

We conducted our experiments using a modified version of the Avida software, which is open source and freely available on GitHub (Lalejini and Ferguson, 2021b). Modifications to Avida included an improved phylogeny tracking system that enabled us to track coalescence events and the addition of custom sensory instructions specific to our experiments. We used Python for data processing, and we conducted all statistical analyses using R version 4 (R Core Team, 2021). We used the tidyverse collection of R packages (Wickham et al., 2019) to wrangle data, and we used the following R packages for analysis, graphing, and visualization: ggplot2 (Wickham et al., 2020), cowplot (Wilke, 2020), Color Brewer (Harrower and Brewer, 2003; Neuwirth, 2014), rstatix (Kasambara, 2021), ggsignif (Ahlmann-Eltze and Patil, 2021), scales (Wickham and Seidel, 2020), Hmisc (Harrell Jr, 2020), fmsb (Nakazawa, 2019), and boot (Canty and Ripley, 2019). We used R markdown (Allaire et al., 2020) and bookdown (Xie, 2020) to generate web-enabled supplemental material. All of the source code for our experiments and analyses, including configuration files and guides for replication, can be found in our supplemental material, which is hosted on GitHub (Lalejini and Ferguson, 2021b). Additionally, our experimental data is available on the Open Science Framework at <https://osf.io/sav2c/> (Lalejini and Ferguson, 2021a).

<b>Metric</b>	<b>Description</b>
Coalescence event count	Number of coalescence events that have occurred, which indicates the frequency of selective sweeps in the population.
Mutation count	Sum of all mutations that have occurred along a lineage.
Phenotypic volatility	Number of instances where parent and offspring phenotypic profiles do not match along a lineage.
Mutational robustness	Proportion of mutations (from the set of all possible one-step mutations) that do not change the phenotypic profile of a focal genotype. We also measured <i>realized mutational robustness</i> , which is the proportion of mutated offspring along a lineage whose phenotypic profile matches that of their parent.
Final novel function count	Count of unique novel functions performed by the representative organism in a final population from experiment phase 2B. This metric can range from 0 to 71 and measures how well the fitness landscape was exploited at a given point in time.
Novel function discovery	Number of unique novel functions ever performed along a given lineage in experimental phase 2B, even if a function is later lost. This metric can range from 0 to 71 and measures a given lineage's level of exploration of the fitness landscape.
Novel function loss	Number of instances along a given lineage from experimental phase 2B where a novel function is performed by a parent but not its offspring. This metric measures how often a given lineage fails to retain evolved traits over time.
Final deleterious function count	Number of times the deleterious function is performed by the representative organism from a final population from experiment phase 2C.
Deleterious function acquisition count	Number of instances along a given lineage where a mutation causes an offspring to perform the deleterious function more times than its parent.

Table 2.1: Metric descriptions.

## 2.3 Results

### 2.3.1 Adaptive phenotypic plasticity slows evolutionary change in fluctuating environments

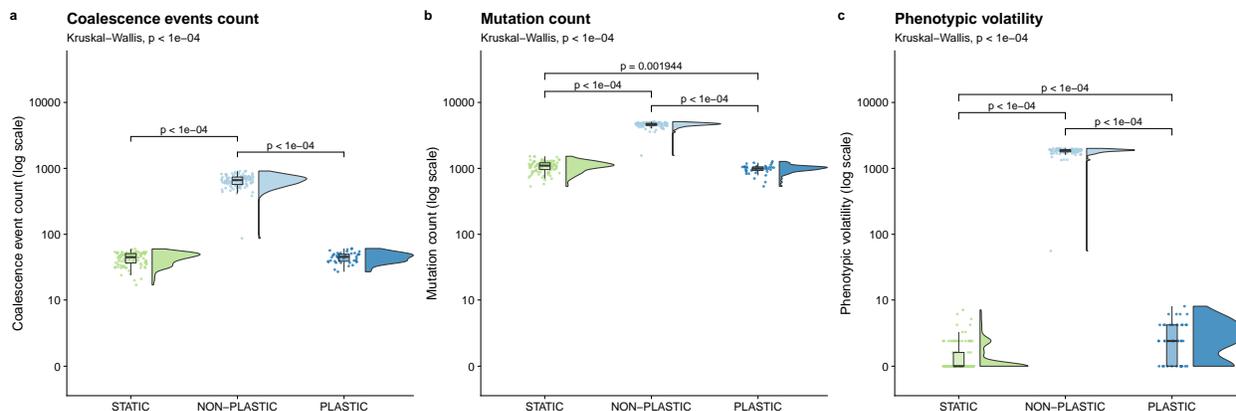


Figure 2.3: Raincloud plots (Allen et al., 2019) of (a) coalescence event count, (b) mutation count, and (c) phenotypic volatility. See Table 2.1 for descriptions of each metric. Each plot is annotated with statistically significant comparisons (Bonferroni-corrected pairwise Wilcoxon rank-sum tests). Note that adaptive phenotypic plasticity evolved in 42 of 100 replicates from the PLASTIC treatment during phase one of this experiment; we used this more limited group to found 42 phase-two PLASTIC replicates from which we report these PLASTIC data.

In experimental phase 2A, we tested whether adaptive phenotypic plasticity constrained or promoted subsequent evolutionary change in a fluctuating environment. First, we compared the total amount of evolutionary change in populations evolved under the PLASTIC, NON-PLASTIC, and STATIC treatments as measured by coalescence event count, mutation count, and phenotypic volatility (Figure 2.3). According to each of these metrics, NON-PLASTIC populations experienced a larger magnitude of evolutionary change than either PLASTIC or STATIC populations. We observed significantly higher coalescence event counts in NON-PLASTIC populations than in PLASTIC or STATIC populations (Figure 2.3a). NON-PLASTIC lineages had significantly higher mutation counts (Figure 2.3b) and phenotypic volatility than PLASTIC or STATIC lineages (Figure 2.3c).

Changing environments have been shown to increase generational turnover (*i.e.*, how rapidly generations elapse) in *Avida* populations (Canino-Koning et al., 2016), which could explain why we observe a larger magnitude of evolutionary change at the end of 200,000 updates of evolution in NON-PLASTIC populations. Indeed, we found that significantly more generations of evolution elapsed in NON-PLASTIC populations (mean of  $41090 \pm 2702$  std. dev.) than in PLASTIC (mean

of  $31016 \pm 2615$  std. dev.) or STATIC (mean of  $30002 \pm 3011$  std. dev.) populations during phase 2A (corrected Wilcoxon rank-sum tests,  $p < 10^{-4}$ ).

To evaluate whether increased generational turnover explains the greater magnitude of evolutionary change in NON-PLASTIC populations, we examined the average number of generations between coalescence events and the realized mutational robustness of lineages (Table 2.1). A coalescence event indicates a selective sweep, which is a hallmark of adaptive evolutionary change. Realized mutational robustness measures the frequency that mutations cause phenotypic changes along a lineage. We expect that static conditions should favor fit lineages with high realized mutational robustness that no longer undergo rapid adaptive change and hence do not trigger frequent coalescence events. Under fluctuating conditions, however, lineages must be composed of plastic organisms if they are to maintain both high fitness and realized mutational robustness. Without plasticity, we expect fluctuating conditions to produce lineages with low realized mutational robustness and frequent coalescence events as populations must continually acquire and fix mutations to readapt to the environment.

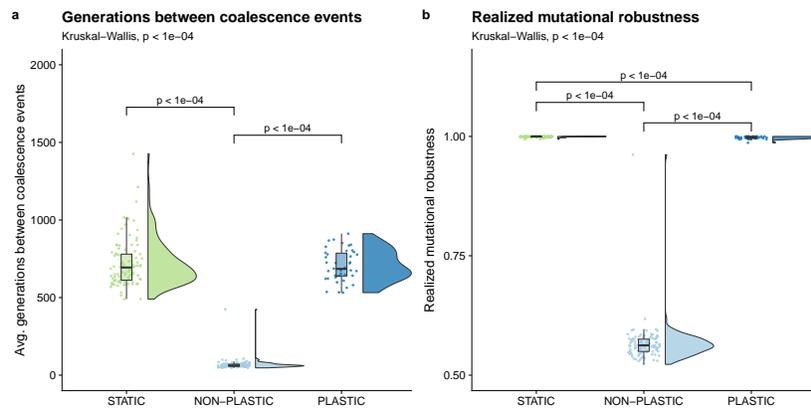


Figure 2.4: Raincloud plots of (a) average number of generations between coalescence events, and (b) realized mutational robustness (Table 2.1). Each plot is annotated with statistically significant comparisons (Bonferroni-corrected pairwise Wilcoxon rank-sum tests).

On average, significantly fewer generations elapsed between coalescence events in NON-PLASTIC populations than in either PLASTIC or STATIC populations (Figure 2.4a). We also found that both STATIC and PLASTIC lineages exhibited higher realized mutational robustness relative to that of NON-PLASTIC lineages (Figure 2.4b); that is, mutations observed along NON-PLASTIC lineages more often caused phenotypic changes in offspring. Overall, our results indicate

that NON-PLASTIC populations underwent more rapid (and thus a greater amount of) evolutionary change than either PLASTIC or STATIC populations.

While both STATIC and PLASTIC lineages exhibited high realized mutational robustness, we found that STATIC lineages exhibited higher realized robustness than PLASTIC lineages (Figure 2.4b). Overall, there were rare instances of mutations that caused a change in phenotypic profile across all PLASTIC lineages. Of these mutations, we found that over 80% (83 out of 102) of changes to phenotypic profiles were cryptic. That is, the mutations affected traits that would not have been expressed in the environment that the organism was born into but would have been expressed had the environment changed.

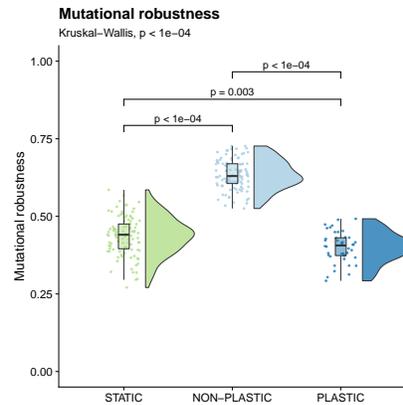


Figure 2.5: Raincloud plot of mutational robustness of each representative genotype (Table 2.1). The plot is annotated with statistically significant comparisons (Bonferroni-corrected pairwise Wilcoxon rank-sum tests).

Given that NON-PLASTIC lineages exhibited the lowest realized mutational robustness of our three experimental treatments, we sought to determine if this effect was driven by differences in evolved genetic architectures. Specifically, did the NON-PLASTIC genetic architectures evolve such that mutations were more likely to result in phenotypic change? Such a mutational bias would trade off descendant fitness in the same environment in exchange for a chance of increasing descendant fitness in alternate environments. This strategy would be an example of diversifying bet-hedging (*i.e.*, reducing expected mean fitness to lower variance in fitness) (Childs et al., 2010). Alternatively, the lower realized mutational robustness in NON-PLASTIC lineages could be due to survivorship bias, as we measured realized mutational robustness as the fraction of mutations observed along *successful* lineages that caused a phenotypic change. That is, our measure of realized mutational

robustness concentrates only on mutations that appear in the representative lineage (*i.e.*, “survivors” of selection), ignoring mutations that did not. Thus, lower realized mutational robustness in NON-PLASTIC lineages could simply be due to phenotype-altering mutations being more frequently favored by selection.

We analyzed the mutational robustness of representative genotypes by calculating the fraction of single-instruction mutations that change the phenotypic profile. We found that mutations to representative genotypes on NON-PLASTIC lineages are *less* likely to result in a phenotypic change than mutations to comparable genotypes on either STATIC or PLASTIC lineages (Figure 2.5). These data provide evidence against NON-PLASTIC lineages engaging in a mutation-driven bet-hedging strategy, and instead, are consistent with the hypothesis that lower realized mutational robustness in the NON-PLASTIC treatment was due to survivorship bias.

In general, adaptive plasticity stabilized PLASTIC-treatment populations against environmental fluctuations, and their evolutionary dynamics more closely resembled those of populations evolving in a static environment. We observed no significant difference in the number and frequency of coalescence events in PLASTIC and STATIC populations. We did, however, observe small, but statistically significant, differences in each of the following metrics: elapsed generations, mutation counts, phenotypic volatility, realized mutational robustness, and mutational robustness between PLASTIC and STATIC populations. We expect that these differences are a result of plastic organisms needing to simultaneously maintain both function and function-regulation machinery, resulting in more genetic components that can be broken by mutations; moreover, many of these components are under relaxed selection in periods between environmental changes. Overall, these differences were not substantial enough to play an obvious role in any of the dynamics we analyzed, but could be examined further in the future.

### **2.3.2 Adaptively plastic populations retain more novel functions than non-plastic populations in fluctuating environments**

We have so far shown that adaptive plasticity constrains the rate of evolutionary change in fluctuating environments. However, it is unclear how this dynamic influences the evolution of novel functions. Based on their relative rates of evolutionary change, we might expect NON-PLASTIC-treatment populations to evolve more novel functions than PLASTIC-treatment populations. But, how much of the evolutionary change in NON-PLASTIC populations is useful for exploring novel

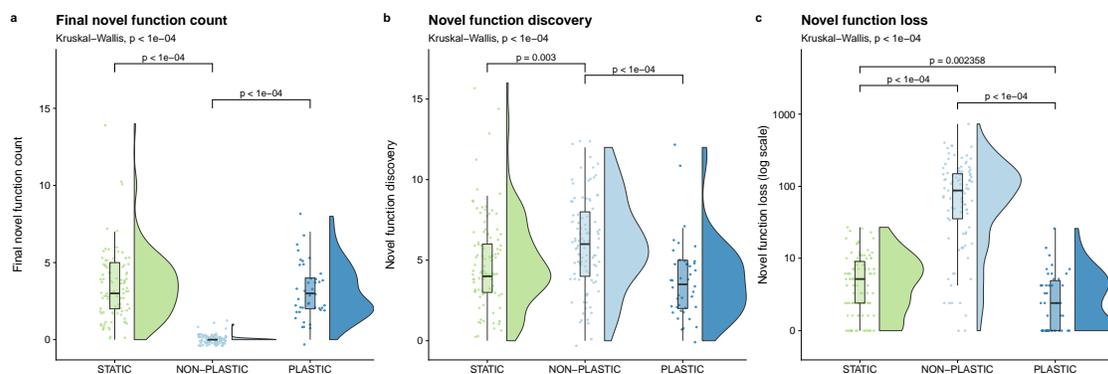


Figure 2.6: Raincloud plots of (a) final novel function count, (b) novel function discovery, and (c) novel function loss. See Table 2.1 for descriptions of each metric. Each plot is annotated with statistically significant comparisons (Bonferroni-corrected pairwise Wilcoxon rank-sum tests). Note that adaptive phenotypic plasticity evolved in 42 of 100 replicates from the PLASTIC treatment during phase one of this experiment; we used this more limited group to seed the resulting 42 phase-two PLASTIC replicates.

regions of the fitness landscape versus continually rediscovering the same regions?

To answer this question, we quantified the number of novel functions performed by a representative organism in the final population of each replicate. We found that both PLASTIC and STATIC populations had significantly higher final function counts than NON-PLASTIC populations at the end of the experiment (Figure 2.6a). The final novel function count in PLASTIC and STATIC lineages could be higher than that of the NON-PLASTIC lineages for several non-mutually exclusive reasons. One possibility is that PLASTIC and STATIC lineages could be exploring a larger area of the fitness landscape when compared to NON-PLASTIC lineages. Another possibility is that the propensity of the NON-PLASTIC lineages to maintain novel traits could be significantly lower than PLASTIC or STATIC lineages. When we looked at the total sum of novel functions discovered by each of the PLASTIC, STATIC, and NON-PLASTIC lineages, we found that NON-PLASTIC lineages generally explored a larger area of the fitness landscape (Figure 2.6b). Although the NON-PLASTIC lineages discovered more novel functions, those lineages also exhibited significantly higher novel function loss when compared to PLASTIC and STATIC lineages (Figure 2.6c).

A larger number of generations elapsed in NON-PLASTIC populations than in PLASTIC or STATIC populations during our experiment (Lalejini and Ferguson, 2021b). Are NON-PLASTIC lineages discovering and losing novel functions more frequently than PLASTIC or STATIC lineages, or are our observations a result of differences in generational turnover? To answer this question, we

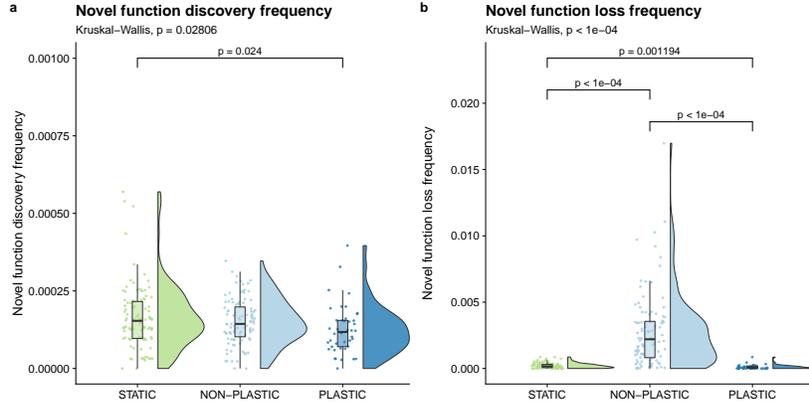


Figure 2.7: Raincloud plots of (a) novel function discovery frequency and (b) novel function loss frequency. Each plot is annotated with statistically significant comparisons (Bonferroni-corrected pairwise Wilcoxon rank-sum tests).

converted the metrics of novel function discovery and novel function loss to rates by dividing each metric by the number of elapsed generations along the associated representative lineages. We found no significant difference in the frequency of novel function discovery between NON-PLASTIC and STATIC lineages, and we found that PLASTIC lineages had a lower frequency of novel function discovery than STATIC lineages (Figure 2.7a). Therefore, we cannot reject the possibility that the larger magnitude of function discovery in NON-PLASTIC lineages was driven by a larger number of elapsed generations. NON-PLASTIC lineages had a higher frequency of function loss than either PLASTIC or STATIC lineages, and PLASTIC lineages tended to have a lower frequency of novel function loss than STATIC lineages (Figure 2.7b).

Next, we examined the frequency at which novel function loss along lineages co-occurred with the loss or gain of any of the six base functions. Across all NON-PLASTIC representative lineages, over 97% (10998 out of 11229) of instances of novel function loss co-occurred with a simultaneous change in base function profile. In contrast, across all PLASTIC and STATIC dominant lineages, we observed that approximately 20% (29 out of 142) and 2% (13 out of 631), respectively, of instances of novel function loss co-occurred with a simultaneous change in base function profile. As such, the losses of novel functions in NON-PLASTIC lineages appear to be primarily due to hitchhiking or epistatic effects where a mutation that knocks out a maladaptive base function (after the environment changes) also knocks out a beneficial novel function.

### 2.3.3 Lineages without plasticity that evolve in fluctuating environments express more deleterious functions

Phenotypic plasticity allows for genetic variation to accumulate in genomic regions that are unexpressed, which could lead to the fixation of deleterious instructions in PLASTIC populations. However, in NON-PLASTIC lineages, we observe a higher rate of novel function loss, indicating that they may be more susceptible to deleterious mutations (Figure 2.7b).

Therefore, in experiment phase 2C, we tested whether adaptive phenotypic plasticity can increase the incidence of deleterious function performance. Specifically, we added an instruction that triggered an explicitly deleterious function and measured the number of times it was executed. Each execution of the `deleterious` instruction reduces an organism’s fitness by 10%. At the beginning of phase 2C, the `deleterious` instruction is not present in the population, as it was not part of the instruction set during phase one of evolution. Accordingly, if a `deleterious` instruction fixes in a population, it must be the result of evolutionary dynamics during phase 2C, including cryptic variation, hitchhiking, or as a result of sign epistasis where a `deleterious` instruction knocks out an even more maladaptive trait.

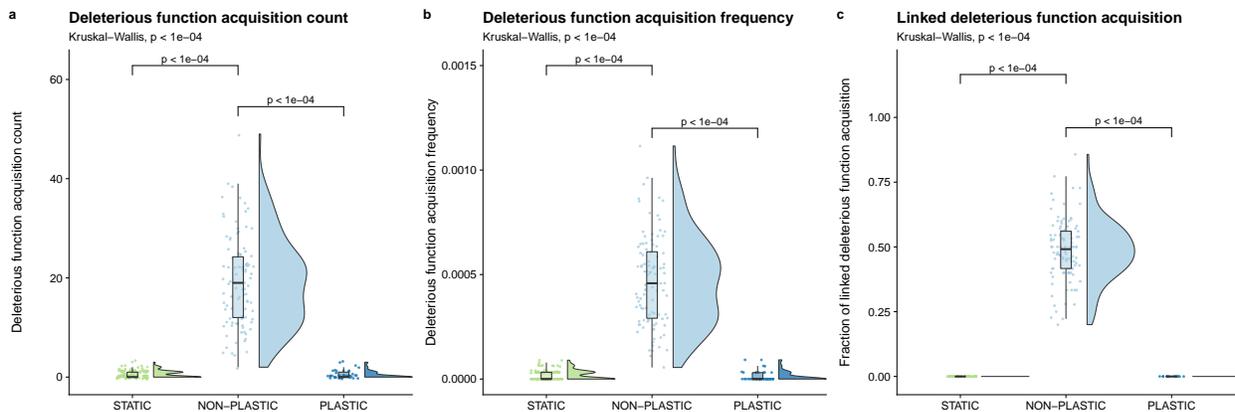


Figure 2.8: Raincloud plots of (a) deleterious function acquisition, (b) deleterious function acquisition frequency, and (c) the proportion of mutations that increase deleterious function expression along a lineage that co-occur with a change in phenotypic profile. Each plot is annotated with statistically significant comparisons (Bonferroni-corrected pairwise Wilcoxon rank-sum tests). Note that adaptive phenotypic plasticity evolved in 43 of 100 replicates from the PLASTIC treatment during phase one of this experiment; we used this more limited group to seed the 43 phase-two PLASTIC replicates.

At the end of our experiment, no representative organisms from the PLASTIC or STATIC treatments performed the deleterious function under any environmental condition; however, repre-

representative organisms in 14% of replicates of the NON-PLASTIC treatment performed the deleterious function at least once. NON-PLASTIC lineages contained significantly more mutations that conferred the deleterious function as compared to PLASTIC or STATIC lineages (Figure 2.8a), and these mutations occurred at a significantly higher frequency in NON-PLASTIC lineages (Figure 2.8b).

Next, we measured how often mutations that increased deleterious function performance co-occurred with changes to the base function profile within representative lineages. A **deleterious** instruction can fix in a population by having a beneficial effect that outweighs its inherent cost (*e.g.*, knocking out a punished function) or through linkage with a secondary beneficial mutation at another site within the genome. Across all NON-PLASTIC representative lineages, we found that approximately 49% (956 out of 1916) of mutations that increased deleterious function expression co-occurred with a change in the base function profile (Figure 2.8c). In all representative lineages from the PLASTIC treatment, only 18 mutations increased deleterious function expression, and none co-occurred with a change in base function profile (Figure 2.8c). Likewise, only 58 mutations increased deleterious function performance in all representative lineages from the STATIC treatment, and none co-occurred with a change in base function profile (Figure 2.8c). We did not find compelling evidence that the few mutations that increased deleterious function expression occurred as cryptic variation in PLASTIC lineages.

We repeated this experiment with 3% and 30% metabolic rate penalties associated with the deleterious function, which produced results that were consistent with those reported here (Lalejini and Ferguson, 2021b).

## 2.4 Discussion

In this work, we used evolving populations of digital organisms to determine how adaptive phenotypic plasticity alters subsequent evolutionary dynamics and influences evolutionary outcomes in fluctuating environments. Specifically, we compared lineages of adaptively plastic organisms in fluctuating environments to both non-plastic organisms in those same environments and other non-plastic organisms in static environments.

### 2.4.1 Evolutionary change

We found strong evidence that adaptive plasticity slows evolutionary change in fluctuating environments. Adaptively plastic populations experienced fewer coalescence events and fewer total

genetic changes relative to non-plastic populations evolving under identical environmental conditions (Figure 2.3). Whereas non-plastic populations relied on *de novo* mutations to adapt to each environmental fluctuation, plastic populations leveraged sensory instructions to regulate function performance. Indeed, in fluctuating environments, selection pressures toggle after each environmental change. We hypothesize that in non-plastic populations such toggling would repeatedly drive the fixation of mutations that align an organism’s phenotypic profile to the new conditions. This hypothesis is supported by the increased frequency of coalescence events in these populations (Figure 2.4a) as well as increased rates of genetic and phenotypic changes observed along the lineages of non-plastic organisms.

Representative lineages in the non-plastic treatment experienced lower realized mutational robustness than plastic and static lineages (Figure 2.4b). We reasoned that this lower realized mutational robustness was due to non-plastic populations evolving a bet-hedging strategy where mutations are more likely to modify the phenotypic profile. However, when we switched from measuring the realized mutational robustness of representative lineages to measuring the mutational robustness of representative genotypes (*i.e.*, what fraction of one-step mutants change the phenotypic profile), we observed that non-plastic genotypes exhibited the highest mutational robustness of all three treatments (Figure 2.5). This result runs contrary to both our expectations and the results of other fluctuating environment studies in *Avida* (Canino-Koning et al., 2019). Canino-Koning et al. (2019) found that mutational robustness is negatively correlated with the number of function-encoding sites in the genome. In our work, most plastic and static genotypes encode all six base functions, while most non-plastic genotypes only encode functions from one environment; this results in fewer function-encoding sites, which may increase mutational robustness in non-plastic genotypes (relative to plastic and static genotypes). Regardless of the cause, this higher mutational robustness in non-plastic organisms indicates that bet-hedging is not driving the low realized mutational robustness observed in non-plastic lineages. Thus, we expect the lower realized mutational robustness in non-plastic lineages to be driven by survivorship bias. Because non-plastic lineages must rely on mutations to adapt to environmental changes, phenotype-altering mutations are often highly advantageous, and their selection decreases the realized mutational robustness of successful lineages.

To our knowledge, this study is the first in-depth empirical investigation into how the *de novo*

evolution of adaptive plasticity shifts the course of subsequent evolution in a cyclic environment. The relative rates of evolutionary change that we observed in non-plastic populations, however, are consistent with results from previous digital evolution studies. For example, Dolson et al. (2020) showed that non-plastic populations that were evolved in cyclically changing environments exhibited higher phenotypic volatility and accumulated more mutations than that of populations evolved under static conditions. Furthermore, Lalejini and Ofria (2016) visually inspected the evolutionary histories of non-plastic organisms evolved in fluctuating environments, observing that mutations along successful lineages readily switched the set of traits expressed by offspring.

Our results are also consistent with conventional evolutionary theory. A trait’s evolutionary response to selection depends on the strength of directional selection and on the amount of genetic variation for selection to act upon (Lande and Arnold, 1983; Zimmer and Emlen, 2013). In our experiments, non-plastic populations repeatedly experienced strong directional selection to toggle which functions were expressed after each environmental change. As such, retrospective analyses of successful lineages revealed rapid evolutionary responses (that is, high rates of genetic and phenotypic changes). Evolved adaptive plasticity shielded populations from strong directional selection when the environment changed by eliminating the need for a rapid evolutionary response to toggle function expression. Indeed, both theoretical and empirical studies have shown that adaptive plasticity can constrain evolutionary change by weakening directional selection on evolving populations (Ghalambor et al., 2015; Paenke et al., 2007; Price et al., 2003).

#### **2.4.2 The evolution and maintenance of novel functions**

In fluctuating environments, non-plastic populations explored a larger area of the fitness landscape than adaptively plastic populations (Figure 2.6b). However, adaptively plastic populations better exploited the fitness landscape, retaining a greater number of novel functions than non-plastic populations evolving under identical environmental conditions (Figure 2.6a). In our experiment, novel functions were less important to survival than the fluctuating base functions. In non-plastic populations, when a mutation changes a base function to better align with current environmental conditions, its benefit will often outweigh the cost of losing one or more novel functions. Indeed, we found that along non-plastic representative lineages, 97% of the mutations associated with novel function loss co-occurred with phenotypic changes that helped offspring adapt to current environmental conditions.

Previous studies have shown that transitory environmental changes can improve overall fitness landscape exploration in evolving populations of non-plastic digital organisms (Nahum et al., 2017). Similarly, changing environments have been shown to increase the rate of evolutionary adaptation in simulated network models (Kashtan et al., 2007). In our system, however, we found that *repeated* fluctuations reduced the ability of non-plastic populations to maintain and exploit functions; that said, we did find that repeated fluctuations may improve overall function discovery by increasing generational turnover. Consistent with our findings, Canino-Koning et al. (2019) found that non-plastic populations of digital organisms evolving in a cyclic environment maintained fewer novel traits than populations evolving in static environments.

Our results suggest that adaptive phenotypic plasticity can improve the potential for populations to exploit novel resources by stabilizing them against stressful environmental changes. The stability that we observe may also lend some support to the hypothesis that phenotypic plasticity can rescue populations from extinction under changing environmental conditions (Chevin et al., 2010).

Our data do not necessarily provide evidence for or against the genes as followers hypothesis. The genes as followers hypothesis focuses on contexts where plastic populations experience novel or abnormally stressful environmental change. However, in our system, environmental changes were cyclic (not novel), and no single environmental change was *abnormally* stressful. Further, the introduction of novel functions during the second phase of the experiment merely added static opportunities for fitness improvement. This addition did not change the meaning of existing environmental cues, nor did it require those cues to be used in new ways.

### 2.4.3 The accumulation of deleterious alleles

We found that non-plastic lineages that evolved in a fluctuating environment exhibited both greater totals and higher rates of deleterious function acquisition than that of adaptively plastic lineages (Figure 2.8). There are several, non-mutually exclusive possibilities that could explain the fixation of explicitly deleterious instructions: random genetic drift, deleterious hitchhiking, epistatic effects, and cryptic variation (in plastic organisms). We find it unlikely that random genetic drift explains our observations. Each time an organism expresses a **deleterious** instruction, the organism incurs a 10% penalty to their replication rate, which results in strong purifying selection against mutations that cause offspring to execute **deleterious** instructions.

In asexual populations without horizontal gene transfer, all co-occurring mutations are linked. As such, deleterious mutations linked with a stronger beneficial mutation (*i.e.*, a driver) can sometimes “hitchhike” to fixation (Buskirk et al., 2017; Smith and Haigh, 1974; Van den Bergh et al., 2018). Natural selection normally prevents deleterious mutations from reaching high frequencies, as such mutants are outcompeted. However, when a beneficial mutation sweeps to fixation in a clonal population, it carries along any linked genetic material, including other beneficial, neutral, or deleterious mutations (Barton, 2000; Smith and Haigh, 1974). Therefore, deleterious genetic hitchhiking could have contributed to **deleterious** instruction accumulation along non-plastic lineages in changing environments.

Epistatic effects (*i.e.*, interactions between genes) could have also contributed to **deleterious** instruction accumulation along non-plastic lineages. On their own, mutations that increase **deleterious** instruction execution are maladaptive; however, if such a mutation were to also knock out an even more harmful function, that mutation may have a net beneficial effect. As such, mutations that confer increased **deleterious** instruction execution could directly drive a selective sweep.

Representative lineages from non-plastic populations in the cyclic environment exhibited higher mutation accumulation (Figure 2.3b), novel function loss (Figure 2.6c), and deleterious function acquisition (Figure 2.8a) than their plastic counterparts. In aggregate, we found that many (~49%; 956 / 1916) mutations that increased **deleterious** instruction execution in offspring co-occurred with mutations that provided an even stronger benefit by adapting the offspring to an environmental change. We expect that an even larger fraction of these deleterious mutations were linked to beneficial mutations, but our analysis only counted mutations that co-occurred in the same generation. Our analyses did not distinguish between epistatic effects and deleterious hitchhiking; however, more fine-grained analyses of secondary effects of mutations that conferred **deleterious** instruction execution could be performed in future work to disentangle these two mechanisms.

Theory predicts that under relaxed selection deleterious mutations should accumulate as cryptic variation in unexpressed traits (Lahti et al., 2009). Contrary to this expectation, we did not find evidence of **deleterious** instructions accumulating as cryptic variation in adaptively plastic lineages. One possible explanation is that the period of time between environmental changes was too brief for variants carrying unexpressed **deleterious** instructions to drift to high frequencies

before the environment changed, after which purifying selection would have removed such variants. Indeed, we would not expect drift to fix an unexpressed trait since we tuned the frequency of environmental fluctuations to prevent valuable traits from being randomly eliminated during the off environment. Additionally, plastic organisms in Avida usually adjust their phenotype by toggling the expression of a minimal number of key instructions, leaving little genomic space for cryptic variation to accumulate.

#### 2.4.4 Limitations and future directions

Our work lays the groundwork for using digital evolution experiments to investigate the evolutionary consequences of phenotypic plasticity in a range of contexts. However, the data presented here are limited to the evolution of *adaptively* plastic populations. Future work might explore the evolutionary consequences of maladaptive and non-adaptive phenotypic plasticity (*e.g.*, Leroi et al. 1994), which are known to bias evolutionary outcomes (Ghalambor et al., 2015).

Additionally in our experiments, sensory instructions perfectly differentiated between ENV-A and ENV-B, and environmental fluctuations never exposed populations to entirely new conditions. These parameters have been shown to influence evolutionary outcomes (Boyer et al., 2021; Li and Wilke, 2004), which if relaxed in the context of further digital evolution experiments, may yield additional insights. Our experiments also focused on asexually reproducing digital organisms. Sexual reproduction has been shown to be advantageous in rapidly changing environments, such as the cyclic environments used in our study (Misevic et al., 2010). Future work could investigate how sexual reproduction affects the evolutionary consequences of adaptive plasticity.

We focused our analyses on the lineages of organisms with the most abundant genotype in the final population. These successful lineages represented the majority of the evolutionary histories of populations at the end of our experiment, as populations did not exhibit long-term coexistence of different clades. Our analyses, therefore, gave us an accurate picture of what fixed in the population. We did not, however, examine the lineages of extinct clades. Future work will extend our analyses to include extinct lineages, giving us a more complete view of evolutionary history, which may allow us to better distinguish adaptively plastic populations from populations evolving in a static environment.

As with any wet-lab experiment, our results are in the context of a particular model organism: “Avidian” self-replicating computer programs. Digital organisms in Avida regulate responses to en-

vironmental cues using a combination of sensory instructions and conditional logic instructions (`if` statements). The `if` instructions conditionally execute a single instruction depending on previous computations and the state of memory. As such, plastic organisms in Avida typically regulate phenotypes by toggling the expression of a small number of key instructions as opposed to regulating cohorts of instructions under the control of a single regulatory sequence (Lalejini and Ferguson, 2021b). This bias may limit the accumulation of hidden genetic variation in Avida genomes. However, as there are many model biological organisms, there are many model digital organisms that have different regulatory mechanisms (*e.g.*, Lalejini and Ofria 2018) that should be used to test the generality of our results.

As with most digital evolution experiments, our mutation rates were high and population sizes were small (3600 individuals) relative to experiments with microbes or conditions common in nature. As such, beneficial mutations can be generated rapidly and selective sweeps can occur quickly. Moreover, our analyses were limited to a single rate of environmental change and simple function reward structures, which likely influenced the rates of selective sweeps observed in our experiments. Future studies could address these limitations by increasing population sizes, decreasing mutation rates, investigating different function rewards and punishments, and altering the time spent in each environment.

## Supplemental Material

The supplemental material for this article is hosted on GitHub and can be found online at <https://github.com/amlalejini/evolutionary-consequences-of-plasticity> (Lalejini and Ferguson, 2021b).

## Data Availability Statement

The datasets generated and analyzed for this study can be found on the Open Science Framework at <https://osf.io/sav2c/> (Lalejini and Ferguson, 2021a).

## Chapter 3

# Potentiating mutations facilitate the evolution of associative learning in digital organisms

Authors: Austin J. Ferguson and Charles Ofria

This chapter is adapted from (Ferguson and Ofria, 2023), which was peer-reviewed and published in the proceedings of the 2023 Conference on Artificial Life.

In this work we use analytic replay experiments to quantify how the probability of evolving associative learning (i.e., its potentiation) changes over the course of evolved digital lineages. Specifically, we replay four case study lineages that successfully evolved learning. We find that potentiation can increase suddenly and identify “potentiating mutations” in each lineage that substantially increase potentiation in a single generation. We then dive into the mechanics of these mutations and how they interact with later mutations to confer associative learning.

### 3.1 Introduction

How likely is the evolution of a particular trait? Researchers have long been interested in predicting evolutionary outcomes, but the inherent stochasticity in the process makes this goal exceptionally challenging. In order to make more accurate predictions, we would need to better understand how and why the underlying probabilities of potential outcomes change over time. Looking purely retrospectively at evolution in nature, this type of analysis is not possible (at least not without a time machine). Leveraging the flexibility and controls available in experimental evolution, however, allows us to empirically test questions that were previously only hypothetical (Kawecki et al., 2012). Here, we focus on Stephen Jay Gould’s idea of “replaying the tape of life” (Gould, 1990). The idea is simple: If we were to start life over again from the same initial conditions, would evolution follow the same pathway? Alas, Gould remarked that this experiment is unfortunately impossible.

While it may be impossible to replay the *entire* tape of life, practitioners of experimental evolution have conducted this experiment on a smaller scale. Travisano et al. (1995), Wagenaar and Adami (2004), and Blount et al. (2008) introduced and refined methods of investigating the role of historical contingency in evolving populations: parallel and analytic replay experiments. By evolving multiple populations from the same starting organisms, researchers can identify the range and distribution of outcomes. These populations can be evolved simultaneously from a given starting

point (parallel replays), however many microbial and digital populations allow us to preserve a “fossil record”, opening up another possibility. Analytic replay experiments systematically revive historical populations to re-evolve them from multiple time points, allowing researchers to identify alternative possibilities after the fact (Blount et al., 2018). For a more thorough review of the story behind analytic replay experiments and an overview of several papers that have used them, see Chapter 1.

In this work we use digital evolution, specifically the evolution of self-replicating computer programs in the Avida Digital Evolution Platform (Ofria and Wilke, 2004), which has previously been used to conduct replay experiments. Yedid et al. (2008) employed this technique to investigate the re-evolution of traits following an extinction episode, while Covert III et al. (2013) used analytic replay experiments to study the importance of individual deleterious mutations in the evolution of complex traits.

We selected associative learning as a complex behavior to study potentiation. Associative learning is a non-trivial capability exhibited by most complex organisms. In digital evolution systems like Avida, it serves as a rare yet evolvable trait (Pontes et al., 2020). For an Avida organism to exhibit associative learning, it must be capable of sensing its environment, taking action, and storing information in memory. The evolution of associative learning has been studied via experimental evolution in both digital (McGregor et al., 2012; Pontes et al., 2020) and natural systems (Dunlap and Stephens, 2014; Mery and Kawecki, 2002), yet many questions remain about how it evolves. While more complex forms of learning are found in nature, associative learning remains an important building block for most others and insights about how it arises may be informative for understanding the broader evolution of intelligence, especially within digital contexts.

In this work, we begin to analyze how the likelihood of evolving a complex trait changes along a successful lineage. Using analytic replay experiments, we identified individual mutations that cause drastic increases in the potentiation of associative learning. We then analyzed those mutations and their mutational neighborhoods to begin characterizing how a mutation is potentiating. While these replay experiments are informative and useful for exploring counterfactual evolutionary possibilities, they are also computationally intensive. As such, we start by focusing on a set of case-study lineages to develop an initial framework for understanding how potentiation can occur.

Analyzing four successful lineages, we find that potentiation can increase suddenly, even due to a single mutation. Since these lineages were selected because they successfully evolved associative

learning, potentiation generally increases in each, though some decreases do occur. Potentiating mutations vary in initial effect, making them challenging to detect. Retrospective analysis allows us to identify them, however, and begin hypothesizing about the dynamics that allow these mutations to potentiate associative learning. This work demonstrates using analytic replay experiments for quantifying potentiation along a lineage and establishes baselines and techniques for future studies.

## 3.2 Methods

Here we describe the digital evolution system and experiment setup used to conduct this work.

### 3.2.1 The Avida Digital Evolution Platform

This work uses the Avida Digital Evolution Platform (Ofria and Wilke, 2004), which was described in Chapter 2. Here we use an early build of version 5.0 of Avida, currently under development as part of the Modular Agent Based Evolver 2 (MABE2) framework (<https://github.com/mercere99/MABE2>). Avida has previously been used to study both associative learning (Grabowski et al., 2010; Pontes et al., 2020) and historical contingency via replay experiments (Covert III et al., 2013; Yedid et al., 2008). Fundamentally, Avida is designed to have tools necessary to conduct work at the scale required for analytic replay experiments.

Avida genomes consist of assembly-like instructions that transfer data between registers, make basic comparisons, perform mathematical operations, *etc.* We use an extended instruction set that includes extra flow control and environment-specific instructions (Ferguson, 2023).

We used Avida populations on a 60x60 toroidal grid, resulting in a population cap of 3,600 organisms. Offspring are placed in a grid cell next to their parent, overwriting any existing organism in that cell; the parent organism is also reset. During reproduction, point mutations occur in offspring at a rate of 0.0075 per instruction, while single-instruction insertion and deletion mutations occur at a rate of 0.05 per reproduction. Organisms reproduce by executing the **Repro** instruction. To prevent organisms from immediately replicating, organisms must execute 1,500 instructions before the **Repro** instruction can be activated.

#### Associative learning

To test the evolution of associative learning, we created a simplified version of the Avida path following environment (Pontes et al., 2020). Instead of navigating 2D space, an organism exists in one of four states: *forward*, *left*, *right*, or the error state, *backward*. Organisms are given a **Sense** instruction, which will give them the nutrient cue of their current state. Using this nutrient cue,

organisms need to execute the appropriate instruction (one per state) to progress along the path. The *forward* and *backward* states have fixed cues (0 and -1, respectively), while, at birth, each organism is assigned random cue values for *left* and *right* in the range of  $[1, 10^6]$ . Organisms can genetically encode *forward* and *backward*, but must learn *left* or *right* in their lifetime to perfectly solve the task. Each path begins with one of four preset starting state sequences, chosen randomly for each organism at birth, followed by additional random states. The four preset paths are the “one-fixed turn” paths from (Pontes et al., 2020), where organisms are guaranteed to encounter a *left* state before a *right*.

If the organism is not in an error state and executes the appropriate instruction, they are rewarded and move to the next state. If an organism executes the wrong instruction (e.g., the **Left** instruction in the *right* state), it is penalized and placed in the error state. While in the error state, the organism must execute the **Backward** instruction to return to the previous state and be allowed to try again; it will be penalized for any other action. A cooldown is applied, however, such that executing the **Backward** instruction causes the organism to wait for the equivalent of 10 additional instruction executions. Organisms are scored based on the number of valid states they successfully traversed minus the number of incorrect moves made, with a maximum score of 300. Fitness is calculated as  $1.25^{\text{score}}$ , so each additional correct movement grants a 25% boost in fitness regardless of the total number of correct movements.

In this environment, optimal behavior requires associative learning in the form of imprinting. Since the paths are guaranteed to have a *left* state before a *right* state, the optimal behavior is to find and store the first positive cue value as the *left* cue. Combined with genetically-encoded *forward* and *backward* logic, storing and using the *left* cue is enough for organisms to identify the *right* cue through a process of elimination. Other possible behaviors involve error correction (assuming all turns are one direction, then correcting when wrong), bet-hedged learning (assuming more about the paths, e.g., that there are no instances of two lefts in a row), and various mixed strategies.

To categorize the behavior of a genotype, we evaluate it in 100 trials to ensure we observe how it performs in all four environments with different random cues. We then classify each of the 100 trials. Trials are classified as learning if the organism correctly handles greater than 90% of the states they were in, error correction if they always successfully navigated one turn state but not the other, and “low activity” if they failed to successfully navigate at least 25 states. To be categorized

as learning or error correction, *all* 100 trials of that genotype must be of that class. If one or more trials were low activity, the genotype was categorized as “bet-hedged learning” or “bet-hedged error correction”. If a genotype displayed at least one learning trial and at least one error correction trial, they were classified as “mixed bet hedging”. Finally, all remaining genotypes were categorized as “low activity”. This categorization system was used across all three phases of this work.

### 3.2.2 Experiment framework

To identify mutations that substantially increased the likelihood that learning evolved, we split the work into three phases (see Figure 3.1 for an overview). First, we seeded 200 initial parallel replicates in the associative learning environment with a default ancestor only capable of reproduction. Each replicate was given 250,000 updates, where one update is the time it takes for all organisms to execute 30 instructions, on average. We identified the most abundant genotype in each final population to represent the replicate and classified its behavior. We then extracted the “dominant lineage”, stretching from the ancestor to the representative genotype. Each step in the lineage corresponds to a change in genotype between parent and offspring, with clonal offspring occupying the same step as their parent. While a step is often a single mutation, it is possible that one step is composed of multiple co-occurring mutations.

To begin analyzing changes in potentiation, we ran exploratory replays replicates on four lineages capable of learning. For each, we seeded independent replicates for every 50<sup>th</sup> step in the lineage, up to step 1,000. All replay replicates evolved in the same associative learning environment, and replays were given the same number of updates as had occurred after that genotype first appeared (e.g., replays for a genotype that appeared at update 150,000 would be evolved for the remaining 100,000 updates). Potentiation was measured as the portion of replay replicates that evolved learning. Because replays were seeded with a single organism, some replay populations went extinct before ever reproducing and thus were not factored in (the minimum number of finished replay replicates from a given lineage step was 38, while three case study lineages had a minimum of 48).

While the exploratory replays provide an overview of how potentiation changed, we dug deeper by running targeted replays to further explore windows of increasing potentiation. Specifically, we found the 50- or 100-step “potentiation window” that sees the largest increase in potentiation in the exploratory results, and seeded additional replays for every step in that range. These targeted

replays were conducted identically to the exploratory replays, only they did not skip steps. Though computationally expensive, these replays illuminated the impact every genotypic change had on potentiation. Running 50 replay replicates per step still results in considerable noise, but we were able to identify mutations that clearly and substantially increased potentiation using these targeted replays.

We hand-analyzed algorithms in all potentiating mutations, here defined as single lineage steps that result in a potentiation increase of 25 percentage points or more. Further, we assessed genotype fitness in context of their lineage to identify if potentiation mutations were beneficial, deleterious, or neutral. Finally, we characterized the local fitness landscape of each genotype (one and two mutations out), measuring the presence and fitness of nearby genotypes capable of learning.

### **3.2.3 Data and software availability**

Both the data and the software used to conduct this work are available in the supplemental material (Ferguson, 2023). Analyses were conducted in the R statistical computing language (R Core Team, 2021) using the *dplyr* package to summarize data (Wickham et al., 2022). Data was visualized using the *ggplot2* and *cowplot* packages (Wickham et al., 2020; Wilke, 2020).

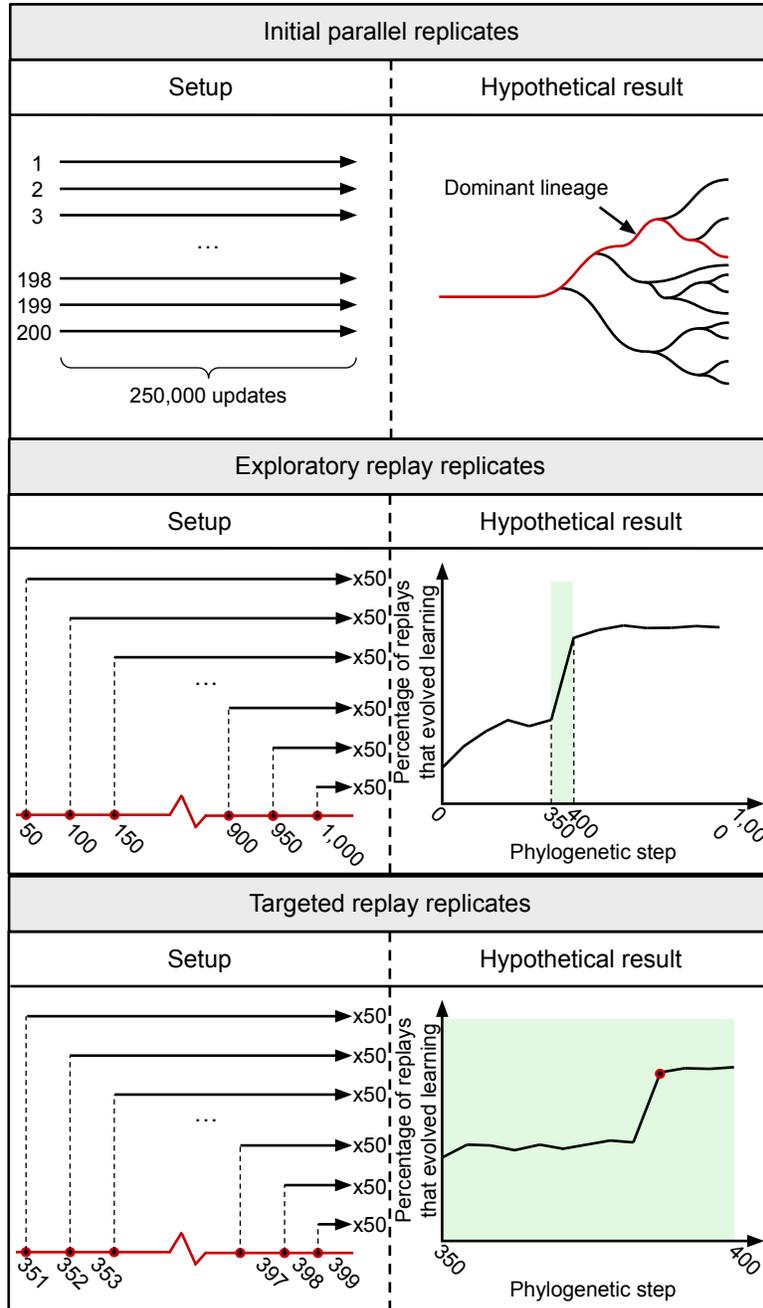


Figure 3.1: Illustration of the experimental design and hypothetical results. The top panes show the 200 initial parallel replicates seeded with the ancestral genotype and evolved for 250,000 updates. We extracted the lineage of the most abundant genotype in the evolved population (the dominant lineage), shown in red. Next, we conducted exploratory replays (middle panes) by launching replay replicates at regular intervals along focal lineages. These exploratory replays give a coarse-grained view of how potentiation changed over a lineage. We identified the window with the largest potentiation gain, shown as the shaded region. Finally, we ran targeted replay replicates for every step in this potentiation window. These fine-grained replay replicates show mutations that resulted in large potentiation increases (shown here with a red dot).

### 3.3 Results

Here we discuss the generation and analysis of the initial replicate runs, followed by the more detailed results from each of the four case study lineages.

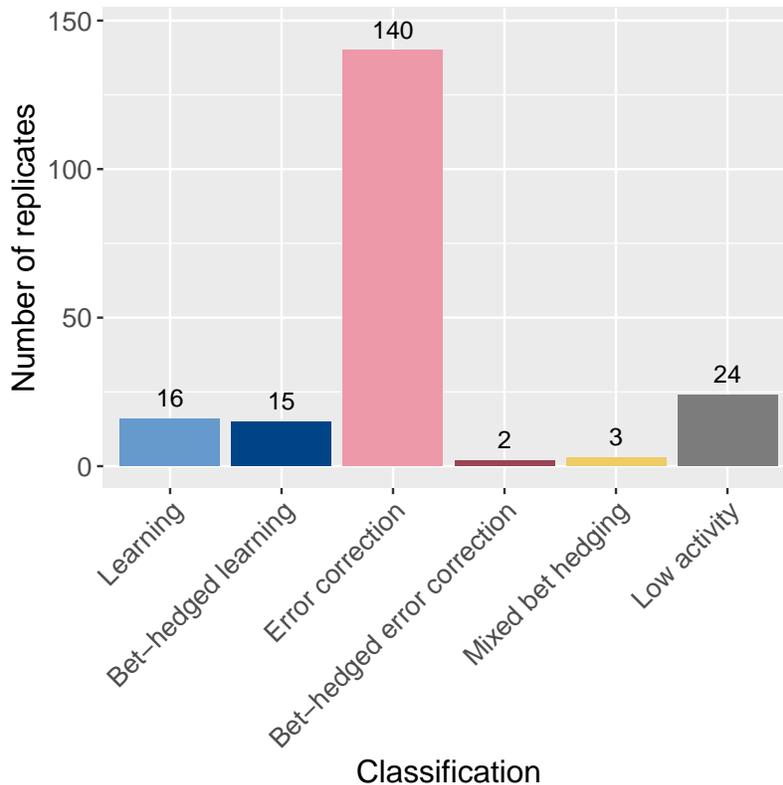


Figure 3.2: Behavior classification of the final dominant genotypes from the 200 initial parallel replicates.

#### 3.3.1 Evolution of learning in the initial replicates

In the first phase of this study, we evolved 200 independent replicates for 250,000 updates (about 3600 generations) in the associative learning domain, each starting with a default ancestor. The distribution of evolved behaviors is shown in Figure 3.2. Only 16 of 200 replicates exhibited associative learning. An additional 15 replicates evolved forms of bet-hedged learning, with two of those replicates gaining and then losing associative learning along their lineage. The majority of replicates relied on some form of error correction, either as a sole strategy (140), a bet-hedged variant (2), or as a fallback due to limited learning (3). Finally, 24 replicates failed to navigate enough states to categorize them, leading us to label them as “low activity”.

We analyzed all 16 replicates that evolved and maintained learning, identifying the length of their lineages from onset of evolution until learning stabilized, no longer showing further improve-

ment. Given the substantial computational power required for each time point studied, we performed replay experiments only on the shortest three such lineages (lineages A-C), plus the shortest lineage that exhibited error correction at some point in its evolutionary history (lineage D). Selecting these particular replicates to replay has the potential to bias the results, as discussed below.

### 3.3.2 Case studies of individual lineages

Below we present the results of the replay experiments performed on the four focal lineages and provide a step-by-step analysis of how key mutations altered both immediate fitness and evolutionary potential (potentiation). Where possible, we explained how these mutations altered the underlying algorithms. For each lineage, potentiation across both exploratory and targeted replays can be found in Figure 3.3.

#### Lineage A

Our first case study is one of the shortest lineages and it contains a quick jump to learning (at step 537) from low activity, with only a brief time spent in bet-hedged learning. Exploratory replays for this lineage revealed a stark jump in potentiation between 400 and 500 steps from the ancestor, which we call the potentiation window. From step 400 to step 450, potentiation increased from 20% of replicates to 44%, and increased again to 84% of replicates by step 500. Before this window, potentiation fluctuated around the original 8% found starting from the default ancestor. After the selected window, potentiation increased once more and then fluctuated between 94% and 100%.

With this in mind, we seeded 50 replays for each genotype in the potentiation window. Even with the noise due to a small sample size, we identified two mutations that conferred sizable increases in potentiation: steps 432 and 484. The mutation at step 432 brought potentiation above 50% for the first observed time in the lineage. Surprisingly, potentiation then decreased (on average) back to a local minimum of 20% of replicates at step 480. Finally, the mutation at step 484 substantially increased potentiation to 92%, where it stayed for all subsequent replays.

Even though the largest jump in potentiation occurred at step 484, learning did not appear in the lineage until step 537. That said, only steps 516 and 525 caused any change in behavior; all other interim mutations occurred in unexecuted regions of the genome. The potentiating mutation at step 484 made a key instruction in the main loop of the genome redundant. It had no immediate effect on fitness, but later (in intermediate step 516) allowed the redundant instruction to be replaced by

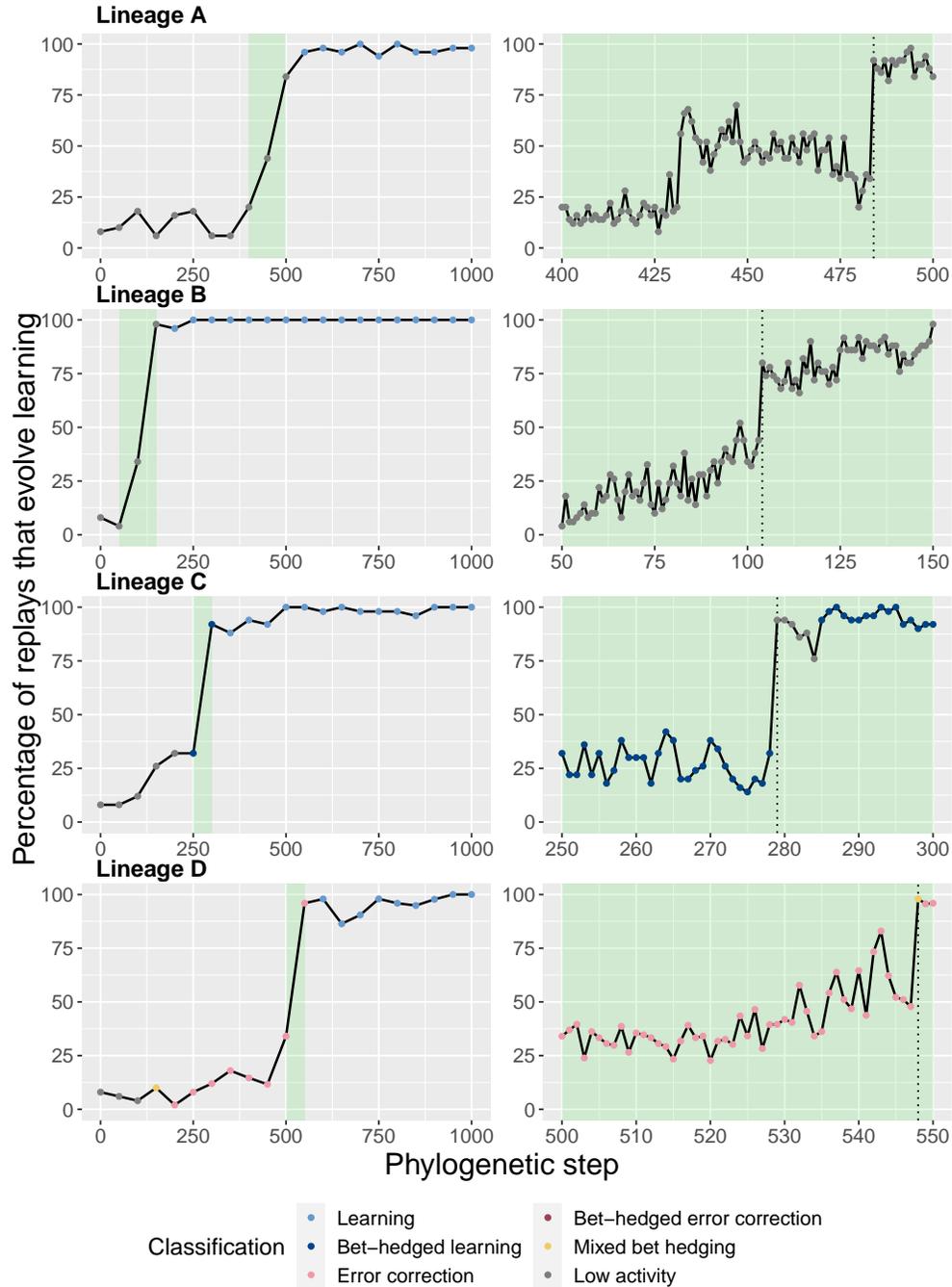


Figure 3.3: Potentiation of associative learning for all case studies, shown as the percentage of replicates that evolve associative learning when replayed from a given genotype. For each case study, the left plot shows the results of the exploratory replays. We identified a window of potentiation gain in each lineage, indicated by the shaded region. Within that window, we conducted targeted replays for every step along the lineage, shown on the right. The color of the points corresponds to the behavior exhibited at that step of the lineage. A dotted line in the targeted replays indicates the step that conferred the most potentiation.

a right turn that granted a small fitness increase as organisms could now navigate until they reached the second left turn. Step 525 further improved navigation, but used a comparison that made an unfounded assumption on whether the left or right random cue is larger. When the assumption was correct organisms were capable of learning the cues, however the assumption is only correct 50% of the time, so this genotype is categorized as bet-hedged learning. Finally, step 537 swapped that comparison with one that makes no assumptions about cue values, enabling the genotype to learn in all environments.

Looking at the local mutational neighborhood, the potentiating mutation at step 484 increased the number of two-step mutations that conferred learning from 2 to 9 (of approximately 56 million). Additionally, the fitness of the learning mutations in the local neighborhood increased by three or four orders of magnitude.

What about the earlier potentiation that was gained and then lost? The mutation that substantially increased potentiation at step 432 introduced a comparison that had no immediate fitness effect. This comparison remained unimportant until step 525 when it became integral in introducing bet-hedged learning. Neither step 432 nor its predecessor had access to learning within a two-step mutational range. Thus, it is likely that the potentiation comes from that comparison given that we observed it being utilized for learning later on.

Why then, did potentiation decrease between steps 432 and 484? At step 432 (and indeed before it), the algorithm had a section where if register B was non-zero, then B stored the cue associated with a left turn. While this information was likely to make the evolution of learning easier, it was unused at that time. As such, the mutations between steps 432 and 484 dismantled that machinery, requiring a replacement to be built before learning could evolve.

## **Lineage B**

Similar to Lineage A, this lineage transitioned from low activity to learning through a brief period of bet-hedged learning. Exploratory replays on this lineage reveal that learning was potentiated almost immediately; by step 150 potentiation had climbed above 95%, where it stayed for the rest of the lineage. As such, the potentiation window included steps 50 through 150.

Unlike Lineage A, the targeted replays reveal a general trend of increasing potentiation, with step 104 as a notable outlier. Mutations from steps 50 through 103 slowly increased potentiation from 4% to 44%, but the mutation at step 104 jumped to 80%. From there, another slow increase

continued to raise potentiation to a peak of 98% at step 150.

Learning did not appear until step 195, over 90 steps beyond the largest potentiating mutation. Given that 34 intermediate mutations altered the encoded algorithm, the mechanistic pathway to achieve learning is more complicated than can be broken down in this work. However, the potentiating mutation at step 104 modified the execution flow of the genome, which appears to have been essential for the later evolution of learning.

While two mutations occurred at step 104, only one caused a functional change: an instruction to swap data between registers was mutated to a left turn. Prior to this mutation, the genome encoded a left turn later on, after which the execution became trapped in an endless loop. The potentiating mutation was immediately beneficial; it allowed organisms to take the left turn earlier, which, in turn, allowed them to avoid the loop. As a side effect, a large portion of the genome that was previously executed was now skipped, and these instructions remained skipped when learning evolved 91 steps later. Looking at the local fitness landscape, learning was neither present in the potentiating step's landscape nor in the step before. We hypothesize that the potentiation came from the change in execution flow, and that skipping over those instructions avoided a pitfall and freed up execution time that may have been useful in evolving learning.

### **Lineage C**

Lineage C has the biggest single-mutation potentiation increase (64 percentage points), and that mutation was deleterious when it occurred at step 279 along the lineage. Learning later appeared at step 305.

The potentiating step mutated a no-operation instruction into a conditional flow control instruction. At step 278 the genotype was capable of bet-hedged learning, but the mutation at 279 knocked out all instances of learning, reclassifying the behavior as “low activity.” The next step restored some fitness, and then step 285 interacted with the mutation at 279 to not only restore fitness, but to dramatically improve it. Ultimately, the potentiating mutation at step 279 allowed the algorithm to more precisely discriminate between the left and right cues. Prior to the mutation, a less-than comparison was used, which only functioned correctly in 50% of instances, specifically those where the *left* cue was less than the *right* cue. The potentiating mutation switched to an equality comparison, which alleviated the assumption and initiated the transition from bet-hedging to learning.

Interestingly, the potentiating mutation lowered both the number and average fitness of learning mutations available in the local fitness landscape,

### **Lineage D**

Of the four replicates we analyzed, Lineage D is the only one that evolved error correction before learning. Like the earlier lineages, the exploratory replays show that almost all potentiation comes from a single window. In this case, potentiation grew from 34% of replicates to 96% between steps 500 and 550. Targeted replays are especially noisy for this lineage, but generally show an increase in potentiation, especially in the latter half of the window. The largest jump in potentiation occurred at step 548, near the end of the window. Several prior mutations also showed notable potentiation increases, but in each case later mutations appeared to counteract them. Specifically, steps 542 and 543 appear to have higher potentiation than the points around them, but potentiation dipped back below 50% before the largest jump at 548.

Out of all four lineages, D has the fewest steps between the largest potentiating step (548) and the first appearance of learning (556). At the time, the potentiating mutations at 548 caused no discernible change in fitness even though they increased potentiation by 50 percentage points. Two mutations occurred at step 548: a point mutation swapped a flow control instruction for a math instruction and an insertion mutation added a comparative conditional instruction into the main execution loop. At step 548 the genotype encoded a naive error correction algorithm: after setup, organisms could always handle *right* states, but always failed *left* states, recovered, and then continued. A mutation at step 556 swapped a sensing instruction with a math instruction, and this combined with the prior comparison instruction from step 548 to allow the organism to move left when needed and shifted the behavior to learning. The local fitness landscape supports the idea that the comparison instruction was useful to the evolution of learning, as the potentiating mutation increased the number of learning genotypes in the two-step mutational neighborhood from under 800 to over 100,000.

Looking back at the apparent false start at steps 542 and 543, it is not clear what algorithmic changes these mutations conferred. These steps did, however, alter the set of learning behaviors that fell within the local mutational neighborhood. At step 541, there were only 324 two-step mutations that conferred learning, and only three of those resulted in a substantial fitness increase (a metabolic rate  $> 10^{15}$ ). After steps 542 and 543, that number rose such that over 900 two-step

mutations could confer learning, with over 500 resulting in a substantial fitness increase (including over 200 that reached a merit  $> 10^{25}$ ). We may be unsure of the exact effects of these mutations on the mechanics of the algorithm, but the changes in the local fitness landscape are profound.

## 3.4 Discussion and Conclusion

### 3.4.1 Potentiation can rise suddenly

We have documented several cases where single mutations dramatically increased the probability of associative learning later evolving. Of the four lineages analyzed, each had a single step in the lineage that resulted in a substantial increase in potentiation (ranging from 36 to 64 percentage points). Indeed, two of the lineages had an additional potentiating mutation that resulted in an increase of over 30 percentage points. Looking only at exploratory replays, each lineage has a 50-step window that resulted in a potentiation increase of at least 40 percentage points.

While four lineages are insufficient to make any strong claims, these results demonstrate that it is *possible* for single mutations to drastically increase potentiation, and provide compelling evidence that they may, in fact, be common. In Lineages B and D, however, we do also observe regions with smaller, incremental increases in potentiation. Further studies are clearly necessary to more fully understand the general patterns and processes by which potentiation rises across different representations and environments.

### 3.4.2 Potentiation can decrease along a successful lineage

In two of the lineages we analyzed (A and D), we see evidence of potentiation decreasing over spans of the lineage. With only 50 replay populations per lineage step, our results are noisy and it is difficult to isolate what is occurring during these periods of potentiation decline. While we were unable to identify any “anti-potentiating” mutations with effects as large as the positive potentiation mutations, it is possible for a single step in a lineage to greatly decrease potentiation.

Since we limited our analyses to runs where associative learning arose in the original replicate, we did not expect a preponderance of anti-potentiating mutations, but were intrigued to see evidence of them, even if at low effect. These same analytic replay experiments could be applied to lineages that failed to evolve the focal behavior, to see if potentiation of that behavior experiences sudden drops. Similarly, our replays targeted windows with substantial increases in potentiation; other windows would be more likely to include decreases. Finally, failed replays from starting points with otherwise high potentiation must have failed for a reason; they too could be used as a likely source

(albeit more artificial) of anti-potentiating mutations.

### 3.4.3 Potentiating mutations can appear innocuous when they first occur

We analyzed the mutational step in each of the four lineages that conferred the greatest increase in potentiation. Of those four mutational events, two were neutral, one was deleterious, and one was beneficial. Even among these few replicates, there is no obvious pattern in the properties of potentiating mutations. Of the two neutral mutations, one made an instruction redundant while the other added a conditional instruction that had no effect when it was initially introduced. The deleterious and beneficial mutations both caused the execution flow to loop back earlier than it did before. Additionally, the number of mutations between the potentiating mutation and the appearance of learning varied wildly between lineages, ranging from 8 steps up to 91. The potentiating mutations in these four lineages are unique, and at the current time there is no pattern emerging among them. So, how are these mutations any different from other mutations? Untangling this mystery could be critical for predicting evolutionary outcomes or accelerating adaptive evolution.

### 3.4.4 We can identify *how* a mutation is potentiating

There are many mechanisms by which a mutation could facilitate the evolution of associative learning. For example, the mutation could provide a building block that is helpful to perform the task. But for a mutation to be potentiating it must notably increase the probability of associative learning appearing in the future. Any change, no matter how helpful, that was already likely to occur would not be considered potentiating. Indeed, it is the earlier mutations that made that change so likely that would be potentiating. Of course, those mutations are also more challenging to identify.

We have three different hypotheses for how a mutation could be potentiating: (1) It is the initial move into a genetic neighborhood with associative learning, (2) It is a shift into a genetic neighborhood with a more valuable version of learning, or (3) It is a “gateway” mutation that unlocks a beneficial pathway to learning, even though learning is not in the immediate genetic neighborhood.

Across the potentiating mutations we analyzed, we have found evidence for each of these hypotheses. The largest potentiating mutation in Lineage D supports Hypothesis 1, as it is the first time in the lineage that learning is only one mutation away. The main potentiating mutation in Lineage A and the earlier potentiation mutations in Lineage D support Hypothesis 2 as both cause drastic increases in the fitness benefit of learning mutations in the two-step neighborhood.

Finally, Lineages B, C and the early potentiating mutation from Lineage A all provide support for Hypothesis 3. The mutations from Lineages A and B both have *zero* learning mutations in their two-step neighborhoods. Interestingly, Lineage C sees a *decrease* in the number and fitness of learning mutations in the local neighborhood.

Hypothesis 3 has many possible mechanisms by which it may work. For example, new traits may produce a single, clear, beneficial pathway of improvements to follow. Alternatively, a new building block may open a larger region with many different ways of evolving associative learning. Finally, the mutation may actually damage existing functionality or remove existing interactions that were impeding further evolution. While all three hypotheses have some support, future work can begin to uncover if a certain hypothesis is seen more often, what conditions might result in each scenario, or if additional analyses are needed to truly characterize these potentiating mutations.

### 3.4.5 Outlook

This work is only an early step, focused on developing techniques and expectations for performing fine-grained analyses of replay experiments. Next, we must expand beyond four lineages, to collect broader, more systematic replay data, automating as much of the process as possible. We conducted this study on associative learning in Avida, but the underlying techniques must be examined broadly in other environments and substrates to ensure that our results are not unique to Avida or the evolution of associative learning. Within the current study system, there are many questions that remain unanswered: We focused on large *increases* in potentiation, but are there more obvious signals associated with *decreases*? How much of the noise that we see in our data is due to limiting ourselves to 50 replicates, and how much of it is due to actual shifts in potentiation with each mutation? What does potentiation look like in replicates that fail to evolve learning? Finally, it would be valuable to compare the specific evolutionary pathways the different replays take. Do they follow the same trend or do they differ? This would allow us to understand if, for example, a potentiating mutation funnels evolution in a fixed direction.

Ultimately, these analytic replay techniques provide us with a tool for examining evolution in a prospective fashion, not just the retrospective approach that we are traditionally limited to. They will allow for the development of new evolutionary theory and predictive capacity that will be invaluable, both for understanding how meaningful complexity is produced in the natural world and for improving evolutionary applications.

## Chapter 4

# A large-scale digital evolution study of historical contingencies in associative learning

Authors: Austin J. Ferguson and Charles Ofria

This chapter is in final preparation before submission.

This chapter directly extends the previous chapter. We expand from analyzing the potentiation dynamics of four case study lineages to replaying 50 lineages. With this larger dataset, we begin to conduct statistically powerful analyses of the the distribution of potentiation changes. In all 50 lineages we find a single-generation step in the lineage where the potentiation of associative learning increases significantly. Further, we verify that these increases are most often driven by a single mutation, though we do find evidence that multiple mutations can interact to increase potentiation. We find that these large potentiation increases are comparable to some, but not all, previous works on potentiation, and discuss those comparisons in detail.

### 4.1 Introduction

In Chapter 3 I conducted four case studies and analyzed the genetic potentiation of associative learning in the Avida digital evolution system. As in the previous chapter, here I use “genetic potentiation” to refer to changes in the genetic background that promote the evolution of a particular trait or behavior, in this case associative learning (Blount et al., 2008). We found that potentiation can increase over a short period of time, including a single mutation between parent and offspring. While I conducted a deep examination of the observed potentiation dynamics in Chapter 3, I limited myself to examining only four case study lineages. Here, I expand this work by again replaying the evolution of associative learning in Avida, but shifting from detailed case studies to aggregate data by analyzing the potentiation dynamics of 50 lineages.

My overall goal for this work is to provide a more comprehensive point of comparison for other studies on potentiation dynamics. While several previous studies of potentiation dynamics exist, they all analyze the dynamics of only one or at most a few lineages. By analyzing 50 lineages, here we shift from a few data points of potentiation measurements to full distributions backed by statistically powerful conclusions. With this work, I aim to accomplish two goals: 1) to allow us to examine previous studies in a new light, possibly identifying trends in potentiation dynamics across systems, and 2) to provide a benchmark of potentiation dynamics for future studies to compare

against.

There are multiple, but not many, papers that directly or indirectly study potentiation dynamics in ways comparable to this work. We are testing the effect of accumulated genetic backgrounds on the subsequent evolution of a particular trait. Thus, we can discuss other papers that use analytic replay experiments (Blount et al., 2018) to test the potentiation of a trait across different points in an evolved lineage. Alternatively, we can leverage mutant studies, where experimental evolution is conducted on organisms with and without a particular set of mutations, but an otherwise identical genetic background. However, we can only compare against studies that look at the binary categorical outcome of “did X trait evolve?” across replicates; we cannot directly compare to similar studies of continuous phenotypes or of evolvability (Travisano et al., 1995; Wagenaar and Adami, 2004; Woods et al., 2011). In practice, you can turn a continuous variable into a categorical variable (e.g., “did the replicates evolve a size greater than  $1\mu m^2$ ”), however such analysis decisions should be made before experimentation; post-hoc analysis can unintentionally inject our own bias, thus creating a statistical skew.

We find that, at both 50 lineage steps and single-step levels, potentiation for associative learning can increase substantially. Further, single-step increases in potentiation are likely to be driven by a single mutation, though we do see both epistatic and non-epistatic interactions between multiple mutations to alter potentiation. Finally, we compare our potentiation dynamics against other works that leverage analytic replay experiments to test the effects of actual accumulated genetic history (Blount et al. (2008), Chapter 3). Additionally, we compare against works that examine constructed one-step mutants (Jochumsen et al., 2016), or varying levels of shared history across evolutionary replicates (Meyer et al., 2012). We find that several, but not all, of these other studies have observed similarly large increases in potentiation. This result provides a baseline point of comparison for future potentiation studies, both *in silico* and *in vitro*.

## 4.2 Methods

This work extends the study described in Chapter 3. As such, most of the methods remain the same. At the core of these studies, we use the Avida digital evolution platform (Ofria and Wilke, 2004) to evolve digital organisms capable of simple associative learning in a path-following domain. Here we provide a brief outline of the methodology, highlighting areas where the methods differ from the previous chapter. All code, configuration scripts, and summarized data are available

online (Ferguson, 2023).

#### 4.2.1 Initial evolutionary replicates

First, we evolved populations capable of associative learning to provide lineages for us to replay. In order to accumulate 50 “learning replicates” (i.e., replicates that were capable of performing associative learning at the end of evolution), we ran batches of 500 replicates until we reached 50 learning replicates. This required 8 batches or 4,000 replicates total, for an ultimate success rate of 1.25%.

The main details of these initial replicates were identical to Chapter 3. We started each replicate with a single Avida organism capable only of reproducing itself, and we capped the population size at 3,600 organisms by using a 60x60 grid for our population. We again evolved each initial replicate for 250,000 Avida updates and then identified a representative genotype – the most abundant genotype at update 250,000 – and extracted the genotypic lineage leading to it from the original ancestor. We used this lineage from each replicate to perform our replay analyses. Mutation rates were kept the same as the previous study: per-site substitution mutations occurred at a rate of 0.0075, while insertion and deletion mutations occurred independently at a rate of 0.05 per reproduction event. Organisms were again able to replicate themselves via the `Repro` instruction, but they were still required to execute 1,500 instructions before being allowed to reproduce.

We evaluate the organisms on the same path-following task as the previous chapter. At each step in the path, we provided the organisms with an integer cue indicating which action they should take to maximize their fitness (move forward, turn left, turn right, or move backward). Organisms must interpret the current cue and execute the correct action, with each action being a single, atomic instruction that we added to the Avida instruction set. The “turn left” and “turn right” cues, however, have randomized values for each organism evaluation, and as such organisms must associate the cue values with the correct actions during their lifetimes (i.e., they must encounter the cues, empirically determine their meaning, and remember them) to perform optimally.

This path-following environment is the same as in Chapter 3 with two exceptions:

First, Chapter 3 used the pre-defined paths from (Pontes et al., 2020), where each organism was placed on one of four “one fixed turn” paths. This configuration guaranteed that organisms would always encounter the *left* cue before the *right* cue. For this follow-up work, however, we placed all of the organisms on completely random paths. We have thus removed the guarantee of seeing one

cue before the other, substantially increasing the difficulty of the learning task.

Second, we have refined the definitions of metabolic rate and fitness. An organism’s score is calculated the same way as before: organisms gain a reward of +1 for each correct movement and a  $-1$  for each incorrect movement. Metabolic rate, however, is now calculated as  $2^{\text{score}}$  instead of  $1.25^{\text{score}}$ . This new calculation means that each additional correct movement doubles metabolic rate and each incorrect movement halves metabolic rate. This change creates a steeper gradient – increasing performance will result in much higher metabolic rate (and thus fitness), but this also makes performance decreases much more deleterious. While metabolic rate is used for determining how often an organism executes instructions during the evolutionary run, here we calculate fitness specifically as metabolic rate divided by the reproduction time of the organism – thus organisms can increase their fitness by improving performance or by reproducing faster.

#### 4.2.2 Exploratory replays

As in Chapter 3, our goal is to identify how the genetic potentiation of learning changes over evolution, as evidenced by the course of the focal lineages. To accomplish this, we split our analyses into two types of analytic replay experiments (Blount et al., 2018): 1) exploratory replays that provide a coarse-grained overview of how potentiation changes over a lineage and 2) targeted replays that identify individual “potentiating” steps in a lineage. To calculate the potentiation level for associative learning in a given genotype, we simply replayed evolution from that genotype 50 times and recorded the percentage of replicates that evolved associative learning. We maintained the same stopping criterion for all replays: evolution would end after a lineage experienced 250,000 total updates (*e.g.*, for a genotype that appeared at update 150,000, replays would evolve for the remaining 100,000 updates).

In the previous work, replay replicates started from a single organism – the genotype along the lineage whose potentiation we were measuring. However, this technique resulted in a few replay replicates that went extinct for various reasons, such as being unable to reproduce if a specific cue sequence was encountered. Here, replays again consisted of 50 evolutionary replicates, but this time we started each replay replicate with a full clonal population of the genotype being tested. This change resulted in all replay replicates completing normally, and has the side effect of reducing the potential for adaptive momentum in our replay replicates (Chapter 5).

We conducted exploratory replays for all 50 replicates that evolved learning. Here we began

our exploratory replays in the same way as before: we identified the step in the lineage that first exhibited learning and replayed that genotype. In Chapter 3, we then started at the ancestral genotype and replayed every 50th step in the lineage until we reached this first learning genotype. Here, after replaying the first learning phenotype, we then iterated *backward*. We would take 50 steps backward (toward the ancestor) along the lineage and run another batch of replay replicates. We repeated this process until we reached a genotype that resulted in potentiation  $\leq 20\%$ . The results of the previous chapter demonstrated that this backward traversal would likely capture the largest increases in potentiation, as all four increases were greater than 20 percentage points. Therefore, the only way that we would miss the largest increase in potentiation would be if there was substantial gain in potentiation followed by substantial loss, which we saw no support for in the previous chapter. This scheme allowed us to avoid replaying the earliest genotypes in the lineage, and since they would have taken the longest to run (more updates per replay), this change substantially reduced the computational resources needed to conduct our exploratory replays.

In addition to our 50 learning replicates, we ran exploratory replays for ten replicates that did not evolve learning. We replayed two randomly-sampled replicates for each of the other five final behaviors (bet-hedged learning, error correction, bet-hedged error correction, mixed bet hedging, and low activity). As these replicates never evolved learning, we resorted to conducting exploratory replays as in Chapter 3: starting at step 50 in the lineage, we replayed every 50<sup>th</sup> step along the lineage. These non-learning replicates were replayed to give us a baseline of how potentiation might change in “unsuccessful” replicates. We recorded both the potentiation of learning and the potentiation for the other behaviors in these additional replicates.

For each replicate (both learning and non-learning), we found the 50-step window with the largest gain in potentiation. We then conducted a two-tailed Fisher’s exact test between the potentiation before and after this window to determine if the potentiation gain of the window was significant.

### 4.2.3 Targeted replays

As in the previous chapter, after conducting our exploratory replays we used that data to go a step further: replaying individual lineage steps. The goal of these targeted replays was to identify the “potentiating step” – the parent-offspring step in the lineage that conferred the largest increase in potentiation.

The exploratory replays provided insight into how potentiation changed over 50-step windows, and we used this information to construct a set of windows for targeted replays. Our goal was to capture as much potentiation gain as possible while remaining computationally feasible. We initialized the set with the 50-step window with the largest potentiation gain as identified by the exploratory replays. However, Chapter 3 showed that there are occasionally multiple windows (usually adjacent) that have similar increases in potentiation, so the potentiating step could be in one of these other windows. To catch these instances, before conducting targeted replays we recursively examined the windows adjacent to the those in the selected set. If an adjacent window had  $\geq 90\%$  of the potentiation gain of the maximum per-window gain, and the final potentiation of the window was above 10%, we then added this window to our set. Of the 50 replicates, the set contained only one window for 46 replicates, and the other four replicates contained exactly two windows.

We then conducted targeted replays by running 50 replay replicates for every genotype in the selected set. The “potentiating step” was then identified as the step in the lineage with the largest absolute increase in potentiation from the previous step. Similar to the exploratory replays, we found the potentiating step for each of the 50 learning replicates and then tested the significance of the potentiation difference. We again performed a two-tailed Fisher’s exact test, this time between the potentiating step and the step before it in the lineage.

#### 4.2.4 Post-replay analyses

After conducting the targeted replays and identifying the maximum potentiating step of each lineage, we then conducted two additional analyses to further disentangle the details of each potentiating step.

First, we re-analyzed the genotypes just before and just after the potentiating step in each lineage. For each pair of genotypes, we evaluated them on the same set of 100 paths. This allowed us to perform a paired Wilcoxon test to look for significant differences in fitness between the original genotype and the mutant. Since fitness effects (e.g., beneficial, deleterious, slightly deleterious, etc) are often loosely defined, this technique allows us to discuss these fitness differences in terms of significance and effect size, while making it possible to cleanly identify mutations that have exactly zero effect on fitness.

For the second analysis we analyzed the replicates with more than one mutation in the poten-

tiating step. By calculating the edit distance (Wagner and Fischer, 1974) between the genotypes before and after the potentiating step, we isolated the  $N$  mutations in the potentiating step. With the individual mutations identified, we ran 50 replay replicates for each of the  $2^N$  possible combinations of mutations for replicates where  $N > 1$ . We then applied a logistic regression model to test the significance of the contributions for the mutations and all possible interactions. This analysis allows us to identify whether the potentiation gain was driven by one of the  $N$  mutations or by a combination of mutations.

### 4.3 Results and Discussion

Here we detail the results of all experiments and place them in the broader landscape of historical contingency literature.

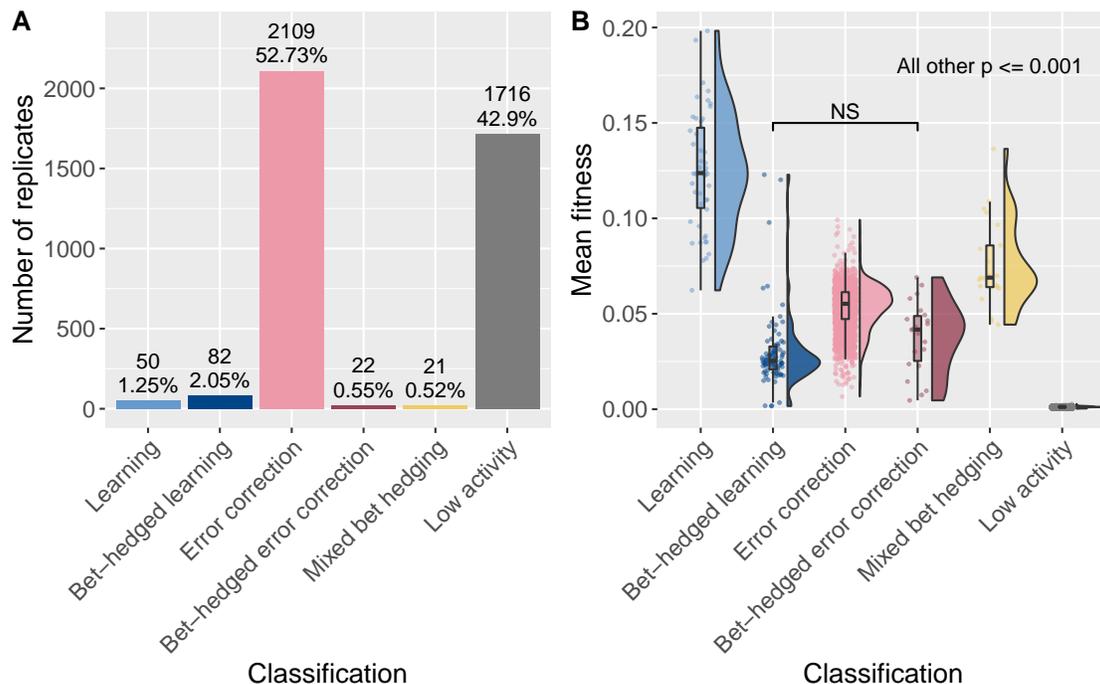


Figure 4.1: Subplot A) Columns show the number of initial replicates whose representative genotype evolved each type of behavior. Above each column, the numbers show the number of replicates that evolved that behavior and that count as a percentage of all 4,000 replicates. Subplot B) Raincloud plots show fitness of all 4,000 initial replicates' representative genotypes, grouped by evolved behavior. All pairwise comparisons are highly significant ( $p \leq 0.001$ ; Pairwise Mann-Whitney with Bonferroni correction) except for the one marked pair (bet-hedged learning and bet-hedged error correction did not have a significant difference).

### 4.3.1 The evolution of learning is exceedingly rare on random paths

As shown in Figure 4.1A, associative learning evolved in only 50 of 4,000 replicates (1.25%). This is considerably lower than the probability of evolving associative learning in previous work, with 8% of replicates evolving learning in Chapter 3 and up to 14% in the “two fixed turns” environment of (Pontes et al., 2020). However, these higher success rates were observed in environments that guaranteed the order of one or more turns at the start of each path. In this work we have removed those guarantees entirely; the next step in the path is always uniformly chosen at random among a left turn, right turn, or a step straight ahead. Thus, we must compare these results to the “random start” experiments in (Pontes et al., 2020). In their work, zero of 50 replicates evolved learning under such conditions. This work therefore marks the first time that associative learning has evolved in Avida without any guarantees about cue ordering. It should be mentioned, however, that this work is more lenient with the `Move Backward` instruction, placing the organism directly back on the path instead of moving them one tile backward and requiring them to reorient on their own, as in (Pontes et al., 2020).

While learning rarely evolved, error correcting behaviors evolved in the majority (53%) of replicates. Similarly, most of the other replicates (43%) were classified as “Low Activity” as, even at the end of the evolutionary replicate, they failed to correctly identify at least 25 cues in any of 100 trials. The remaining replicates evolved bet-hedged behaviors, with bet-hedged learning, bet-hedged error correction, and mixed bet hedging each present in less than 3% of final evolved populations.

Finally, we see that organisms that perform learning have significantly higher fitness than organisms performing any other strategy (Figure 4.1B, all  $p \ll 0.001$ , pairwise Mann Whitney with Bonferroni corrections for multiple comparisons). Mixed bet hedging has the next highest fitness, falling between its two constituent parts, learning and error correction. Bet-hedged learning and bet-hedged error correction both have lower fitness, on average, than their non-bet hedged counterparts. Finally, “Low activity” has very low fitness, as expected. These fitness values show us that learning was not rare because it was a poor solution to the problem, but rather due to the rugged nature of the underlying fitness landscape.

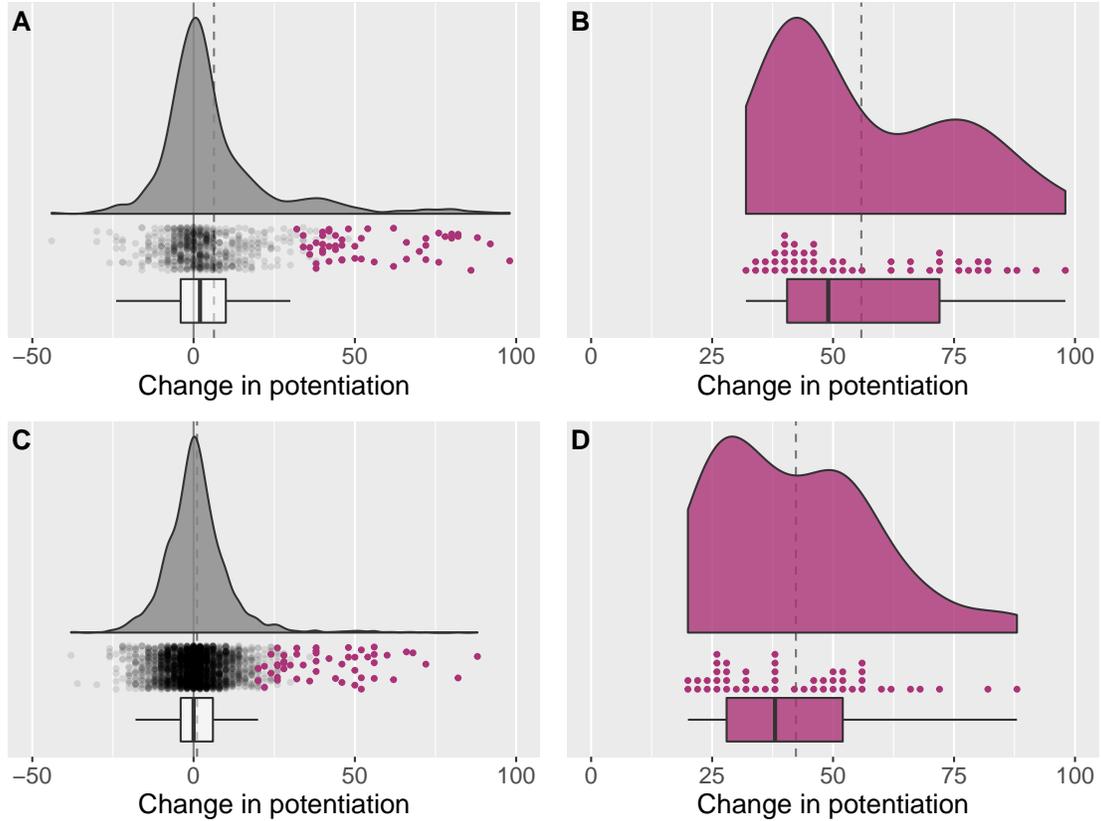


Figure 4.2: Aggregate data for the exploratory replays (A,B) and targeted replays (C,D) of learning replicates. Subplots A and C) Raincloud plots showing the distribution of potentiation changes for the cumulative effects across all replayed windows (A) or the individual effects of all replayed lineage steps (C). The solid vertical line shows zero change and the dashed vertical line shows the mean of the potentiation changes. For each of the 50 replicates, a purple point shows the maximum single-window (A) or single-step (C) potentiation increase. These 50 points are then shown in isolation in (B) and (D). Subplots B and D) Raincloud plots showing the distribution of maximum single-window (B) or single-step (D) potentiation changes for each replicate (one point for each of the 50 learning replicates). The dashed vertical line shows the mean of these changes.

### 4.3.2 Substantial increases in potentiation are common

By number of replays conducted, this work is the largest study of potentiation via analytic replay experiments to date. Here we detail the key finding of this work: large increases in potentiation are not only possible, but are *common* in this system. This result is supported at multiple levels, as we detail below. For all 50 replicates, figures showing potentiation over time for both exploratory and targeted replays are available in the appendix.

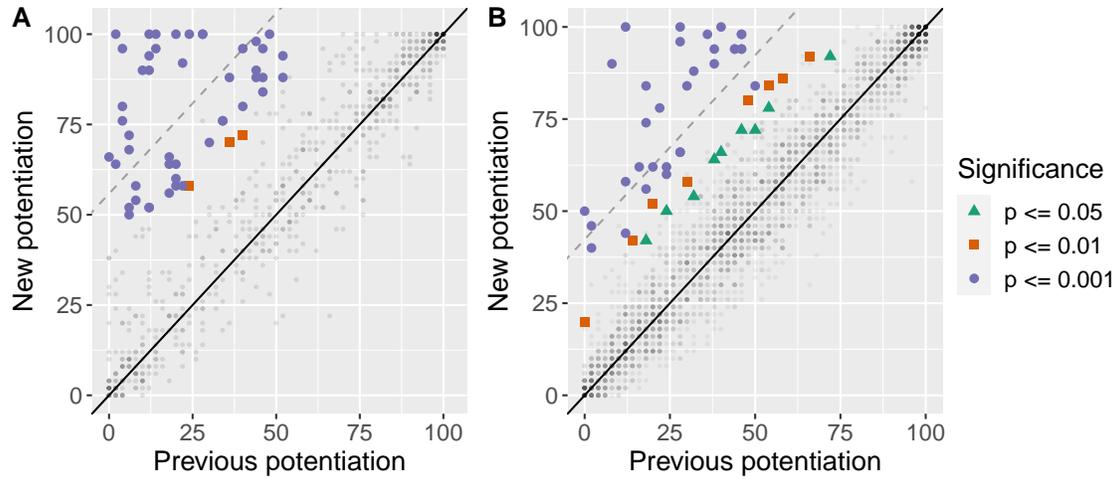


Figure 4.3: Scatter plot showing the potentiation before and after each exploratory window (Subplot A) or lineage step (Subplot B). The color and shape of the 50 main points show significance of the potentiation change (Fisher’s exact test). Semi-transparent points show the rest of the per-window (A, 10% opacity) or per-step (B, 5% opacity) potentiation changes. The line shows  $y = x$ , or no potentiation change. Potentiation gain is the vertical distance to this line. The dashed line shows the maximum potentiation gain of each replicate, averaged across all 50 replicates.

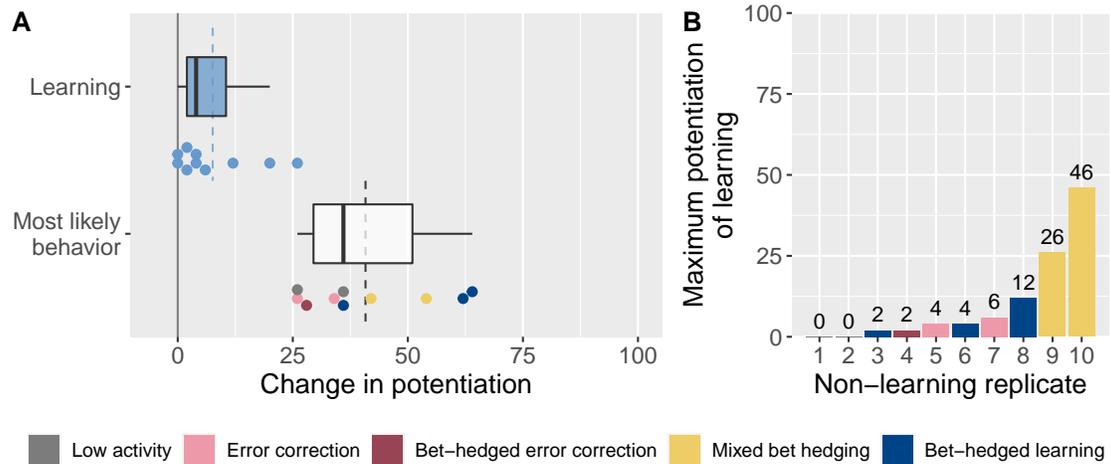


Figure 4.4: Subplot A) Distribution of the maximum per-window change in potentiation for learning (top) and most likely behavior to evolve (bottom) of each of the 10 replayed non-learning replicates. Colors for the bottom distribution indicate the most likely behavior for each replicate. Vertical dashed lines show the mean for each group. Subplot B) The maximum potentiation level of learning observed in each of the 10 non-learning replicates, sorted.

### All replicates have windows of significant potentiation gain

Exploratory replays show us that, on average, the 50-step windows we replay show little potentiation gain (mean  $\approx 6$ , median = 2 percentage points) (see Figure 4.2A). However, we see that

these replays also cover a wide range, from -44 (i.e., 44 percentage points of potentiation *loss*) to +98 (i.e., effectively full gain in potentiation in only 50 lineage steps).

When we consider these 50 learning replicates, we can also examine the maximum per-window potentiation gain of each replicate. These maximum gains are highlighted in Figure 4.2A and shown in isolation in Figure 4.2B. When only considering these maximum per-replicate gains, we see a mean gain of approximately +56, a median gain of +49, and a range of [+32, +98] percentage points. Figure 4.3A shows all potentiation gains as a combination of their previous and new potentiation values. We see that potentiation increases significantly for all 50 learning replicates, even with only 50 replays each ( $p \leq 0.01$ , Fisher's exact). Indeed, 47 of these replicates are highly significant ( $p \leq 0.001$ ). Thus we have further evidence that potentiation of learning can increase considerably in a relatively short amount of time.

### **Significant single-step increases in potentiation exist in all replicates**

While we see these large increases in potentiation over 50-step intervals, we can also look at smaller timescales. Indeed, the targeted, single-step replay results tell a similar story. Figure 4.2D shows the potentiation gains at *all* replayed steps. Here, the mean is approximately +1, the median is 0, and the range is [-38, +88] percentage points. Thus we see a similar trend to the exploratory replays: the bulk of the distribution residing near zero, but with a short tail in the negative direction and a long tail in the positive direction.

Focusing on only the maximum potentiating step for each lineage (Figure 4.2D), we see a mean of approximately +42, a median of +38, and a range of [+20, +88] percentage points. Immediately we see that, in many replicates, a single step in the dominant lineage can account for 50 percentage points of the potentiation of associative learning. However, not all replicates see these huge jumps; multiple replicates have a potentiating step of less than 25 percentage points. When we examine the individual potentiating steps (Figure 4.3B), we see that all of these changes are significant. Of the 50 replicates, 32 were significant at  $p \leq 0.001$ , nine more at  $p \leq 0.01$ , and the final nine additional replicates at  $p \leq 0.05$ .

### **Non-learning replicates see windows of potentiation increase in other behaviors**

We also see evidence of large gains in potentiation of behaviors other than learning. Figure 4.4 details the results of our ten non-learning replicates that were replayed. Of these 10 replicates, the single-window potentiation gains for learning were relatively small, with a mean increase of

+7.6, median of +4, and a range of [0, +26] percentage points. These low values are not surprising, however, as these lineages did not ultimately evolve learning. The potentiation for learning was never actualized and eventually lost. It should be noted that, even so, a potentiation gain of 26 percentage points over 50 lineage steps is substantial, which is highly significant ( $p \leq 0.001$ , Fisher's exact). Indeed, we see that one of the ten replicates reached a maximum learning potentiation of 46% at one point in its evolutionary history, though seven replicates never crossed more than a 6% chance of evolving learning (Figure 4.4B).

While the non-learning replicates did not see large increases in the potentiation of learning, they did see large increases in the potentiation of other behaviors. Figure 4.4A also shows the largest per-window potentiation increase in the most likely behavior at the end of evolution. Here we see a mean of +40.8, a median of +36, and a range of [+26, +64] percentage points. All of these differences are significant ( $p \leq 0.01$ , Fisher's exact), and eight of them highly significant ( $p \leq 0.001$ ). Thus, the large increases in potentiation that we have observed are *not* limited to the evolution of learning. Our expectation is that we would observe large single-step potentiation increases for these non-learning behaviors too if we conducted targeted replays on these lineages. Figures of how potentiation changed over time for all 10 non-learning replicates are available in the appendix.

### Comparison to other works

In addition to a deeper investigation into the underlying pattern associated with historical contingency, the original motivation for this work was to validate the findings of Chapter 3 and provide evidence as to their generality. That is, we wanted to test if the large per-window and per-step increases we observed in that work were common or if they were mere flukes. Indeed, we can now confirm that both the per-window and single-step potentiation increases from that chapter are common in this particular system. The maximum per-step potentiation increases we observed in the previous chapter (range of [+36, +64] percentage points) fall comfortably within the range that we observe here ([+20, +88] percentage points).

It is more difficult to compare these results to other works, however. While several works have leveraged replay experiments to study the potentiation of a particular trait or behavior, only one has applied replays to multiple points in a single lineage (Blount et al., 2008). Blount et al. replayed multiple time points of frozen *E. coli* from the Ara-3 strain to empirically test the potentiation of citrate metabolism. The largest jump in potentiation they found was  $\sim +13.3\%$ , shifting from 0/30

replicates at generation 31,500 to 4/30 replicates at generation 32,000 This is a much smaller jump in potentiation, measured over a much longer period of time compared to our results in this work as the tests were 500 generations apart.

While not directly pulling from an evolving lineage, Jochumsen et al. (2016) tested the effects of individual mutations on the evolution of colistin resistance in *Pseudomonas aeruginosa*. Via experimental evolution, they found evidence that two individual mutations (*phoQ* and *pmrB*) each potentiated the evolution of colistin resistance at the  $16 \frac{\mu\text{g}}{\text{ml}}$  concentration level. These two mutations each increased the potential to evolve this colistin level from a  $< 5\%$  chance to a  $> 70\%$  chance. These results directly align with our results – that a single mutation can drastically increase the probability of a focal behavior evolving.

For other works, Turner et al. (2015) performed similar analytic replay techniques, but never saw the focal results – the extinction of the  $\text{Cit}^-$  clade in the Long Term Evolution Experiment – in any of the replay replicates. Meyer et al. (2012) replayed combinations of Phage  $\lambda$  and *E. coli*. While some of their results are striking (e.g., the “all-or-nothing“ pattern of OmpF receptor targeting across certain treatments), the comparisons were across strains and thus not a direct comparison. In prior digital work, Covert III et al. (2013) showed that a single deleterious mutation could increase the evolution of a complex behavior in Avida (the EQU logic task) from 13/20 (65%) to 20/20 (100%) of evolutionary replicates, providing further evidence of substantial potentiation gain from a single mutation. Similarly, Bryson (2012) showed multiple instances of  $\geq 50$  percentage points of increase in short-term (always 10,000 updates) replays of EQU in Avida.

### 4.3.3 Increases in potentiation are often driven by a single mutation

For each of the learning lineages, we have identified the potentiating step; that is, the step in the phylogeny that conferred the largest increase in potentiation. Since Avida has a per-site rate for substitution mutations, and insertion and deletion mutations that can occur simultaneously, these lineage steps can consist of more than one mutation. Even so, 60% of the potentiating steps (30 of 50) are comprised of only one mutation. Of the remaining 20 potentiating steps, 14 steps contained two mutations, three steps contained three mutations, one step contained four mutations, and two steps contained five mutations.

Of course, the 30 potentiating steps that consisted of a single mutation must be driven by that particular mutation. For each of the other 20 potentiating steps (where multiple mutations

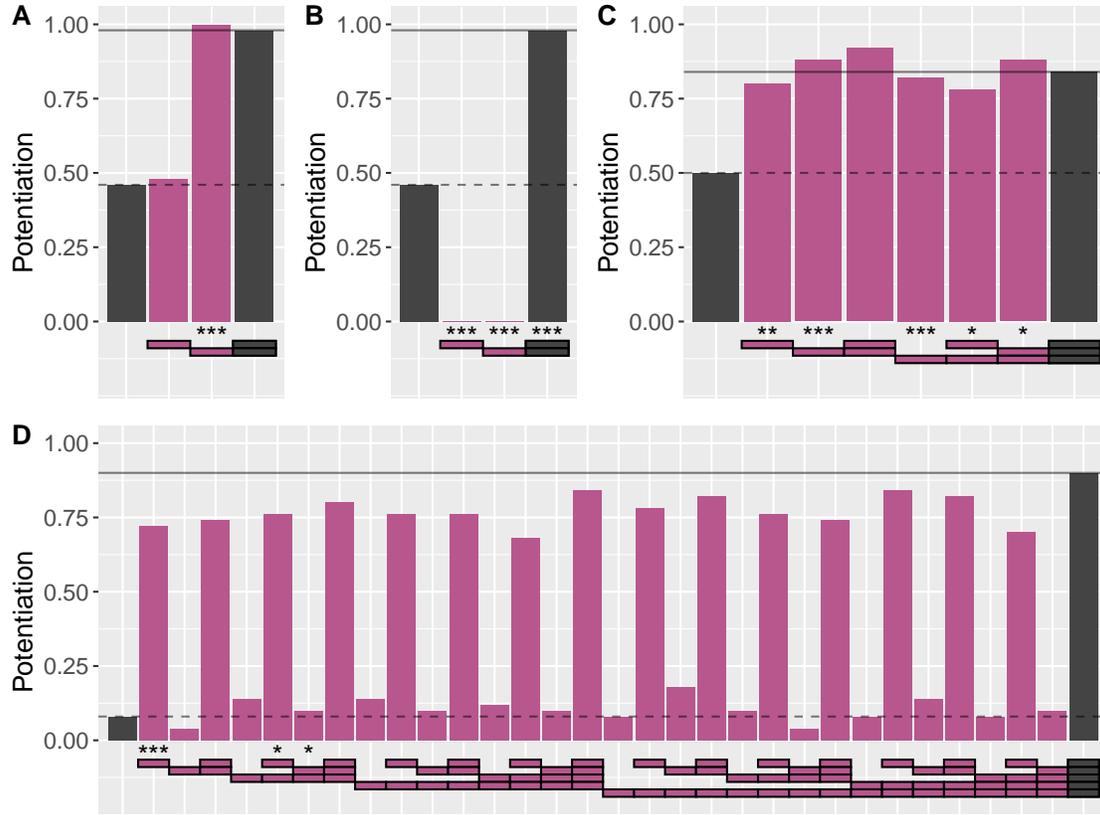


Figure 4.5: Selected results from the mutation split experiment. For each subplot, the rightmost column shows the potentiation immediately after the largest potentiating step in that replicate (also shown via solid horizontal line), and the leftmost column shows the potentiation immediately before that step in the lineage (also shown via dashed horizontal line). This potentiating step involved  $N$  mutations, where  $N$  is 2 for subplots A and B, 3 for subplot C, and 5 for subplot D. There are a total of  $2^N$  bars in each subplot, representing the  $2^N$  possible combinations of these mutations, as indicated by the small boxes below each bar. The left bars show potentiation with none of the  $N$  mutations presents, the right bars have all  $N$  mutations present, and the middle bars (purple) show the potentiation of every other possible combination of these mutations. Significance of each mutation combination in the logistic regression is shown under each column (\* for  $p \leq 0.05$ , \*\* for  $p \leq 0.01$ , \*\*\* for  $p \leq 0.001$ ). In subplots A and D potentiation is driven by just one of the mutations; in B potentiation requires both mutations (which are individually lethal); and in C all three mutations are independently-potentiating.

co-occurred), we ran replay experiments to split the mutations and determined that not all were necessary to increase potentiation. Indeed, in 14 cases we identified exactly one mutation that significantly increased potentiation on its own. The two steps involving five-mutation are included here, as a single mutation explained the potentiation gain in both cases. Data from an example of potentiation being driven by one of two co-occurring mutations can be seen in Figure 4.5A, and an

example of one driving mutation out of five total mutations can be seen in Figure 4.5D.

As discussed above, other work has shown that single mutations can have a large impact on potentiation. This has been shown in colistin resistance in *Pseudomonas aeruginosa* (Jochumsen et al., 2016) and in the evolution of EQU in *Avida* (Covert III et al., 2013). Thus we have instances where potentiation increases drastically with single mutations in both digital and natural organisms.

#### 4.3.4 Mutation interactions influence potentiation

While we are able to confidently say that many replicates saw potentiation increase with a single mutation, we also observed instances where potentiation changes appeared to require interactions across multiple simultaneous mutations, or other unexpected interactions. We identified one replicate where two mutations independently increased potentiation, and combined they accounted for the full step’s potentiation gain. In one of the steps containing three mutations, all three mutations independently increased potentiation significantly, but the combination of mutations did not further increase potentiation (Figure 4.5C). In two replicates we see strong sign epistasis; each step consisted of two mutations, and when analyzed independently these mutations proved to be lethal (and thus potentiation is 0), but when the mutations were combined the resulting mutants were not only viable, but potentiated (Figure 4.5B). This is akin to the case study of EQU potentiation in (Covert III et al., 2013), where two deleterious mutations (one nearly lethal) combined to actualize EQU. Finally, of the 50 replicates, two were found to have no significant mutation effects, and further replicates would be required to disentangle the effects of the mutations and their interactions.

#### 4.3.5 No real-time hallmarks of potentiating mutations were evident in an evolving population

As discussed in Chapter 3, while we can retrospectively identify potentiating mutations, we have yet to find strong indicators of these mutations *during evolution*. Any such consistent indicators would be invaluable for problem solving with evolutionary computation or other forms of applied evolution.

Our early hypothesis was that these potentiating mutations are likely to be deleterious, or at least neutral. Our reasoning was that if the mutations are beneficial, they would already be likely to be selected and thus they would already be factored into evolutionary potential. Of the four case study lineages in Chapter 3, we saw that one potentiating step was deleterious, two were neutral, and one was beneficial.

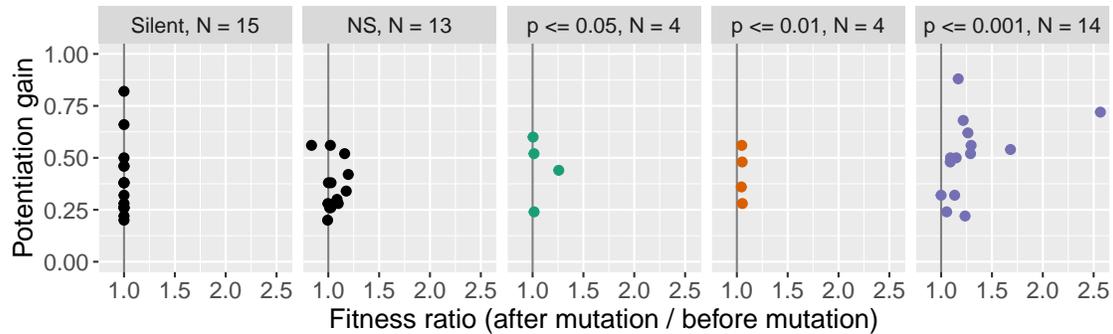


Figure 4.6: The fitness effect and potentiation gain of the potentiating step of all 50 learning replicates. To calculate fitness, both the potentiation step and the step before it were evaluated on the same set of 100 random seeds. We ran a paired Wilcoxon rank-sum test for these two fitness groups for each replicate, and divided and colored the results based on these significance findings. Non-significant points are split into those that are truly silent mutations (i.e., zero change in fitness across all 100 trials) and those that are not (fitness values changed, but not significantly). The fitness ratio, used for the x-axis, is calculated as the average post-mutation fitness divided by the average pre-mutation fitness. The vertical line shows a fitness ratio of one, which indicates no change in the average effect.

Here, we have refined our methods to compare the potentiating step to the step before it, evaluating both genotypes on the same set of 100 random paths to conduct a pairwise comparison (Figure 4.6). Using a paired Wilcoxon test, we find that in 28 of 50 lineages the potentiating step did not have a significant effect on realized fitness. Indeed, in 15 of these replicates, the two genotypes displayed *exactly* the same behavior – the mutations were truly silent, while in 13 changes to fitness occurred, but not enough to meaningfully change the distribution. The truly neutral mutations may be modifying instructions that are executed but do not directly effect the phenotype, or they could be cryptic genetic variation occurring in unexpressed regions of the genotype (Chapter 2). Of the remaining 22 replicates with significant changes to fitness, we find that 21 significantly increase fitness, and only one mutation was significantly deleterious. Further, our one deleterious mutation was significant ( $p < 0.001$ ) but not substantial, with less than a 1% change in fitness.

The fitness effects of the significantly beneficial mutations varied wildly, with many conferring small impacts on fitness. Two mutations had a greater than 50% increase in fitness, one of which had a fitness over 250% of the pre-mutation fitness. Thus, we have strong evidence against our hypothesis that potentiating mutations are likely deleterious. We observe that potentiating mutations are often neutral or nearly-neutral, though they can also drastically increase fitness. In a

genetic space as complicated as Avida’s, even highly-beneficial mutations are not guaranteed to be found, let alone sweep. Given the number of alternative possibilities, some of which may be highly advantageous themselves, even beneficial mutations can be highly potentiating. Indeed, even if deleterious mutations would open up new pathways through the genotype space, the low probability of them gaining substantial numbers in the population may eliminate the possibility of them being potentiating.

One final hypothesis for identifying potentiation mutations *as they occur* is to look for mutations that alter the overall behavioral phenotype. Of the 50 potentiating lineage steps that we analyzed, we found that 43 of 50 left the underlying behavioral phenotype unchanged. Of the seven that did change the phenotype, all were originally error correction behaviors. After the potentiating step, five then shifted to a mixed bet hedging strategy – they still performed error correction on some paths but were capable of learning on others. One more replicate evolved learning *with the potentiating step*, likely representing a low probability mutation that was actualized. The final behavior-changing mutation shifted the behavior from error correction to bet-hedged error correction. While this replicate might represent an interesting case – the degradation of behavior – the majority of potentiating mutations did not change the behavior and thus behavior changes are not a reliable indicator of potentiating mutations when they occur.

## 4.4 Conclusion

Here we have conducted a study of potentiation with the largest number of analytic replay experiments to date. By replaying 50 lineages of digital organisms that evolved associative learning, we have provided further evidence that potentiation can increase suddenly, with strong evidence that a single mutation can result in large increases in potentiation. We have also shown that potentiation can increase as a result of the co-occurrence of multiple mutations, with or without epistatic interactions. Finally, we were unable to find a signal for these potentiating mutations when they appear – as such, we are still only able to identify these mutations retrospectively.

### 4.4.1 Limitations and applicability to other systems

In both this work and Chapter 3 we primarily investigate the potentiation of one behavior, associative learning, in the Avida digital evolution framework. As such, we cannot claim that the potentiation dynamics observed in this work are globally applicable. Indeed, we can only claim these results as evidence that large, single-mutation increases in potentiation are *possible*, and that,

in certain systems, they might be common.

While this work is among the first to explore *patterns* in potentiation across many evolved lineages, it is far from the first to explore potentiation dynamics. Do the results of other studies match the large increases in potentiation we observed for the evolution of associated learning in Avida? In some cases it appears that short spans of evolution, and in fact even single mutations, confer substantial increases in potentiation (Covert III et al., 2013; Jochumsen et al., 2016). These increases are not universal, as other works have seen maximum increases in potentiation substantially lower than observed here (Blount et al., 2008). In conducting these comparisons, however, we must consider that the sample sizes for these other works are quite low and that may bias our comparisons. Additionally, all replay replicates in (Blount et al., 2008) evolved for the same number of generations regardless of the starting point, while we ensured that all replay replicates experienced the same number of Avida updates as the original lineage experienced after the sampled genotype appeared. We would expect the former scheme to observe substantially lower potentiation in replicates founded from earlier points in the population’s history. Finally, it is possible that the lineages in these other studies *did* see substantial increases in potentiation, but those particular changes were missed in those experiments.

If these potentiation dynamics truly differ, what might explain these differences? While Avida has a rich and complex genetic architecture compared to many other digital evolution experiments, these digital organisms are still vastly less complex than even the most simple natural bacteria. An Avida organism in this work, with our original genome length of 100 instructions, has roughly 9,000 possible one-step mutations that can occur. With a naive assumption that every nucleotide could mutate to any other nucleotide at every site in the *E. coli* genome (~4.6 million base pairs, Blattner et al. (1997)), we would expect over 13 million one-step substitution mutations alone. While this is only a rough estimate, it is clear that the likelihood of a particular mutation appearing is much lower in natural organisms, which can bias the potentiation changes. This difference only compounds as we consider the likelihood of multiple necessary mutations arising and fixing. Therefore, we expect that natural organisms will typically see lower levels of potentiation for equivalent genetic distances from the focal behavior.

Additionally, there is likely a difference in how the focal behavior arises and is classified. We are stringent in our system and only classify a behavior as learning if it is already well-refined, and thus

even the first organism in the population to be classified as having evolved learning is likely to have high fitness. In natural systems, biologists will often denote a non-trivial trait or behavior as present even if it still requires considerable refinement. For example, in the *E. coli* studies in the LTEE, citrate metabolism was not highly advantageous initially, but rather needed considerable refinement before it rose to prominence in the population (Blount et al., 2012). Thus, even genotypes capable of weak citrate metabolism may still have low potentiation for maintaining that trait in the longer term, as it may be lost before it is refined. Follow-up work to our study should consider analyzing the potentiation of all forms of learning, including bet-hedged learning and mixed bet hedging strategies that combine learning and error correction, to see if potentiation dynamics change. Further, learning is a terminal behavior in our system – the lack of open-endedness dictates that once learning evolves, it is unlikely to be replaced with a more effective survival strategy. Combined, these factors may create larger increases in potentiation, and should be dissected in future digital studies.

Finally, our system uses small population sizes (3,600 organisms) compared to natural systems (which are often in excess of tens of millions of organisms), and our replay techniques may also bias the changes in potentiation. By using a smaller population size, we increase the effect of genetic drift on our population and increase the likelihood that neutral or even deleterious mutations can spread before being purified. In fact, by replaying populations via a full clonal population of an genotype from the dominant lineage, we are resetting all population dynamics from the potentiation measurements. During normal evolution of a population, not only does a mutation need to occur, but the organism in which it resides also needs to successfully replicate at least once. By using clonal populations for our replay experiments, we ensure that every mutation establishes, eliminating possible intermediate potentiation values.

#### **4.4.2 Future work**

In this chapter we described our deep analysis of the potentiation dynamics behind the evolution of associative learning in Avida. First and foremost, future work should investigate potentiation dynamics in a range of other systems. Digital systems as simple as bitstring evolution problems could illuminate how the dynamics are affected by factors such as mutation rate and population size. Slightly more complicated systems could investigate characteristics of the genetic space such as connectedness, dimensionality, mutational operators, or crossover. Finally, more wet-lab studies of potentiation across a wider range of environments or in a more varied selection of natural organisms

can provide more real-world grounding for these results, though such systems are considerably more labor-intensive.

As we alluded to in Chapter 3, not only are we interested in identifying mutations that substantially increase potentiation, but we are also interested in the mutations that significantly *decrease* potentiation. In this work we replayed ten replicates that did not evolve associative learning, to see how their potentiation dynamics differed from “successful” replicates. Even within those ten replicates we observed one replicate that reached a height of 46% potentiation for associative learning. Extending this idea, if we have 50 replicates that evolved learning, we might expect 100 replicates to have reached 50% potentiation but only half of them to have succeeded in that coin flip. This conjures the question: do we see large decreases in potentiation similar to the increases observed in this work? Are there single mutations that doom a would-be promising lineage to an “unsuccessful” fate? For example, an immediately beneficial mutation might sweep through a population trapping it on a local optima and preventing it from following a pathway that would lead a population to higher levels of fitness. Identifying these anti-potentiating mutations in real time could potentially be as valuable as identifying potentiating mutations as they could allow us to slow the evolution of pathogenic or otherwise harmful populations. While additional studies would be needed to fully delve into these dynamics, early work should dive into this dataset. When conducting replay experiments, we are effectively quantifying the potentiation of *each* possible behavior. Thus we have non-learning potentiation data, but analyzing this data is outside the scope of this chapter.

Finally, additional work is needed to identify techniques that reduce the cost of performing replay experiments. This work has taken one step: iterating backward from the appearance of the focal behavior during the exploratory replay phase. Additional optimizations could include minimizing the original number of replicates per replayed time step and then running additional replicates for those that seem interesting. Finding more efficient ways of searching for potentiating steps is critical. We cannot directly perform a binary search for finding the largest potentiating step, as potentiation does not increase monotonically and thus large increases in potentiation can be hidden by decreases in the same window. However, there may be variations of binary search that will provide the maximum potentiating step within some margin of error or the maximum *sustained* increase in potentiation. In particular, using digital systems to find and verify these ideas might make future *in vitro* experiments more tractable to perform.

Overall, these results provide a useful benchmark for any future projects that use replay experiments to study historical contingency. As more studies of mutations are conducted, we can refine our ideas about how potentiation changes over the course of a population's history. As we progress, we can leverage our understanding of potentiation dynamics to better infer what happened in the evolution of life around us, better prepare for how populations will evolve under climate change, and better leverage evolution to solve technical problems.

## Chapter 5

# Predicting the unpredictable: Using replay experiments to disentangle how evolutionary outcomes are altered by adaptive momentum

Authors: Austin J. Ferguson, Charles Ofria, and Clifford Bohm

This chapter is adapted from a paper that has been peer-reviewed and accepted to appear in the proceedings of the 2024 Conference on Artificial Life.

In this work we use the lens of historical contingency and potentiation to provide a new vantage point in understanding the adaptive momentum effect. According to the adaptive momentum framework, when a population is in disequilibrium some advantaged organisms can buffer more deleterious mutations than usual. This increased buffering increases evolutionary exploration until equilibrium is reestablished. Here we replay the evolution of select populations evolving in an idealized environment: a one-dimensional sawtooth function. At each step we are able to quantify their potential to cross the next deleterious fitness valley. We find that populations experiencing disequilibrium, here caused by a selective sweep, see their potential to cross increase drastically with mutations near the leading edge of the sweep. Conversely, populations in equilibrium rely purely on chance to cross – only the last mutations before a valley cross confer an increase in potential. Finally, we also performed “shuffled” replays to demonstrate that the structure of the population, not just its genetic makeup, is key in increased valley-crossing potential in this system.

### 5.1 Introduction

In Chapter 1 we discussed how researchers are beginning to use analytic replay experiments to quantify how long-term evolutionary outcomes changed over the course of a population’s history. Indeed, in Chapters 3 and 4 we employed these replay techniques to identify key “potentiating” mutations in the evolution of associative learning in *Avida*. Here, we use these same replay techniques to understand the “adaptive momentum” effect (Bohm et al., 2024).

Previously, replay experiments have been employed to measure the potentiation of complex traits and behaviors, both in microbial organisms (Blount et al., 2008; Gupta et al., 2022; Jochumsen et al., 2016; Meyer et al., 2012) and in digital organisms (Chapters 3 and 4). Here we instead focus on the potentiation of a significantly simpler task: crossing deleterious fitness valleys in a one-dimensional sawtooth function. These fitness valleys require exactly six mutations to cross, which,

due to the one-dimensional nature of this system, must be acquired in order. Therefore, we already know that any potentiating mutations must be a subset of these six mutations, and thus we are not searching for them as we were in the previous two chapters. However, we do not know the *amount* of potentiation conferred by each of these six mutations. By comparing these potentiation dynamics of two categories of populations (those in disequilibrium and those in equilibrium), we can directly observe the effect that adaptive momentum has on evolutionary exploration and how these effects change over time.

### 5.1.1 Adaptive Momentum

The adaptive momentum framework suggests that periods of disequilibrium resulting from phenomena like selective sweeps and range expansions can enhance genetic exploration (Bohm et al., 2024). We use analytic replay experiments to investigate this effect in selective sweeps. Consider the appearance of a beneficial mutation in an asexual spatial population. If this mutation establishes and is sufficiently strong, it can trigger a selective sweep where the genotypes with the beneficial mutation come to dominate the population. During the sweep, there will be a boundary between individuals with the beneficial mutation and those without (wild type). If advantaged individuals along this boundary accrue relatively small deleterious mutations, they may still have a combined fitness benefit over the wild type. Thus, individuals along the leading edge of the sweep will have an increased potential to accumulate deleterious mutations that, in turn, increase the potential to explore genetic space and facilitate genetic discovery across fitness valleys. Adaptive momentum persists until the wild type is eradicated and equilibrium is reestablished. However, if during fixation a new beneficial mutation is discovered, the state of disequilibrium will persist, thus extending the “momentum window” (the period of adaptive momentum).<sup>1</sup>

In the work presented here, we create an ideal environment for adaptive momentum by using a one-dimensional spatial population evolving on a rising sawtooth fitness function. We use replay experiments to measure how the state of the population affects the potential to cross the next fitness valley. Replay experiments show that adaptive momentum increases the potential for populations

---

<sup>1</sup>Adaptive momentum describes how disequilibrium during evolution can result in periods of increased mutation buffering. Such disequilibrium can be caused by several conditions, including sweeps, range expansions, and increases in carrying capacity in both spatial and well-mixed populations. The adaptive momentum framework further considers how the increased potential for mutational buffering can also affect large scale evolutionary rates via increased genetic exploration and subsequent genetic discovery.

to cross fitness valleys, an effect that diminishes over time.

Using simple assumptions about the structure of the leading edge of a selective sweep, we generated a predictive model of the potential for valley crossing during adaptive momentum. While this estimation is highly accurate for early steps into the fitness valley, it loses accuracy when the leading edge is deeper in the valley. The time required for a population to reach deeper mutations in the fitness valley is likely also increasing the number of stochastic events that cause differences from the assumptions of the model. Finally, by shuffling organism positions in our population snapshots we can disrupt population structure. This shuffled analysis allowed us to verify that the organization of the population, not just its genetic composition, is vital to valley-crossing potential. Overall, this work refines the framework proposed by adaptive momentum while advancing the methodology of analytic replay experiments as a tool for studying historical contingency, exposing sections of both that are ripe for further study.

## 5.2 Methods

### 5.2.1 Evolution system

In order to isolate the adaptive momentum effect and keep computational costs feasible, we used a minimal agent-based evolution model where each organism’s genome consists of a single integer, with value  $x$ . Fitness is based in a repeated “sawtooth” function, similar to the one used in (Bohm et al., 2024), to create a landscape with regular fitness peaks separated by valleys. Each organism is assigned a quality score ( $s(x)$ ), and then that score exponentiated to determine fitness ( $f(x) = 10^{s(x)}$ ). Score is calculated using the sawtooth function:

$$s(x) = \frac{x - (x \bmod w)}{w}G - (x \bmod w)D$$

where  $G$  is the gain per peak,  $D$  is the decrease per step into the valley, and  $w$  is the valley width (number of mutational steps from one peak to the next). Effectively, the score function calculates the quality of the highest peak achieved and subtracts the cost of the current mutational step into the valley. We used  $G = 1.0$ ,  $D = 0.05$ , and  $w = 6$ . This means that peaks appear every six steps; in this case at  $x = 6$ ,  $x = 12$ ,  $x = 18$ , and so forth. We refer to these peaks by their height, so those three peaks would be  $p_1$ ,  $p_2$ , and  $p_3$ , respectively. We refer to the steps between peaks by the prior peak and an offset (*e.g.*,  $x = 17$  is  $p_2 + 5$ , the last step before  $p_3$ ). Figure 5.1 illustrates this

sawtooth function and the relevant peaks.

Crossing a valley increases an organism’s fitness by a factor of 10, and each step into the valley reduces fitness by  $\sim 12\%$ , relative to the previous peak. We selected these drastic fitness differences to clearly demonstrate the effect. While this degree of selective benefit may be rare in biological systems, cases such as antibiotic resistance can show fitness increases of this magnitude (Gullberg et al., 2011).

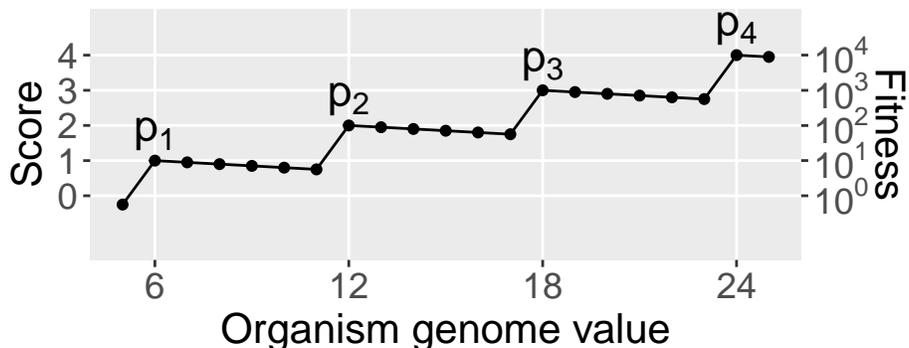


Figure 5.1: The sawtooth function used in this work, with the four peaks mentioned in this work labeled. Score,  $s(x)$ , and fitness are both shown, with fitness being  $10^{s(x)}$ .

We evolved populations of 512 organisms in a one-dimensional spatial population – a single line of organisms – that did not wrap. This structure maximizes fixation times and makes selective sweeps easy to track, as they could only move left or right. The population evolved via discrete, non-overlapping generations. To fill positions in the next generation, we performed rounds of spatial roulette selection, each involving three organisms: the organism in that position in the previous generation and its two immediate neighbors (unless the organism is on the edge, in which case the roulette is between only two organisms). We then copied the selected individual to produce the offspring, with a 0.0125 chance to mutate; mutations either increment or decrement the offspring’s genome value by one. Selective sweeps are thus limited to advancing one position in the population per generation, limiting the growth rate and therefore the fixation time of a beneficial mutation.

### 5.2.2 Experiment design

We conducted this work in three stages: 1) we validated that our model demonstrates adaptive momentum; 2) we ran “benchmarking” data to create an expectation of how potentiation changes during a momentum window; and 3) we conducted analytic replay experiments to observe changes

in potentiation in evolved lineages. Here we outline these experiments in more detail.

### Experiment I: Model Validation

To ensure that our model could produce the adaptive momentum effect, we replicated the primary experiment from (Bohm et al., 2024) comparing crossing times in populations starting from equilibrium versus those starting during the disequilibrium of a previous selective sweep. We ran 500,000 evolutionary replicates, where each replicate started with 512 organisms at  $p_1$  and ran for 5,000 generations. In replicates where  $p_2$  was discovered, we recorded the generation of discovery and extended the run duration for another 5,000 generations beyond the point of discovery. If  $p_3$  was discovered before time ran out, we recorded this time as well. This methodology produced two time distributions: time to first crossing and time between first and second crossing, both capped at 5,000 generations.

### Experiment II: Benchmarking

The adaptive momentum framework posits that populations in disequilibrium experience an increased rate of adaptation. In spatial populations experiencing a selective sweep, the disequilibrium should manifest near the leading edge of the sweep. Specifically, adaptive momentum allows deleterious mutations to accumulate within the advantaged subpopulation along the leading edge. This temporarily expanded mutant cloud increases genetic exploration, accounting for an observed increase in the rate of adaptation.

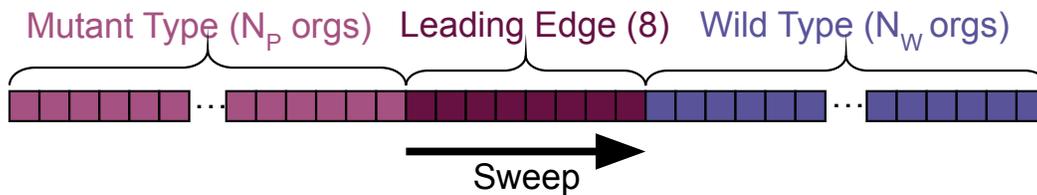


Figure 5.2: Starting layout for experiment II populations, with three clearly-defined sections:  $N_P$  “post-sweep” mutant organisms at peak  $p_2$  (left),  $N_W$  “wild type” organisms at peak  $p_1$  (right), and 8 “leading edge” organisms at a treatment-specific position in the valley past  $p_2$  (middle).

We created idealized scenarios to study the dynamics of adaptive momentum as they unfold. Each population had organisms on  $p_2$  sweeping across organisms on  $p_1$ , with a well-defined leading edge (Fig. 5.2). Experimental treatments used all combinations of how far into the next fitness valley the leading edge started (from  $x = p_2$  to  $x = p_2 + 5$ ), and how far the sweep had progressed

across the population (from  $N_P = 0$ , at the beginning of a sweep, to  $N_P = 504$  at the end.) We ran each replicate for  $768 - N_P$  total generations; the reduced number for larger  $N_P$  was used to make the comparison fair, subtracting off the minimum time it could have taken to establish  $N_P$  post-sweep organisms. For each condition, we measured how often replicates crossed the next valley to reach  $p_3$ .

We recorded the number of replicates that successfully crossed to  $p_3$  in each treatment. This measurement provided an expectation of a population’s potential to cross the valley *based only on the initial state of the leading edge*. Additionally, we also ran “shuffled” controls under otherwise identical conditions, but where we removed population structure (and thus the notion of a leading edge) by randomly shuffling the organisms before each evolutionary replicate.

### **Experiment III: Analytic replays of evolved lineages**

Finally, we focused on the treatment from Experiment II that represented the start of a selective sweep; that is, no post-sweep organisms ( $N_P = 0$ ) and a leading edge that just made it to  $p_2$ . We ran 500 replicates under these conditions, saving snapshots at each generation, allowing us to perfectly recreate the population at any time point. We randomly selected 10 replicates that failed to reach  $p_3$ , 10 random replicates that did reach  $p_3$  (but not further), and all four replicates that crossed two valleys to reach  $p_4$ . For each of these 24 replicates, we performed 1000 analytic replays at every fourth generation, restarting evolution from a given time point with different random seeds to investigate the role of chance in determining *the distribution of potential evolutionary outcomes* (Blount et al., 2018). Next, we selected one representative sample from each of the three categories to study at high resolution, replaying 10,000 replicates from every generation.

Using these replays, we recorded the probability that a replicate would cross the valley to  $p_3$  or  $p_4$  at each time point. These data allow us to calculate how the crossing probabilities changed over time. We also ran 10,000 equilibrium replicates and replayed representative replicates in the same manner. Finally, just as in the shuffle benchmarking experiment, we performed “shuffled” replays on disequilibrium replicates that crossed a single valley. Specifically, we shuffled the population before starting each replay to measure the role of the population’s spatial organization in crossing potential.

### 5.2.3 Data and software availability

All code, analyses, summarized data, and figures not included in this work are available in the supplemental material (Ferguson, 2024). The model was built using the Modular Agent Based Evolver version 2.0 (MABE2) (<https://github.com/mercere99/MABE2>). All analyses and plots were generated using the R statistical computing language version 4.1.2 (R Core Team, 2021) and the ggplot2 (Wickham et al., 2020), dplyr (Wickham et al., 2022), Hmisc (Harrell Jr, 2020), tidyr (Wickham and Girlich, 2022) packages.

## 5.3 Results

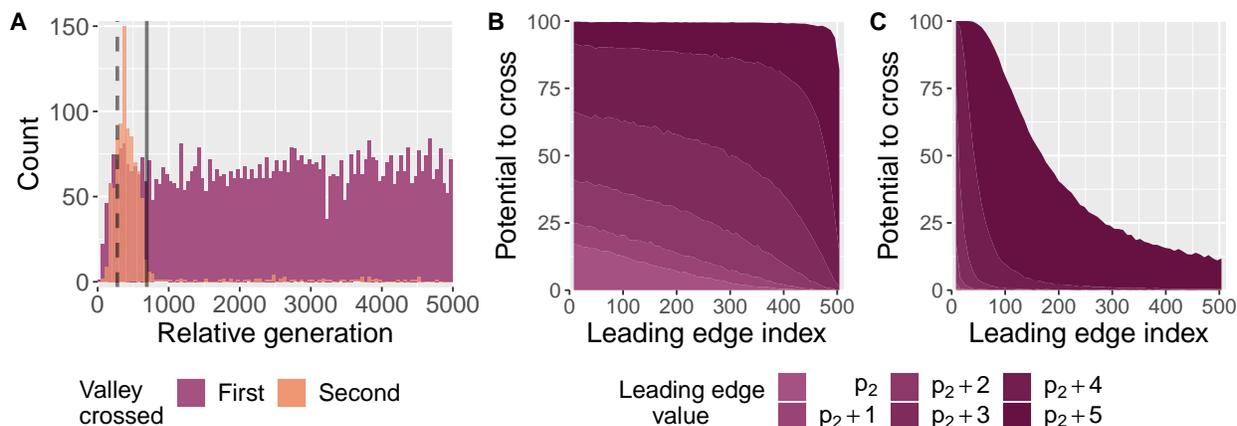


Figure 5.3: (A) Distribution of the number of generations required to cross valleys in the validation experiment. Relative generations refer to the elapsed time (in generations) for the first cross (purple), and from the first cross to the second cross (orange). The dashed and solid vertical lines show minimum and maximum fixation times, respectively. (B) Benchmarking data showing the potential of a leading edge to cross a valley, with a range of leading edge starting positions and values (see Fig. 5.2). Each point represents 10,000 replicates. (C) Shuffled benchmarking data. Each replicate is shuffled prior to evolution, otherwise identical to center plot.

#### 5.3.1 Validation of adaptive momentum effect

First, we measured the time it took for replicates starting from a full population of  $p_2$  to cross the valley to  $p_3$  and, when relevant, from  $p_3$  to  $p_4$ . Figure 5.3A shows the timing distributions of populations that crossed the first valley over the first 5,000 generations (purple) as well as the timing distributions of second crossings that occur within 5,000 generations of a first (orange). We see that the time of first crossing appear uniformly distributed across the 5,000 generations, while the second crosses are strongly skewed toward shorter time periods. Indeed, of 500,000 replicates, 6,485 crossed the first valley within 5,000 generations ( $\sim 1.3\%$ ) with a mean cross time of  $\sim 2,569$

generations and a median of 2,602 generations. Based on these values, we would expect roughly 84 replicates to cross twice ( $6,485 \times 1.3\%$ , or  $0.0169\%$  of all 500,000 replicates), but instead we see 902 replicates ( $\sim 0.18\%$ ) cross the second valley, a much higher rate than expected if the probabilities of first and second crossing were equal. In addition to a higher than expected rate of crossing, the mean cross time between first and second crossings is  $\sim 579$  generations and the median cross time is 401 generations, substantially lower than the first crossing times. Finally, when we consider only those second crossing times that occurred more than 1000 generations after a first crossing, we find that the rate of these second crossings is similar to the first crossing rate ( $\sim 1.23\%$ ). These results comport with the framework of adaptive momentum. They show that an initial beneficial discovery can quickly lead to additional discoveries (during the fixation period), but if the second discovery does not happen before equilibrium is reestablished, the rate of valley crossing is better predicted by the first valley crossing times.

### 5.3.2 Empirical benchmarks

The empirical benchmark data (Fig. 5.3B) illustrate how the initial state of the population affects the potential to cross valleys. As expected, steps further into the valley increase the probability of crossing, regardless of where the leading edge is. On the other hand, the probability of crossing decreases as the ratio of  $p_2$  organisms (mutant type) increases relative to  $p_1$  organisms (wild type) – as the selective sweep progresses, fewer opportunities remain for additional mutations to accumulate. While the potential to cross varies considerably with the type of organisms in the leading edge, these data clearly show that all conditions describing early sweep conditions (*i.e.*, having a leading edge and a significant ratio of the remaining population on  $p_1$ ) substantially increase the probability of crossing the valley compared to populations that are close to reaching equilibrium on  $p_2$ .

We use these results to provide a baseline prediction for the analytic replay experiments.

### 5.3.3 Analytic replay experiments

Of our 500 initial replicates to find candidates for replay experiments, four replicates (0.8%) crossed two valleys in the allotted time, 83 crossed exactly one valley (16.6%), and the remaining 413 did not cross any valleys (82.6%). We randomly-sampled 10 no-cross replicates, 10 single-cross replicates, and took all four double-cross replicates to run coarse-grained replays. All plots are available in the supplement (Ferguson, 2024).

From these coarse-grained replays, we selected one representative replicate from each category to re-run, conducting 10,000 replay trials at *every* time point to give a fine-grained view. For each replayed replicate, we show the potential of crossing the next valley over time, paired with a Muller plot (Muller, 1932) of the initial replicate that was replayed. Figures 5.4, 5.5, and 5.6 show the single-cross, double-cross, and no-cross replicates, respectively. In each plot, the replay results are overlaid on an image generated from the benchmark data. To generate the background image, we treat the benchmark data as a lookup function. When we start a replay replicate, we initialize the population using the snapshot from the target generation of the initial replicate. These snapshots show that the leading edge does not perfectly advance one position every generation; there are many generations where the leading edge either fails to advance or is pushed back one position by the wild type organisms. To make this comparison fair, we find the leading edge in that particular snapshot and then look up the corresponding expectation values from the benchmark data. This adjustment ensures that we are comparing against the correct benchmark data regardless of the motion of the leading edge.

Overall, the potentiation observed in all three replicates closely match the benchmark expectations. The potentiation occasionally increases or decreases suddenly; tracing these changes to the Muller plots typically shows that these events correlate to the gain or loss of a mutation at (or near) the leading edge at that time. For example, the two temporary peaks in potential in Figure 5.4, at roughly generations 250 and 375, can be directly traced to the leading edge temporarily dipping to  $p_2 + 4$ . The potentiation at any particular step in the valley decreases over time, as the selective sweep progresses and the adaptive momentum window closes. Figure 5.6 shows that, while the replicate had substantial potentiation at times (briefly above 50%), it failed to capitalize before the window closed. This failure was exacerbated by two leading mutations from  $p_2 + 3$  to  $p_2 + 2$ , the second of which dropped the potential from over 25% to under 10%, after which the population never recovers.

All three replicates show periods of potentiation higher than what our benchmark data would predict given their leading edge genomic value. The benchmarking data modeled a leading edge of eight organisms, but the Muller plots show that the leading edge grows and shrinks over time. The Muller plots show that underprediction is generally associated with either an expanded leading edge or an excess of individuals with lower fitness behind the leading edge. We hypothesize that these

underpredictions are generally observed in deeper steps into the valley because of the historical contingency required to reach that point (*i.e.*, to reach  $p_2 + 5$  the leading edge must have passed through  $p_2 + 4$ , which may still exist behind the leading edge).

For comparison, we also ran 10,000 replicates that started at equilibrium (*i.e.*, not in a momentum window). Of those, we replayed the one replicate that crossed twice, 10 randomly sampled replicates from 19 that crossed once, and 10 random replicates from the 9,980 that failed to cross. Figure 5.7 shows the replicate that crossed twice. The first crossing in this replicate occurred quite early, crossing the valley on generation 168. However, the potential before crossing is similar to all first-cross replicates: while the potential in an adaptive momentum window starts at  $\sim 17\%$ , the potential for replicates outside momentum windows starts at  $\sim 0.2\%$ . This low potential continues for the first 143 generations, followed by 16 generations with an average potential of  $\sim 1.8\%$ , four generations between 15% and 20%, and then the cross. This pure chance, “all or nothing” potentiation is not unique to this replicate; the same dynamic can be seen in the ten other successful replicates we analyzed (Ferguson, 2024).

Finally, we also show the potentiation of crossing the second valley, which in some cases is realized (Figs. 5.5 and 5.7), and other times is not (Figs. 5.4 and 5.6). As potentiation is probabilistic in nature, the potential to cross the second valley is the probability of crossing the first valley from the current state of the population times the probability of crossing a second valley from a naïve starting position (*i.e.*, the elevated potentiation of a population at the beginning of a sweep). We see this dynamic early in the replays – increases in first-cross potential are reflected, at smaller scales, in second-cross potential, with a much larger increase upon successful crossing of the first valley in Figures 5.4 and 5.5. This result is consistent with the adaptive momentum framework, which posits that the discovery of a new peak will initiate a new adaptive momentum window. Strikingly, this same dynamic holds true for Figure 5.7 – while the replicate did not start in a momentum window, the first cross creates a window which increases the potential of the second cross.

### 5.3.4 Shuffled population experiments

To test the importance of population structure for potentiation, we repeated the benchmarking and replay experiments with shuffled populations. In both cases, we kept the experiments identical except for an additional shuffle step: before starting a replicate, we shuffled the order of organisms in the population and then proceeded with evolution as normal.

The shuffled benchmark data in Figure 5.3C shows that only populations with a leading edge of  $p_2 + 5$  are able to maintain greater than a 10% chance to cross if the leading edge has swept more than half the population. Two differences can decrease crossing potential: 1) multiple leading edges can form, allowing faster fixation and 2) the eight leading-edge organisms are more likely to encounter higher-fitness mutant-type organisms and thus be purified faster.

A representative sample from the single-cross replicates is shown in Figure 5.8. The replay data indicate that potential to cross remains relatively low, never reaching a 20% chance, until a critical mass of nearly-crossed organisms skyrockets the potential from 9.2% to 100%. The earlier spikes in potentiation always correspond to the appearance of a single organism that is one step away from crossing the valley, which was then lost in the original population. These trends are consistent across all 10 replicates that were replayed (Ferguson, 2024).

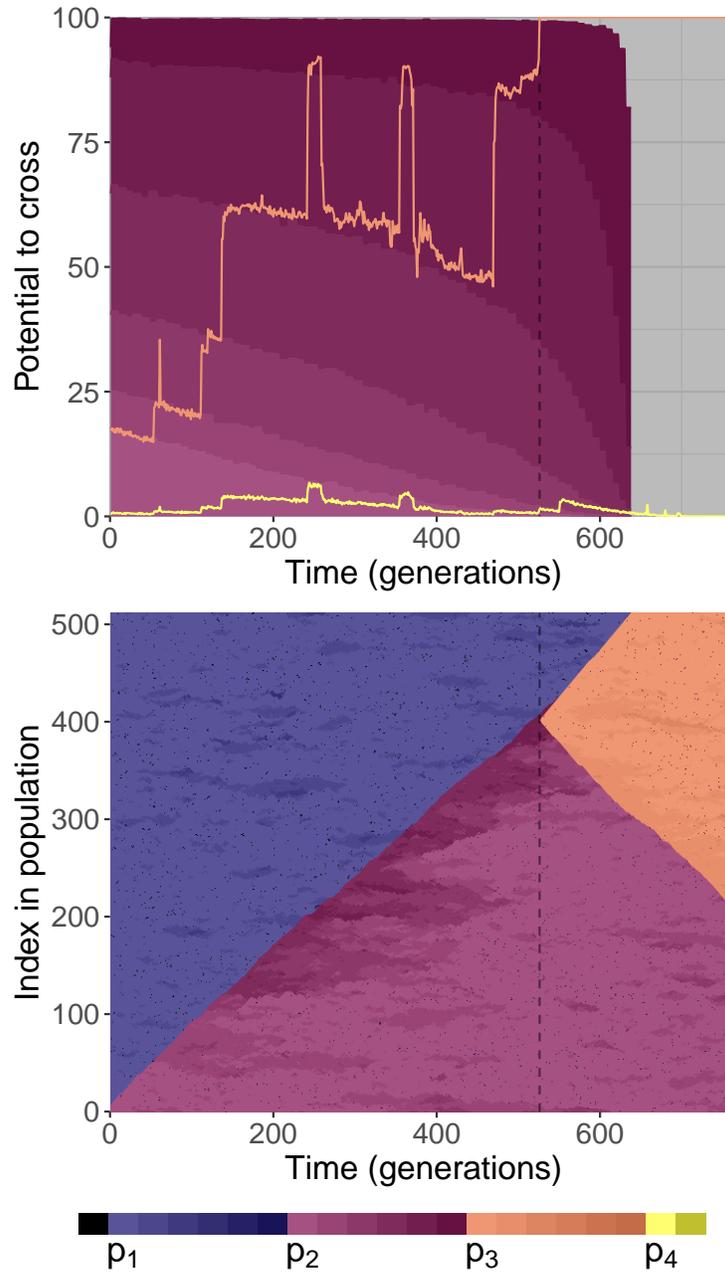


Figure 5.4: Top plot: Analytic replay data for a representative replicate that crossed one valley during a momentum window, overlaid on baseline data. Lines show the probability of crossing both the first valley (orange line; initially-higher) and the second valley (yellow line; initially-lower). The background shows the expected potential to cross (as shown in Fig. 5.3B); data here are shifted to align with the realized leading edge position. Bottom plot: A Muller plot of all organisms in the original 1D population over time. Dark colors show descent into valleys, with hue identifying the valley being crossed. In both plots, vertical dashed lines show initial valley crosses. Colors in the legend apply to all plots.

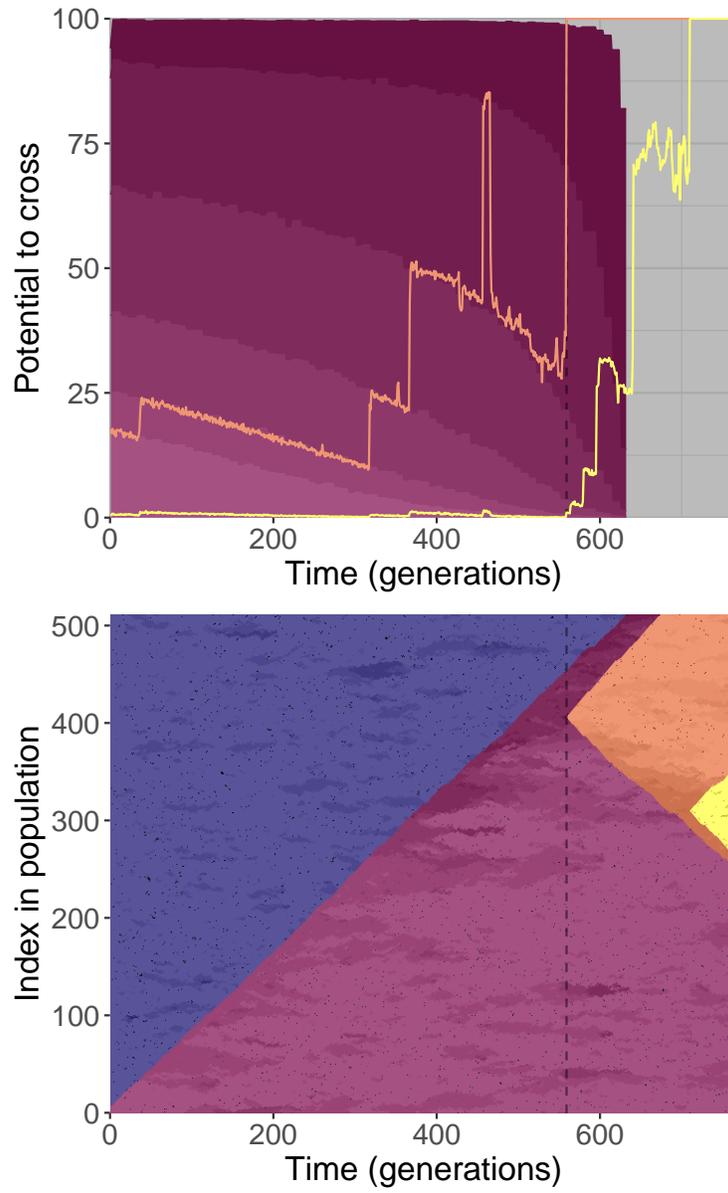


Figure 5.5: The analytic replay data for a representative replicate that crossed two valleys during a momentum window. See Figure 5.4 for details and legend.

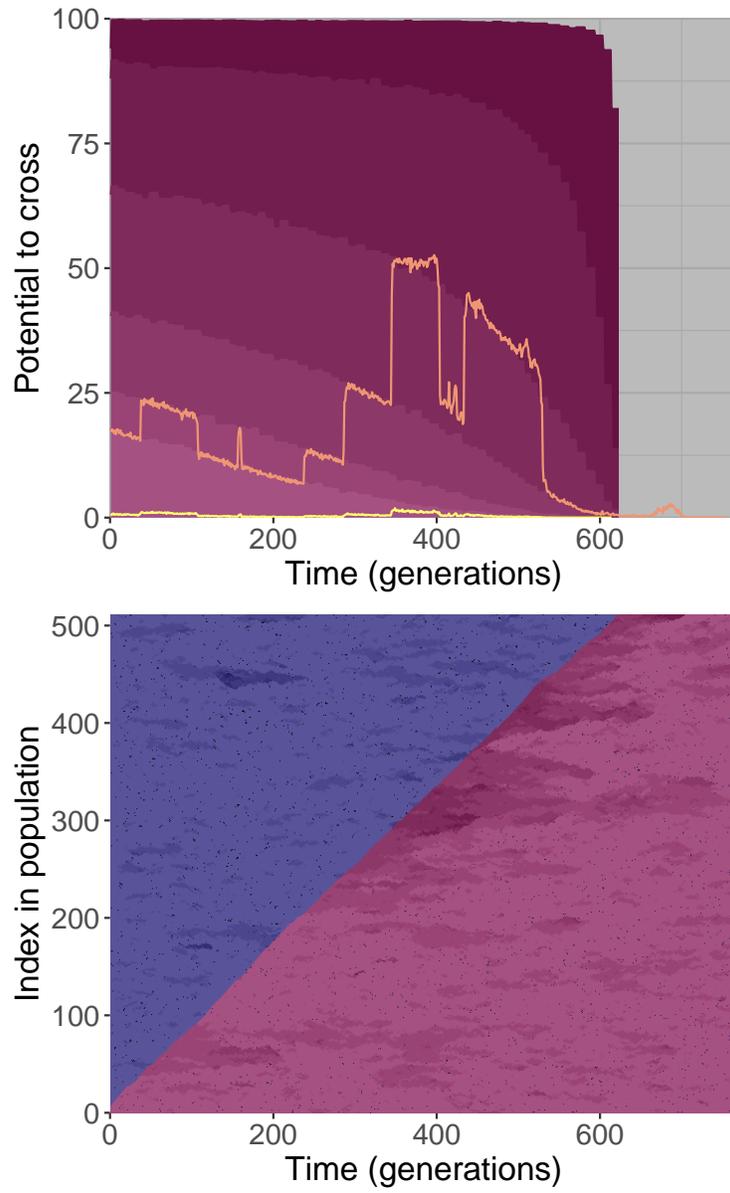


Figure 5.6: The analytic replay data for a representative replicate that failed to cross a valley during a momentum window. See Figure 5.4 for details and legend.

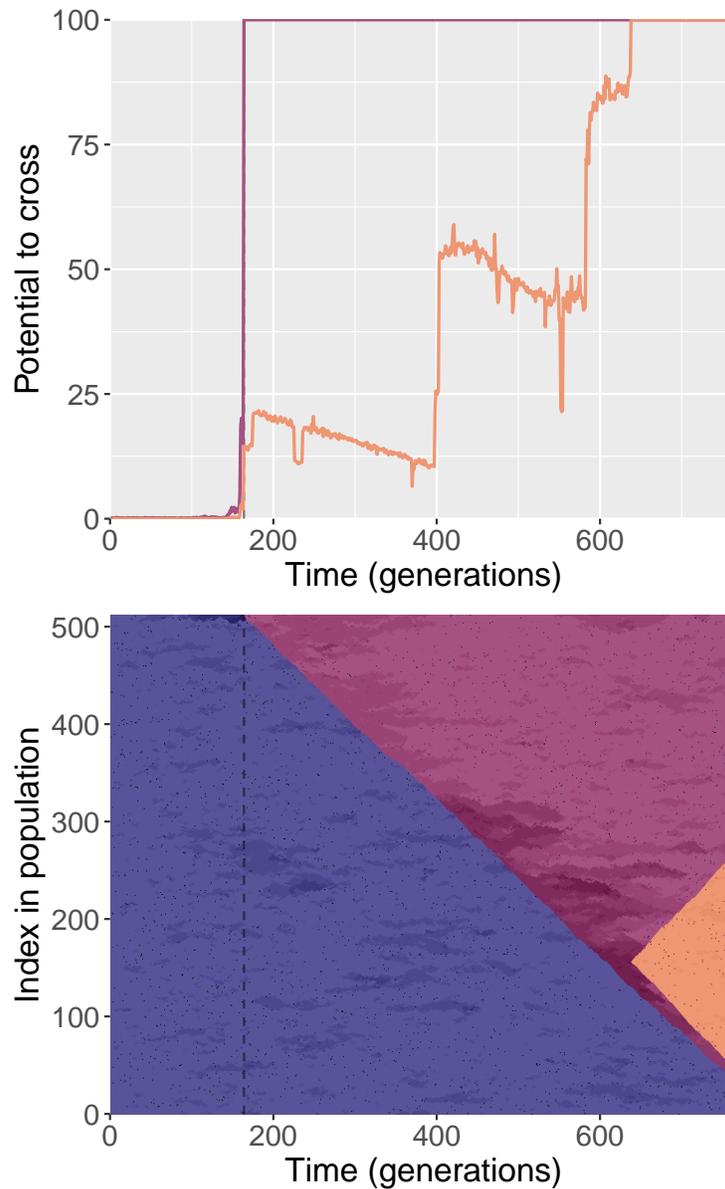


Figure 5.7: **Top plot:** The analytic replay data for the sole replicate that did *not* start in a momentum window, but still managed to cross a valley, and indeed crossed twice. The red line shows the potential to cross the first valley, while the orange line shows the potential to cross the second valley, exhibiting the hallmarks of adaptive momentum due to the first cross. **Bottom plot:** a Muller plot of the original population, like in Figure 5.4.

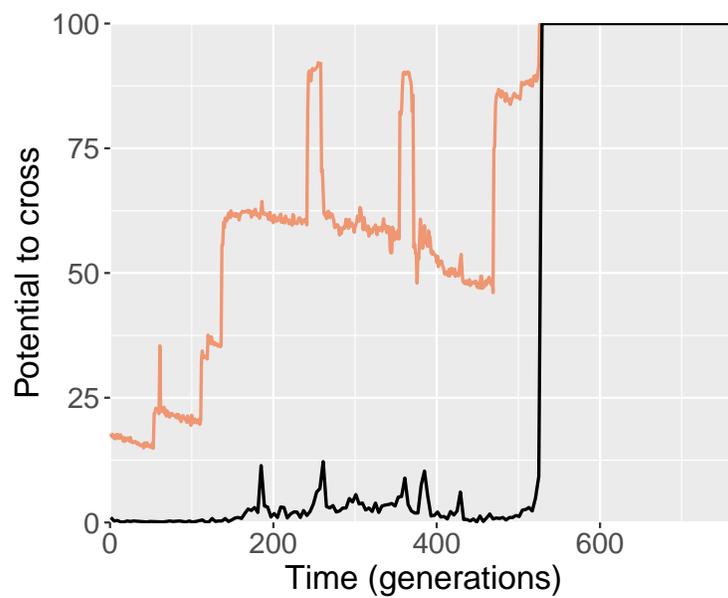


Figure 5.8: The black line (bottom) shows the potential for a shuffled population to cross a valley. The orange line (top) shows the standard potential for the population to cross (same data as Fig. 5.4). The shuffled line consists of 1,000 samples every 4 generations.

## 5.4 Discussion and Conclusion

### 5.4.1 Potentiation exhibits adaptive momentum

In this work, we have corroborated adaptive momentum’s benefit to valley crossing (as outlined in (Bohm et al., 2024)) and expanded our understanding of the dynamic. Our initial experiments demonstrated that our system can undergo adaptive momentum, and that disequilibrium is the key driver. However, our main goal in this paper is to provide an alternate vantage point from which to view the dynamics of adaptive momentum. While the original paper validated the effect via aggregated data, here we analyze the underlying dynamics in action on individual populations by quantifying potentiation via replay experiments.

We have shown that populations outside of momentum windows must rely on chance alone to cross a valley. During these “equilibrium” periods, early mutations into the valley, although required for a valley crossing, have no substantial impact on the population’s probability of crossing. Conversely, populations in momentum windows immediately see a drastically higher chance to cross, and every mutation in the leading edge is either potentiating or anti-potentiating, depending on the direction. This result is highlighted in Figure 5.7, where the first cross was pure chance while the second cross was driven by adaptive momentum. This work is only a beginning; future work should apply these techniques to examine the role of adaptive momentum in more complex and realistic fitness landscapes.

### 5.4.2 The leading edge in a spatial selective sweep determines potentiation

The adaptive momentum framework posits that disequilibrium in a population can reduce the selection pressure on the advantaged subpopulation (Bohm et al., 2024). In spatial populations, disequilibrium is focused at the leading edge of selective sweeps, where advantaged mutants encroach on the wild type. Indeed, we observe that the genotype of the organism at the leading edge of the sweep is a strong predictor of potentiation. Moreover, as we can see in the Muller plots, there are often mutated organisms lagging the leading edge. When the leading edge is further into the valley, those lagging organisms have a non-negligible chance to accumulate sufficient mutations to finish crossing the valley, thus increasing the potential.

We aimed to create the simplest system for studying valley crossing in spatial populations. In these one-dimensional populations, our artificially-started sweeps have exactly one leading edge. These edges become more complex in two-dimensional digital systems and only increase in com-

plexity moving toward more natural systems. While the identification and measurements of the leading edge may become more difficult, we expect similar dynamics to hold. It is critical that we continue to improve our understanding of adaptive momentum in these simple systems to build a solid theoretical foundation.

### 5.4.3 Population heterogeneity and structure affect potentiation

Previous experiments have conducted analytic replays starting from clonal populations (Blount et al., 2008; Ferguson and Ofria, 2023). Here, however, we used perfect population snapshots that record every organism in every generation. Our technique provides a more fine-grained look into how potentiation changes. Furthermore, previous work has often used the most abundant genotype at a time point to seed replay experiments. Across all of our experiments, the most abundant genotype was always on one of the peaks. Replays looking at the potential of the first cross would see effectively zero probability for  $p_1$  and  $p_2$  and 100% probability for  $p_3$  and  $p_4$ , which have already crossed, missing all nuance and change in this potential.

The nuance gained by replaying population snapshots provides important insight into using analytic replay experiments, and puts the large jumps in potentiation seen in previous work into question (Ferguson and Ofria, 2023). While those large jumps in potentiation are valid, they are likely missing important intermediate genotypes or population dynamics. In effect, they show genetic effects isolated from effects of population structure. Future work looking to leverage replay experiments should be aware that population structure and composition can affect evolutionary outcomes, and should thus carefully consider how replays are initialized.

Further, by shuffling the population snapshots we have shown that it is not just the portion of the population with each genotype that matters, but also the structural relationships and interactions among the organisms (Figure 5.8). Other computational studies will likely be needed to tease apart when and how this organization matters as perfectly preserving structure and composition of natural populations for so many replicates is currently impossible. As such, we should leverage computational models to develop techniques possible in both digital and natural populations, possibly finding a middle ground between a single clonal sample and full population snapshots.

### 5.4.4 Outlook

This work not only provides evidence to support our understanding of adaptive momentum, but further clarifies the underlying mechanisms. By quantifying the potentiation of valley crosses and

relating this measure to population composition and structure, we have provided insights into the historical contingencies and long-term trajectories of populations experiencing adaptive momentum. Further, we advance the methodology of analytic replay experiments by conducting them at a greater scale than previously seen, demonstrating a new use case, and leveraging perfect population snapshots to seed the replay experiments. These advances combine to show that replay experiments can extend genetic potentiation to include the effects of population dynamics. All together, this work constitutes both a step toward better understanding adaptive momentum and a methodological refinement of replay experiments for understanding historical contingency.

## Chapter 6

### Conclusions and outlook

This dissertation constitutes an important step in understanding *how* historical contingency can shape long-term evolutionary outcomes. Using digital evolution, I have shown that the evolution of phenotypic plasticity can act as an evolutionary stabilizer, negating the effects of environmental variation on future evolutionary dynamics. By leveraging analytic replay experiments, I have shown that a single mutation can shift the likelihood of an evolutionary outcome from incredibly unlikely to almost certain. Further, I have demonstrated that the same replay techniques can be used to clearly illustrate basic evolutionary dynamics via illuminating the long-term impacts of short-term changes.

There remains an enormous amount work to fully understand the role of history in evolution. Thankfully, however, this work has given me some insight into how we think about historical contingency, potentiation, and replay experiments, as well as the future work that should be done to further our understanding. This chapter is dedicated to recapitulating the lessons learned throughout this dissertation and discussing how we can build upon them to continue enhancing our understanding of historical contingency in evolution.

#### 6.1 Contributions

Here I summarize the contributions of each research chapter.

In **Chapter 2** I experimentally measured the impact of evolved phenotypic plasticity on further evolution. I used the digital evolution software *Avida* to evolve both plastic and non-plastic populations in a cyclically changing environment. Once evolved, I moved populations to three novel environments to observe the difference that plasticity made in each. First, I found that phenotypic plasticity slows the rate of evolutionary change compared to the non-plastic populations that must continuously adapt to environmental changes. Second, I showed that plastic populations accumulated novel beneficial traits at the same rate as non-plastic populations, but plastic populations were able to retain the traits while non-plastic populations continuously lost them. Finally, I examined how, in their continuous struggle to adapt to the environment, non-plastic populations accumulate deleterious traits significantly more often than plastic populations. Overall this chapter identified how the evolution of one trait, phenotypic plasticity, shifts the dynamics of further evolution by drastically reducing the impact of environmental changes.

Next, in **Chapter 3**, I quantified how the likelihood of evolving associative learning (i.e., the potentiation of associative learning) changed over the course of four case study lineages. To do this, I employed “analytical replay experiments”, restarting evolution from various points along each lineage. Across the four lineages, I found substantial single-step increases in potentiation, ranging from 36 to 64 percentage points. I then analyzed the individual mutations that increased potentiation, finding a variety of dynamics in these potentiating mutations. In this chapter, I demonstrated that digital evolution can be used to perform studies on potentiation dynamics that are intractable in living organisms. In this case I used replay experiments to track down individual potentiating mutations.

I then extended this work in **Chapter 4**, analyzing the potentiation dynamics of 50 new lineages that were able to evolve associative learning. I found that the large single-step potentiation increases in Chapter 3 were not a fluke. While the distribution of all observed single-step potentiation increases was centered around zero, the maximum per-replicate increase was significant for all 50 replicates. Indeed, looking at the distribution of maximum per-step potentiation increase across the 50 replicates, I found a range of [20, 88], a mean of approximately 42, and a median of 38 percentage points. I also replayed ten lineages that did not evolve learning, and in each I found significant single-step increases in the potentiation of the most likely behavior at the end of evolution. Finally, I found that in 44 of 50 replicates the largest potentiation gain was caused by a single mutation, and all but one of these mutations were either beneficial or neutral, with 15 mutations causing no change to the organism’s phenotype.

**Chapter 5** pivots to using replay experiments and potentiation to illustrate the adaptive momentum effect in a simple digital model. The adaptive momentum framework posits that disequilibrium, such as a selective sweep, can temporarily increase evolutionary exploration. By quantifying the potentiation of crossing a deleterious fitness valley, I show that early mutations into the valley drastically increase potentiation when the population is in disequilibrium, yet have no detectable effect when the population is in equilibrium. By performing replays with the organisms in the population shuffled at random before starting, I show that population structure plays a critical role in this scenario. This work also demonstrates how replaying population snapshots instead of clonal populations can improve the granularity of the potentiation dynamics, and in fact is especially relevant to the dynamics of adaptive momentum.

All together, these chapters 1) advance our understanding of the role that history plays in evolution 2) showcase the power of replay experiments in finding critical points in a population’s history and 3) demonstrate the advantage of using digital models to test and refine methodologies in evolutionary biology before applying them to natural organisms in the lab.

## 6.2 Insights for wet lab studies

While conducting the research outlined in this dissertation, I have had time to reflect on how the outcomes of these studies – all of which are digital – can be transferred to natural organisms and wet lab studies of their evolution. This question is, by far, the most common question I receive from evolutionary biologists who do not work with digital organisms themselves. As an initial summary, you can consider these digital studies as a rapid form of prototyping and benchmarking. These results are valid, and should be considered when trying to understand the role of history in evolution. However, my overall goal for my ongoing digital research is two-fold:

First, I wish to refine the methodologies behind techniques such as replay experiments. Studies leveraging these techniques are costly; evolving microbial population can take months or years *for a single experiment*. By trying these techniques at a larger scale, I hope to identify time-saving optimizations, identify possible pitfalls, and develop meaningful analyses. This way, future wet lab studies can conduct similar studies more efficiently and confidently. For example, Chapters 3 and 4 have demonstrated that a two-phase replay experiment can be used to find potentiating steps while saving substantial amounts of time.

Second, I hope to provide a baseline that future studies – be they *in vitro* or *in silico* – can compare against. The need for these benchmark data is evident in this dissertation. Upon completion of Chapter 3, while I was confident in the experimental setup and the evolution system, I was surprised and a mildly concerned by how large the potentiation increases were. This result was my original motivation behind conducting the scaled-up experiments for Chapter 4; to ensure that observing such changes was not a probabilistic anomaly. Unfortunately, conducting replay replicates at the single-mutation resolution for that many replicates, is currently impossible in wet lab systems. As such, this work can serve as the first large-scale benchmark, allowing for comparisons until such large studies are possible to conduct with natural organisms.

Over the course of my PhD, I have shifted from considering myself mainly a “computer scientist” to primarily a “computational evolutionary biologist”. With this mindset, the goal in my research

has become to facilitate a synergistic interplay between wet lab experimental evolution and digital evolution. Wet lab researchers are limited in what experiments they can conduct due to time, money, and experimental control. Digital evolution researchers are limited in what their studies can say about evolution in the natural world. By capitalizing on the strengths of each discipline, we can continue to cover the challenges of the other.

### 6.3 Future work

We will first consider the obvious extensions to this work. Foremost, while the chapters of this dissertation dive deep into the role that history plays in evolution, I have analyzed only three experimental setups (two related environments in Avida and a simple one-dimensional integral landscape). We clearly need to expand this work to other types of organisms and environments. My long-term hope is for adventurous and determined wet lab evolutionary biologists to conduct more studies in this vein. However, I have barely begun to scratch the surface of digital evolution models. As such, here I discuss two axes we should vary: the digital evolution model and the task or environment within it. We can also vary the parameters of the system (e.g., the mutation rate, population size, etc), which I discuss in detail below.

There exist dozens of mature digital evolution, evolutionary computation, and similar software models in use today. One particular avenue of exploration is to see how potentiation dynamics change between these different models. For example, Chapters 3 and 4 study the evolution of associative learning in Avida using a path following task. Would the potentiation of path-following associative learning in Markov Brains (Hintze et al., 2017) or PushGP (Spector and Robinson, 2002) follow similar trends to those observed in Avida? The solutions would certainly look quite different, but future experiments are needed to empirically compare the potentiation dynamics.

Even if we limit ourselves to studies within Avida, potentiation dynamics may drastically differ across environments. Would we see large increases in potentiation of optimal plasticity in the plasticity environment from Chapter 2 or in the potentiation of the logic trait EQU in the classic “logic 9” environment (Ofria and Wilke, 2004)? Comparison across these environments is favorable: we could use roughly the same instruction set (barring environment-specific instructions), population size, mutation operators, etc. I find it likely, however, that potentiation dynamics would still differ. As discussed in Chapter 4, I hypothesize that potentiation is influenced by the optimization needed *after* the first versions of a behavior appear, which is possible, for example, with suboptimal

plasticity. In the “logic 9” environment, some logic tasks are building blocks of others. It is possible their potentiation is thus highly correlated, but it is also possible that the potentiation for a simple task actually *decreases* when it appears, as it actualizes in a way that makes it likely to be lost as the more complex tasks evolve.

Below I detail more specific ideas that have arisen during the creation of this thesis.

### 6.3.1 On the multiplicative nature of potentiation

Chapters 3 and 4 identify “potentiating mutations” that caused the largest increases in potentiation for a given lineage. However, this is only one interpretation of the “largest increase” in potentiation, focusing on the absolute percentage point change. We could, however, shift to a *relative* view of potentiation. To illustrate the difference in these two values, imagine we have two mutations. The first mutation increases potentiation from 10% to 40%, while the second increases potentiation from 40% to 80%. Under our current framing, the second mutation is our largest potentiating mutation, as it increases potentiation by 40 percentage points (compared to 30 for the first mutation). However, an alternate framing is that the first mutation quadrupled the probability of evolving our focal trait, while the second mutation only doubled the probability.

To begin contemplating the ramifications of using absolute vs relative potentiation change, we can consider the simple fitness landscape in Figure 6.1. If we assume a scenario with a low mutation rate and strong selection pressure, since fitness increases from the root node to the leaves, evolutionary theory dictates that all populations should evolve to a leaf node. Depending on the starting node, the population can encounter branching nodes over the course of evolution. Since each branching node has two neighboring nodes with higher yet equal fitness, we expect the population to go down each branch with 50% probability. Here we consider the potentiation of the population evolving to reach the star node. In this example, potentiation for the star node from nodes between node B and the star are all approximately 100%. Potentiation at the node B is expected to be 50%, as the population will either evolve down the left or right branch. Similarly, potentiation is expected to be 25% at node A and 12.5% at the root node. Thus, in terms of absolute potentiation differences along a trajectory from the root to the target node, potentiation will increase most moving from node B to the left at a 50% increase. However, if we instead look at relative increases, moving off of the root, node A, and node B are all the same, doubling potentiation. This difference is demonstrated in Figure 6.2

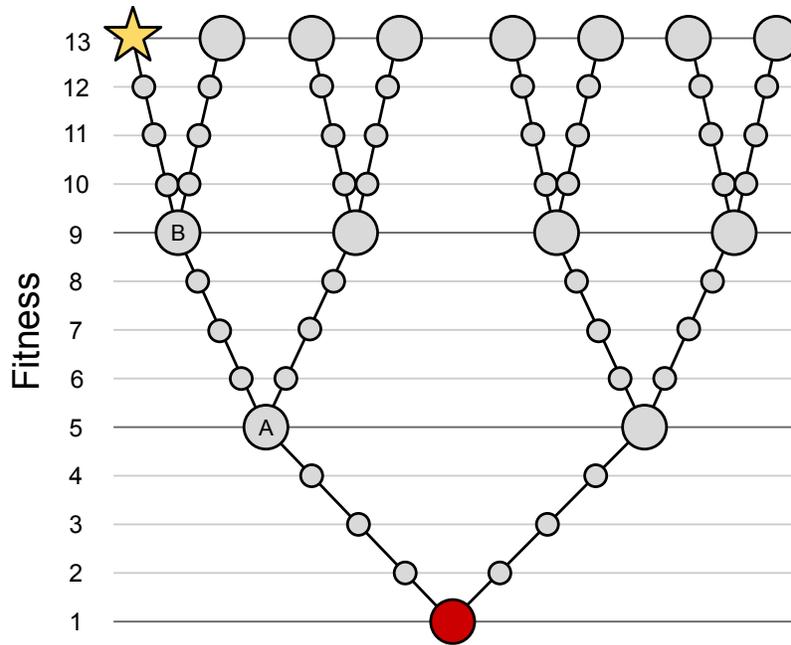


Figure 6.1: A hypothetical fitness landscape in the form of a binary tree. Each node represents a particular genotype, and edges represent available mutations. Branch points, which have two higher-fitness neighbors, are shown larger than other nodes. The root node is shown in red, and two nodes are marked (A and B) for the in-text example. We can consider potentiation relative to the star, which can be reached with a series of mutations that use the left option at each branch point.

Considering these differences as absolute or relative are both valid approaches, in the same way that other measures such as epistasis can be measured either additively or multiplicatively. In Chapters 3 and 4 I took an absolute approach as I wanted to analyze the overall potentiation of associative learning and how it changed. While the “decision points” encountered by evolution are interesting, they were not my main focus. What did prove relevant, however, is that because those experiments were conducted in Avida, I was limited to the resolution of my data. I strained the computation resources available in order to do 50 replays for each potentiation measurement, meaning that I had a base resolution of 2% intervals, before even considering noise. A 5% to 10% increase can be significant and impactful, but the number of replicates needed to determine that difference with a high degree of statistical confidence were computationally infeasible. Future work should consider both approaches however, as large relative increases in potentiation might happen much earlier in the population’s history and might provide additional insight.

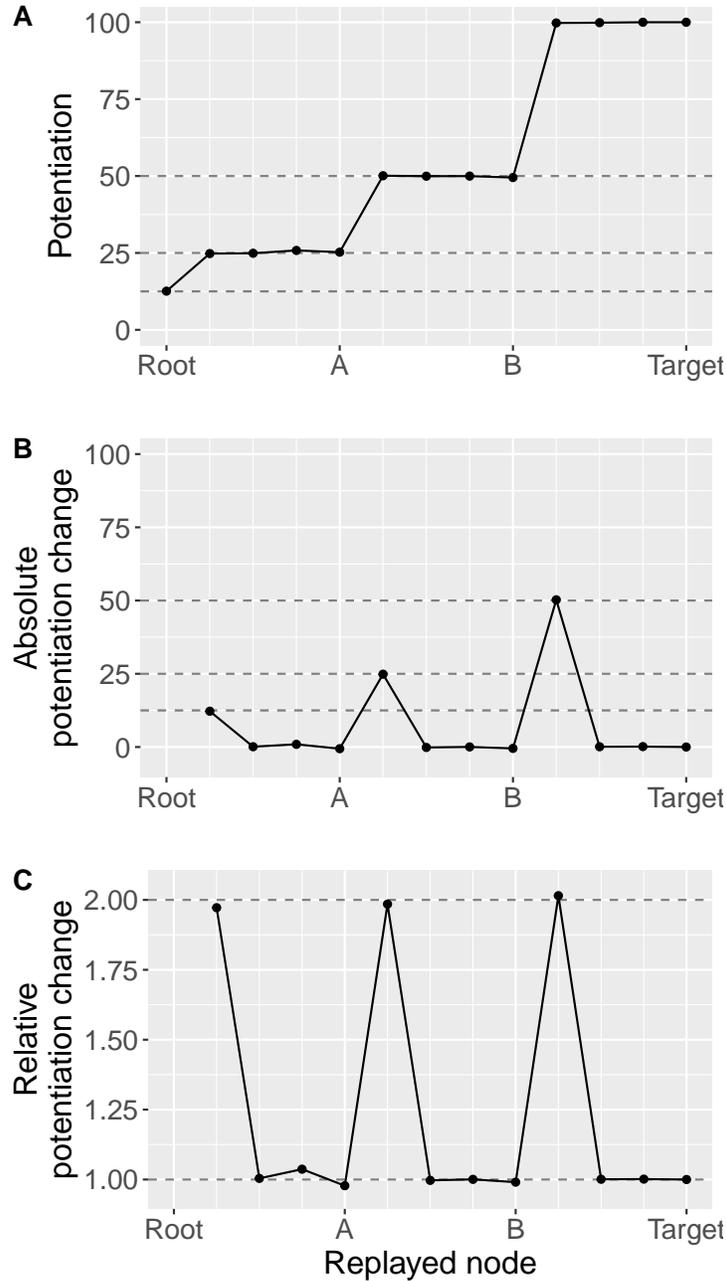


Figure 6.2: Empirically measured potentialization of nodes from the fitness landscape in Figure 6.1. Subplot A shows potentialization for each node on the left side of the tree. Potentialization for a node is calculated as the percentage of replicates (out of 10,000) started with a full population of 100 organisms on that node that evolve to the target node after 500 generations of roulette selection. Subplot B shows the absolute potentialization change when moving to each node, calculated as the difference between potentialization measured from the current position and the previous position. Subplot C shows the relative potentialization change, calculated as the potentialization measured at the current position divided by potentialization measured at the previous position.

### 6.3.2 Understanding factors that influence potentiation

In Chapter 3, I detailed the inner mechanics of the digital organisms and how the potentiating mutations paved the way for the evolution of learning. However, which dynamics factor into potentiation are still poorly understood. For example, my initial hypothesis was that most potentiating mutations would be deleterious or neutral, as beneficial mutations are likely to get selected and thus unlikely to influence the overall probability that the focal trait evolves. Chapter 4 provided strong counter evidence, with 49 of 50 potentiating mutations being either beneficial or having no significant fitness effect. This raises the question: what other factors are at play in potentiation?

Three factors I want to draw attention to are the size of the mutational neighborhood, the mutation rate, and the size of the population. Combined with the topography of the fitness landscape, these factors dictate the likelihood that a given mutation will appear. Specifically, as the number of neighbors increases, the mutation rate decreases, or the population size decreases, it becomes less likely that any given mutation will ever occur. As such, even the mutation with the highest possible fitness can be potentiating, as there is a chance *it never appears* and thus selection cannot act upon it. Additionally, even if a highly beneficial mutation appears, there is always the chance that it is lost to genetic drift. Future studies should examine how varying each of these parameters in isolation changes the potentiation dynamics. In designing the experiment in Chapter 3, I chose to use Avida partially because I was unsure how much complexity was needed to observe interesting potentiation changes. Chapter 5 has demonstrated that this complexity is not required. While future work should observe potentiation in other complex systems, potentiation should also be studied in the simplest possible systems to conduct controlled studies like the ones mentioned here.

### 6.3.3 Improving evolutionary computation

This dissertation is focused on improving our understanding of evolution as a process and the refinement of techniques to be applied back to wet lab studies. However, these techniques and ideas can also be applied to the problem-solving field of evolutionary computation (EC). In order to improve efficiency and problem-solving ability, EC practitioners are constantly refining their evolving systems. Even beyond the initial design phase, most systems have a collection of parameters and operators to fine-tune the system for the particular task at hand. Optimizing the

design decisions and parameter values of these systems is a continuous effort in EC research.

How might a better understanding of historical contingency and potentiation improve our understanding in these systems? Let us imagine that we want to test the effects of adding new mutational operators to our EC system. A proper first step would be to run a two treatment experiment: one batch of experiments that have access to these new operators and a control treatment that does not. With enough replicates, we should see if our changes improve the problem solving success of our system. Indeed, we could dive further and test these new operators in isolation or in combination, to disentangle which operators improve our success.

To consider the benefit of adopting a history-based mindset, we must switch from asking *if* these new mutational operators improve our problem solving success to asking *how* the new operators improve our evolutionary search. We can, and often do, make inferences about what events might have been critical to a replicate's success. For example, we might identify a gene duplication event occurring shortly before the focal trait evolves. However, only by experimentally testing the effect of accumulated history can we verify which specific events were pivotal in the population's ultimate success and which were mere coincidence. In our example, the gene duplication mentioned above might have helped, but it also may have followed another mutation whose importance is harder to detect. It is this interplay of dynamics that we aim to clarify. By leveraging these techniques, we may be able to nail down interactions between constituent parts, such as our mutation operators, that are beneficial and that we want to encourage via our design choices and parameters settings.

However, with this increased analytic power comes an increase in computational requirements. Techniques such as replay experiments have the potential to deliver critical insights into EC systems, but they are too computationally expensive to conduct regularly. As such, future work should aim to identify cases where these techniques are most useful, so practitioners can leverage them when they are the correct tool for the job.

#### **6.3.4 Quantifying evolvability**

Generally, analytic replay experiments have focused on the potentiation of beneficial traits. This leads to a simple observation: if the focal trait is beneficial, an increase in its potentiation is likely to correspond to an increase in the long-term fitness of the lineage, depending on the alternative trajectories that could have been taken. Though we lack an agreed upon definition of evolvability (Pigliucci, 2008), here I use the definition of Payne and Wagner (2019): “the ability of

a biological system to produce phenotypic variation that is both heritable and adaptive”. Thus, I argue that an increase in average long-term fitness constitutes an increase in evolvability.

I am not the first to combine the ideas of evolvability and historical contingency. Woods et al. (2011) were among the first to leverage analytic replay experiments, and they used these replays to test the long-term evolvability of two clades of *E. coli*. In the original replicate, the clade that eventually won had lower fitness at the time of sampling. Replay experiments started from that original point demonstrated that the “eventual winner” clade consistently evolved higher fitness despite its lower initial fitness – it was more *evolvable*.

In conducting multiple replay experiments, I have realized that these systems are ripe for further studies in evolvability. We must first consider the concept of “evolvability-enhancing mutations” as introduced by (Wagner, 2023). Evolvability-enhancing mutations are defined as mutations whose one-step mutational neighborhood has a higher average fitness than the one-step neighborhood of the wild type. As noted by the author, we could also define these evolvability-enhancing mutations using a two-step, three-step, or larger mutational neighborhood. Increasing the range of the mutational neighborhood would overcome issues such as evolvability-enhancing mutations that lead to local fitness peaks, limiting long-term fitness.

Unfortunately, performing a  $N$ -step mutant analysis becomes intractable on most fitness landscapes for, at most,  $N > 2$ . Thankfully, we can make a simplification: what if we only consider the mutational trajectories that are most likely? This is exactly what replay experiments do; each replay replicate samples the fitness landscape that is reachable from the starting condition in a given amount of time. While we are only sampling a fraction of the possible trajectories, the sample should be representative of what will evolve from the starting condition.

Finally, conducting these studies requires *little to no additional work*. If a replay experiment is conducted to observe the potentiation dynamics of a lineage for a particular trait, the only additional step is to also evaluate fitness at the end of each replay replicate. In Chapter 4, I had to quantify fitness in order to determine if the replay replicates evolved learning. As such, *these data already contain information on evolvability*. Unfortunately, a full analysis of these data are out of scope for this dissertation. A preliminary examination is shown in Figure 6.3, exhibiting the long-term fitness of the replays from two lineages. One lineage, which had only a 20 percentage point increase in potentiation, sees little change in the long-term fitness at the potentiating step. The second lineage,

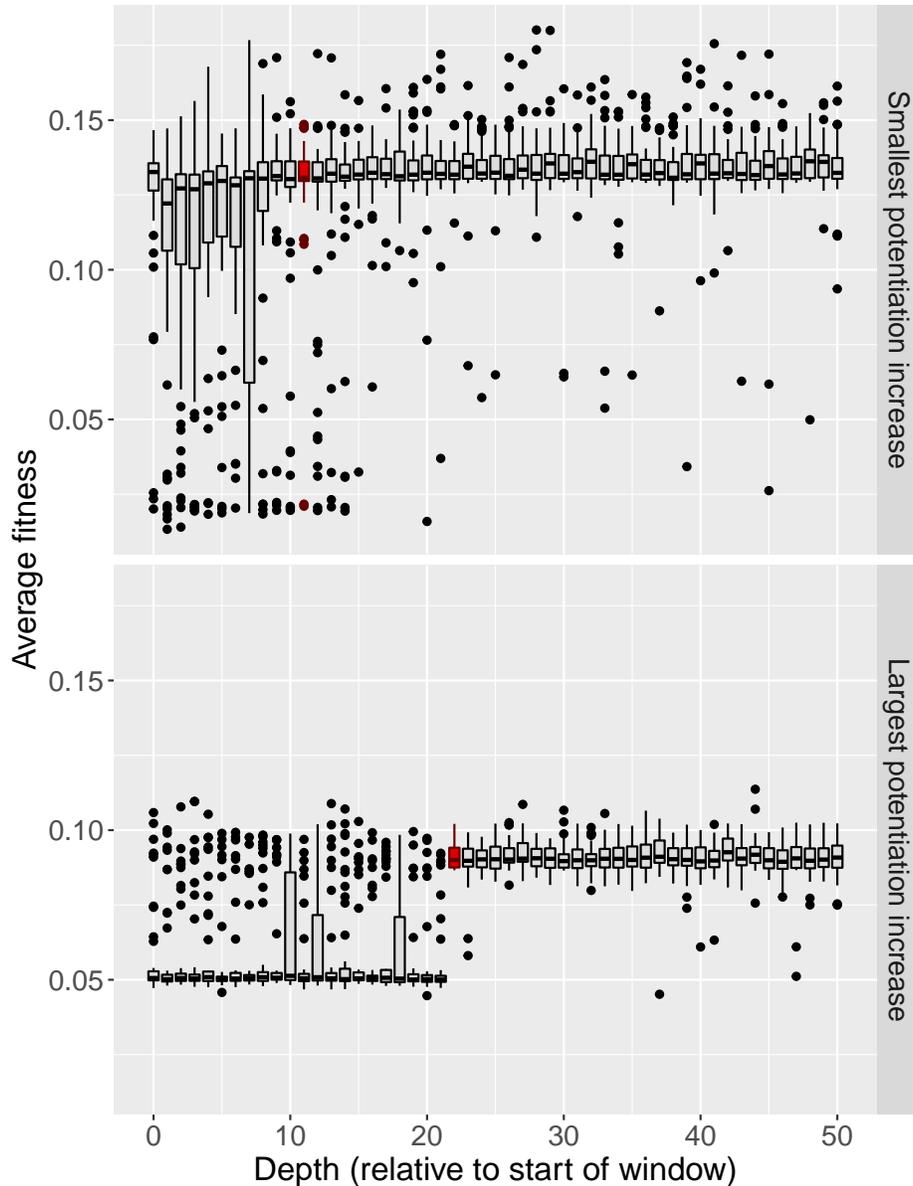


Figure 6.3: Two examples of evolvability data from the targeted replays of Chapter 4. A boxplot is shown for each depth replayed in the targeted replays, corresponding to the average fitness at the end of each of the 50 replicates replayed from that depth. The red boxplot shows the potentiating step. Two lineages are shown, one with a small potentiation increase (20 percentage points) and little change in long-term fitness, and another with a large potentiation increase (88 percentage points) and a stark increase in long-term fitness with the potentiating step.

which had the largest single-step potentiation increase (88 percentage points), sees a substantial increase in long-term fitness. When we consider these results through the lens of potentiation, we see that the higher-fitness trajectories were always present, potentiating mutations are simply

biasing evolution toward these high-fitness areas, and thus, increasing evolvability. Future work should further explore these dynamics; my hypothesis is that the increase in potentiation and the increase in long-term fitness should be correlated in systems where the potentiation trait being measured represents a highly beneficial adaptation.

## BIBLIOGRAPHY

- Ahlmann-Eltze, C. and Patil, I. (2021). *Ggsignif: Significance Brackets for 'Ggplot2'*.
- Allaire, J.J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., and Iannone, R. (2020). *Rmarkdown: Dynamic Documents for R*.
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., and Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, 4:63.
- Ancel, L. W. (2000). Undermining the Baldwin Expediting Effect: Does Phenotypic Plasticity Accelerate Evolution? *Theoretical Population Biology*, 58(4):307–319.
- Barrett, R. and Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1):38–44.
- Barton, N. H. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1403):1553–1562.
- Beaumont, H. J. E., Gallie, J., Kost, C., Ferguson, G. C., and Rainey, P. B. (2009). Experimental evolution of bet hedging. *Nature*, 462(7269):90–93.
- Bedhomme, S., Lafforgue, G., and Elena, S. F. (2013). Genotypic but not phenotypic historical contingency revealed by viral experimental evolution. *BMC Evolutionary Biology*, 13(1):46.
- Blattner, F. R., Plunkett III, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., et al. (1997). The complete genome sequence of *Escherichia coli* K-12. *science*, 277(5331):1453–1462.
- Blount, Z. D., Barrick, J. E., Davidson, C. J., and Lenski, R. E. (2012). Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature*, 489(7417):513–518.
- Blount, Z. D., Borland, C. Z., and Lenski, R. E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 105(23):7899–7906.
- Blount, Z. D., Lenski, R. E., and Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life’s tape. *Science*, 362(6415):eaam5979.
- Bohm, C., Ragusa, V. R., Ofria, C., Lenski, R. E., and Adami, C. (2024). Reduced selection during sweeps lead to adaptive momentum on rugged landscapes. *bioRxiv : the preprint server for biology*.
- Boyer, S., Hérisant, L., and Sherlock, G. (2021). Adaptation is influenced by the complexity of environmental change during evolution in a dynamic environment. *PLOS Genetics*, 17(1):e1009314.
- Braught, G. and Dean, A. (2007). The effects of learning on the roles of chance, history and adaptation in evolving neural networks. In *Proceedings of the 3rd Australian Conference on Progress in Artificial Life, ACAL’07*, pages 201–211, Berlin, Heidelberg. Springer-Verlag.
- Bryson, D. M. (2012). *The Evolutionary Potential of Populations on Complex Fitness Landscapes*. PhD thesis, Michigan State University.

- Bundy, J. N., Ofria, C., and Lenski, R. E. (2021). How the footprint of history shapes the evolution of digital organisms. Preprint, *Evolutionary Biology*.
- Burke, E. K., Newall, J. P., and Weare, R. F. (1998). Initialization Strategies and Diversity in Evolutionary Timetabling. *Evolutionary Computation*, 6(1):81–103.
- Buskirk, S. W., Peace, R. E., and Lang, G. I. (2017). Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. *Proceedings of the National Academy of Sciences*, 114(31):8330–8335.
- Canino-Koning, R., Wisner, M. J., and Ofria, C. (2016). The Evolution of Evolvability: Changing Environments Promote Rapid Adaptation in Digital Organisms. In *Proceedings of the Artificial Life Conference 2016*, pages 268–275, Cancun, Mexico. MIT Press.
- Canino-Koning, R., Wisner, M. J., and Ofria, C. (2019). Fluctuating environments select for short-term phenotypic variation leading to long-term exploration. *PLOS Computational Biology*, 15(4):e1006445.
- Canty, A. and Ripley, B. D. (2019). *Boot: Bootstrap R (S-plus) Functions*.
- Card, K. J., LaBar, T., Gomez, J. B., and Lenski, R. E. (2019). Historical contingency in the evolution of antibiotic resistance after decades of relaxed selection. *PLOS Biology*, 17(10):e3000397.
- Card, K. J., Thomas, M. D., Graves, J. L., Barrick, J. E., and Lenski, R. E. (2021). Genomic evolution of antibiotic resistance is contingent on genetic background following a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 118(5):e2016886118.
- Chadwick, W. and Little, T. J. (2005). A parasite-mediated life-history shift in *Daphnia magna*. *Proceedings of the Royal Society B: Biological Sciences*, 272(1562):505–509.
- Chan, B. K., Siström, M., Wertz, J. E., Kortright, K. E., Narayan, D., and Turner, P. E. (2016). Phage selection restores antibiotic sensitivity in MDR *Pseudomonas aeruginosa*. *Scientific Reports*, 6(1):26717.
- Chevin, L.-M. and Lande, R. (2010). When do adaptive plasticity and genetic evolution prevent extinction of a density-regulated population? *Evolution; international journal of organic evolution*, 64(4):1143–1150.
- Chevin, L.-M., Lande, R., and Mace, G. M. (2010). Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory. *PLoS Biology*, 8(4):e1000357.
- Childs, D. Z., Metcalf, C.J.E., and Rees, M. (2010). Evolutionary bet-hedging in the real world: Empirical evidence and challenges revealed by plants. *Proceedings of the Royal Society B: Biological Sciences*, 277(1697):3055–3064.
- Clune, J., Ofria, C., and Pennock, R. T. (2007). Investigating the Emergence of Phenotypic Plasticity in Evolving Digital Organisms. In Almeida e Costa, F., Rocha, L. M., Costa, E., Harvey, I., and Coutinho, A., editors, *Advances in Artificial Life*, volume 4648, pages 74–83. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Covert III, 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.

- Crispo, E. (2007). The Baldwin effect and genetic assimilation: Revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution; international journal of organic evolution*, 61(11):2469–2479.
- Cummins, E. A., Snaith, A. E., McNally, A., and Hall, R. J. (2021). The role of potentiating mutations in the evolution of pandemic *Escherichia coli* clones. *European Journal of Clinical Microbiology & Infectious Diseases*.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. Murray, London.
- Dolson, E., Lalejini, A., Jorgensen, S., and Ofria, C. (2020). Interpreting the Tape of Life: Ancestry-Based Analyses Provide Insights and Intuition about Evolutionary Dynamics. *Artificial Life*, 26(1):58–79.
- Douglas, S. M., Chubiz, L. M., Harcombe, W. R., and Marx, C. J. (2017). Identification of the potentiating mutations and synergistic epistasis that enabled the evolution of inter-species cooperation. *PLOS ONE*, 12(5):e0174345.
- Dunlap, A. S. and Stephens, D. W. (2014). Experimental evolution of prepared learning. *Proceedings of the National Academy of Sciences*, 111(32):11750–11755.
- Dunn, A. M., Hogg, J. C., Kelly, A., and Hatcher, M. J. (2005). Two cues for sex determination in *Gammarus duebeni* : Adaptive variation in environmental sex determination? *Limnology and Oceanography*, 50(1):346–353.
- Ferguson, A. J. (2023). Supplemental material for chapters 3 and 4. Zenodo and Github. <https://doi.org/10.5281/zenodo.7953692>.
- Ferguson, A. J. (2024). Supplemental material for chapter 5. Zenodo and Github. <https://doi.org/10.5281/zenodo.11507982>.
- Ferguson, A. J. and Ofria, C. (2023). Potentiating Mutations Facilitate the Evolution of Associative Learning in Digital Organisms. In *ALIFE 2023: Ghost in the Machine: Proceedings of the 2023 Artificial Life Conference*. MIT Press.
- Fisher, R. A. (1919). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, 52(2):399–433.
- Flores-Moya, A., Rouco, M., García-Sánchez, M. J., García-Balboa, C., González, R., Costas, E., and López-Rodas, V. (2012). Effects of adaptation, chance, and history on the evolution of the toxic dinoflagellate *Alexandrium minutum* under selection of increased temperature and acidification. *Ecology and Evolution*, 2(6):1251–1259.
- Forsman, A. (2015). Rethinking phenotypic plasticity and its consequences for individuals, populations and species. *Heredity*, 115(4):276–284.
- Fry, B. G., Vidal, N., Norman, J. A., Vonk, F. J., Scheib, H., Ramjan, S. F. R., Kuruppu, S., Fung, K., Blair Hedges, S., Richardson, M. K., Hodgson, W. C., Ignjatovic, V., Summerhayes, R., and Kochva, E. (2006). Early evolution of the venom system in lizards and snakes. *Nature*, 439(7076):584–588.

- Ghalambor, C. K., Angeloni, L. M., and Carroll, S. P. (2010). Behavior as phenotypic plasticity. In Westneat, D. and Fox, C. W., editors, *Evolutionary Behavioral Ecology*, pages 90–107. Oxford University Press, New York, NY.
- Ghalambor, C. K., Hoke, K. L., Ruell, E. W., Fischer, E. K., Reznick, D. N., and Hughes, K. A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, 525(7569):372–375.
- Ghalambor, C. K., McKay, J. K., Carroll, S. P., and Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, 21(3):394–407.
- Gibert, P., Debat, V., and Ghalambor, C. K. (2019). Phenotypic plasticity, global change, and the speed of adaptive evolution. *Current Opinion in Insect Science*, 35:34–40.
- Gibson, G. and Dworkin, I. (2004). Uncovering cryptic genetic variation. *Nature Reviews Genetics*, 5(9):681–690.
- Gifford, D. R., Furió, V., Papkou, A., Vogwill, T., Oliver, A., and MacLean, R. C. (2018). Identifying and exploiting genes that potentiate the evolution of antibiotic resistance. *Nature Ecology & Evolution*, 2(6):1033–1039.
- Goldsby, H. J., Knoester, D. B., Ofria, C., and Kerr, B. (2014). The Evolutionary Origin of Somatic Cells under the Dirty Work Hypothesis. *PLoS Biology*, 12(5):e1001858.
- Gould, S. J. (1990). *Wonderful Life: The Burgess Shale and the Nature of History*. WW Norton & Company.
- Gould, S. J. and Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the royal society of London. Series B. Biological Sciences*, 205(1161):581–598.
- Grabowski, L. M., Bryson, D. M., Dyer, F. C., Ofria, C., and Pennock, R. T. (2010). Early Evolution of Memory Usage in Digital Organisms. In *ALIFE*, pages 224–231.
- Gullberg, E., Cao, S., Berg, O. G., Ilbäck, C., Sandegren, L., Hughes, D., and Andersson, D. I. (2011). Selection of resistant bacteria at very low antibiotic concentrations. *PLoS pathogens*, 7(7):e1002158.
- Gunnell, G. F. and Simmons, N. B. (2005). Fossil Evidence and the Origin of Bats. *Journal of Mammalian Evolution*, 12(1):209–246.
- Gupta, A., Zaman, L., Strobel, H. M., Gallie, J., Burmeister, A. R., Kerr, B., Tamar, E. S., Kishony, R., and Meyer, J. R. (2022). Host-parasite coevolution promotes innovation through deformations in fitness landscapes. *ELife*, 11:e76162.
- Gupta, A. P. and Lewontin, R. C. (1982). A Study of Reaction Norms in Natural Populations of *Drosophila pseudoobscura*. *Evolution; international journal of organic evolution*, 36(5):934.
- Harrell Jr, F. E. (2020). *Hmisc: Harrell Miscellaneous*.
- Harrower, M. and Brewer, C. A. (2003). ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal*, 40(1):27–37.

- Hendry, A. P. (2016). Key Questions on the Role of Phenotypic Plasticity in Eco-Evolutionary Dynamics. *Journal of Heredity*, 107(1):25–41.
- Heyduk, K., Moreno-Villena, J. J., Gilman, I. S., Christin, P.-A., and Edwards, E. J. (2019). The genetics of convergent evolution: Insights from plant photosynthesis. *Nature Reviews Genetics*, 20(8):485–493.
- Hintze, A., Edlund, J. A., Olson, R. S., Knoester, D. B., Schossau, J., Albantakis, L., Tehrani-Saleh, A., Kvam, P., Sheneman, L., and Goldsby, H. (2017). Markov brains: A technical introduction. *arXiv preprint arXiv:1709.05601*.
- Huey, R. B., Hertz, P. E., and Sinervo, B. (2003). Behavioral Drive versus Behavioral Inertia in Evolution: A Null Model Approach. *The American Naturalist*, 161(3):357–366.
- Jochumsen, N., Marvig, R. L., Damkiær, S., Jensen, R. L., Paulander, W., Molin, S., Jelsbak, L., and Folkesson, A. (2016). The evolution of antimicrobial peptide resistance in *Pseudomonas aeruginosa* is shaped by strong epistatic interactions. *Nature Communications*, 7(1):13002.
- Kashtan, N., Noor, E., and Alon, U. (2007). Varying environments can speed up evolution. *Proceedings of the National Academy of Sciences*, 104(34):13711–13716.
- Kassambara, A. (2021). *Rstatix: Pipe-friendly Framework for Basic Statistical Tests*.
- Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M. C. (2012). Experimental evolution. *Trends in Ecology & Evolution*, 27(10):547–560.
- Keller, S. R. and Taylor, D. R. (2008). History, chance and adaptation during biological invasion: Separating stochastic phenotypic evolution from response to selection. *Ecology Letters*, 11(8):852–866.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217:624–626.
- King, J. L. and Jukes, T. H. (1969). Non-Darwinian Evolution: Most evolutionary change in proteins may be due to neutral mutations and genetic drift. *Science (New York, N. Y.)*, 164(3881):788–798.
- Kortright, K. E., Chan, B. K., Koff, J. L., and Turner, P. E. (2019). Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. *Cell Host & Microbe*, 25(2):219–232.
- Kruskal, W. H. and Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47(260):583–621.
- Lahti, D. C., Johnson, N. A., Ajie, B. C., Otto, S. P., Hendry, A. P., Blumstein, D. T., Coss, R. G., Donohue, K., and Foster, S. A. (2009). Relaxed selection in the wild. *Trends in Ecology & Evolution*, 24(9):487–496.
- Lalejini, A. and Ferguson, A. J. (2021a). Data for chapter 2. Available via the Open Science Foundation. <https://osf.io/sav2c/>.
- Lalejini, A. and Ferguson, A. J. (2021b). Supplemental material for chapter 2. Zenodo and Github. <https://doi.org/10.5281/zenodo.4642704>.
- Lalejini, A., Ferguson, A. J., Grant, N. A., and Ofria, C. (2021). Adaptive phenotypic plasticity stabilizes evolution in fluctuating environments. *Frontiers in Ecology and Evolution*, page 550.

- Lalejini, A. and Ofria, C. (2016). The Evolutionary Origins of Phenotypic Plasticity. In *Proceedings of the Artificial Life Conference 2016*, pages 372–379, Cancun, Mexico. MIT Press.
- Lalejini, A. and Ofria, C. (2018). Evolving Reactive Agents with SignalGP. In *The 2018 Conference on Artificial Life*, pages 368–369, Tokyo, Japan. MIT Press.
- Lande, R. and Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. *Evolution; international journal of organic evolution*, 37(6):1210.
- Lenski, R. E., Ofria, C., Collier, T. C., and Adami, C. (1999). Genome complexity, robustness and genetic interactions in digital organisms. *Nature*, 400(6745):661–664.
- Lenski, R. E., Ofria, C., Pennock, R. T., and Adami, C. (2003). The evolutionary origin of complex features. *Nature*, 423(6936):139–144.
- Lenski, R. E., Rose, M. R., Simpson, S. C., and Tadler, S. C. (1991). Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American Naturalist*, 138(6):1315–1341.
- Leroi, A. M., Bennett, A. F., and Lenski, R. E. (1994). Temperature acclimation and competitive fitness: An experimental test of the beneficial acclimation assumption. *Proceedings of the National Academy of Sciences*, 91(5):1917–1921.
- Levis, N. A. and Pfennig, D. W. (2016). Evaluating ‘Plasticity-First’ Evolution in Nature: Key Criteria and Empirical Approaches. *Trends in Ecology & Evolution*, 31(7):563–574.
- Li, Y. and Wilke, C. O. (2004). Digital Evolution in Time-Dependent Fitness Landscapes. *Artificial Life*, 10(2):123–134.
- Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K., and Rodríguez-Schettino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, 279(5359):2115–2118.
- Mayr, E. (1983). How to carry out the adaptationist program? *The American Naturalist*, 121(3):324–334.
- McGrath, K. M., Russell, S. J., Fer, E., Garmendia, E., Hosgel, A., Baltrus, D. A., and Kaçar, B. (2024). Fitness benefits of a synonymous substitution in an ancient EF-Tu gene depend on the genetic background. *Journal of Bacteriology*, 206(2):e00329–23.
- McGregor, S., Vasas, V., Husbands, P., and Fernando, C. (2012). Evolution of Associative Learning in Chemical Networks. *PLOS Computational Biology*, 8(11):e1002739.
- Mery, F. and Kawecki, T. J. (2002). Experimental evolution of learning ability in fruit flies. *Proceedings of the National Academy of Sciences*, 99(22):14274–14279.
- Meyer, J. R., Dobias, D. T., Weitz, J. S., Barrick, J. E., Quick, R. T., and Lenski, R. E. (2012). Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*, 335(6067):428–432.
- Misevic, D., Ofria, C., and Lenski, R. E. (2010). Experiments with Digital Organisms on the Origin and Maintenance of Sex in Changing Environments. *Journal of Heredity*, 101(Supplement 1):S46–S54.

- Moreno, M. A., Yang, C., Dolson, E., and Zaman, L. (2024). Trackable Agent-based Evolution Models at Wafer Scale.
- Muller, H. J. (1932). Some Genetic Aspects of Sex. *The American Naturalist*, 66(703):118–138.
- Nahum, J. R., West, J., Althouse, B. M., Zaman, L., Ofria, C., and Kerr, B. (2017). Improved adaptation in exogenously and endogenously changing environments. In *Proceedings of the 14th European Conference on Artificial Life ECAL 2017*, pages 306–313, Lyon, France. MIT Press.
- Nakazawa, M. (2019). *Fmsb: Functions for Medical Statistics Book with Some Demographic Data*.
- Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*.
- Ofria, C., Bryson, D. M., and Wilke, C. O. (2009). Avida: A Software Platform for Research in Computational Evolutionary Biology. In Komosinski, M. and Adamatzky, A., editors, *Artificial Life Models in Software*, pages 3–35. Springer London, London.
- Ofria, C. and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial life*, 10(2):191–229.
- Oke, K. B., Cunningham, C. J., Quinn, T. P., and Hendry, A. P. (2019). Independent lineages in a common environment: The roles of determinism and contingency in shaping the migration timing of even- versus odd-year pink salmon over broad spatial and temporal scales. *Ecology Letters*, 22(10):1547–1556.
- Paaby, A. B. and Rockman, M. V. (2014). Cryptic genetic variation: Evolution’s hidden substrate. *Nature Reviews Genetics*, 15(4):247–258.
- Paenke, I., Sendhoff, B., and Kawecki, T. J. (2007). Influence of Plasticity and Learning on Evolution under Directional Selection. *The American Naturalist*, 170(2):E47–E58.
- Payne, J. L. and Wagner, A. (2019). The causes of evolvability and their evolution. *Nature Reviews Genetics*, 20(1):24–38.
- Pennock, R. T. (2007). Models, simulations, instantiations, and evidence: The case of digital evolution. *Journal of Experimental & Theoretical Artificial Intelligence*, 19(1):29–42.
- Pigliucci, M. (2006). Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*, 209(12):2362–2367.
- Pigliucci, M. (2008). Is evolvability evolvable? *Nature Reviews Genetics*, 9(1):75–82.
- Pontes, A. C., Mobley, R. B., Ofria, C., Adami, C., and Dyer, F. C. (2020). The evolutionary origin of associative learning. *The American Naturalist*, 195(1):E1–E19.
- Price, T. D., Qvarnström, A., and Irwin, D. E. (2003). The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1523):1433–1440.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ratcliff, W. C., Denison, R. F., Borrello, M., and Travisano, M. (2012). Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences*, 109(5):1595–1600.

- Rice, W. R. (1989). Analyzing Tables of Statistical Tests. *Evolution; international journal of organic evolution*, 43(1):223.
- Santos-Lopez, A., Marshall, C. W., Haas, A. L., Turner, C., Rasero, J., and Cooper, V. S. (2021). The roles of history, chance, and natural selection in the evolution of antibiotic resistance. *eLife*, 10:e70676.
- Schaum, C. E. and Collins, S. (2014). Plasticity predicts evolution in a marine alga. *Proceedings of the Royal Society B: Biological Sciences*, 281(1793):20141486.
- Schlichting, C. D. (2008). Hidden Reaction Norms, Cryptic Genetic Variation, and Evolvability. *Annals of the New York Academy of Sciences*, 1133(1):187–203.
- Schlichting, C. D. and Wund, M. A. (2014). Phenotypic Plasticity and Epigenetic Marking: An Assessment of Evidence for Genetic Accommodation. *Evolution; international journal of organic evolution*, 68(3):656–672.
- Schmidt, M. D. and Lipson, H. (2009). Incorporating expert knowledge in evolutionary search: A study of seeding methods. In *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, GECCO '09, pages 1091–1098, New York, NY, USA. Association for Computing Machinery.
- Schwander, T. and Leimar, O. (2011). Genes as leaders and followers in evolution. *Trends in Ecology & Evolution*, 26(3):143–151.
- Simões, P., Fragata, I., Seabra, S. G., Faria, G. S., Santos, M. A., Rose, M. R., Santos, M., and Matos, M. (2017). Predictable phenotypic, but not karyotypic, evolution of populations with contrasting initial history. *Scientific Reports*, 7(1):913.
- Smith, C. E., Smith, A. N. H., Cooper, T. F., and Moore, F. B.-G. (2022). Fitness of evolving bacterial populations is contingent on deep and shallow history but only shallow history creates predictable patterns. *Proceedings of the Royal Society B: Biological Sciences*, 289(1982):20221292.
- Smith, J. M. and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, 23(1):23–35.
- Spector, L. and Robinson, A. (2002). Genetic Programming and Autoconstructive Evolution with the Push Programming Language. *Genetic Programming and Evolvable Machines*, 3(1):7–40.
- Taylor, T., Bedau, M., Channon, A., Ackley, D., Banzhaf, W., Beslon, G., Dolson, E., Froese, T., Hickenbotham, S., Ikegami, T., McMullin, B., Packard, N., Rasmussen, S., Virgo, N., Agmon, E., Clark, E., McGregor, S., Ofria, C., Ropella, 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.
- Taylor, T. and Hallam, J. (1998). Replaying the tape: An investigation into the role of contingency in evolution. In *Proceedings of Artificial Life VI*, pages 256–265. MIT Press, Cambridge, MA.
- Travisano, M., Mongold, J. A., Bennett, A. F., and Lenski, R. E. (1995). Experimental tests of the roles of adaptation, chance, and history in evolution. *Science*, 267(5194):87–90.
- Turner, C. B., Blount, Z. D., and Lenski, R. E. (2015). Replaying evolution to test the cause of extinction of one ecotype in an experimentally evolved population. *PLoS One*, 10(11):e0142050.

- Van den Bergh, B., Swings, T., Fauvart, M., and Michiels, J. (2018). Experimental Design, Population Dynamics, and Diversity in Microbial Experimental Evolution. *Microbiology and Molecular Biology Reviews*, 82(3):e00008–18, /mibr/82/3/e00008–18.atom.
- Vignogna, R. C., Buskirk, S. W., and Lang, G. I. (2021). Exploring a local genetic interaction network using evolutionary replay experiments. *Molecular biology and evolution*, 38(8):3144–3152.
- Wagenaar, D. A. and Adami, C. (2004). Influence of Chance, History, and Adaptation on Digital Evolution. *Artificial Life*, 10(2):181–190.
- Wagner, A. (2023). Evolvability-enhancing mutations in the fitness landscapes of an RNA and a protein. *Nature Communications*, 14(1):3624.
- Wagner, R. A. and Fischer, M. J. (1974). The String-to-String Correction Problem. *Journal of the ACM*, 21(1):168–173.
- Wennersten, L. and Forsman, A. (2012). Population-level consequences of polymorphism, plasticity and randomized phenotype switching: A review of predictions. *Biological Reviews*, 87(3):756–767.
- West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press.
- West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences*, 102(Supplement 1):6543–6549.
- West-Eberhard, M. J. (2008). Phenotypic Plasticity. In Jørgensen, S. E. and Fath, B. D., editors, *Encyclopedia of Ecology*. Elsevier.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., and Dunnington, D. (2020). *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*.
- Wickham, H., François, R., Henry, L., and Müller, K. (2022). *Dplyr: A Grammar of Data Manipulation*.
- Wickham, H. and Girlich, M. (2022). *Tidyr: Tidy Messy Data*.
- Wickham, H. and Seidel, D. (2020). *Scales: Scale Functions for Visualization*.
- Wilcoxon, F. (1992). Individual Comparisons by Ranking Methods. In *Breakthroughs in Statistics*, pages 196–202. Springer New York, New York, NY.
- Wilke, C. O. (2020). *Cowplot: Streamlined Plot Theme and Plot Annotations for Ggplot2*.
- Wilke, C. O. and Adami, C. (2002). The biology of digital organisms. *Trends in Ecology & Evolution*, 17(11):528–532.
- 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.

- Winger, B. M., Auteri, G. G., Pegan, T. M., and Weeks, B. C. (2019). A long winter for the Red Queen: Rethinking the evolution of seasonal migration. *Biological Reviews*, 94(3):737–752.
- Woods, R. J., Barrick, J. E., Cooper, T. F., Shrestha, U., Kauth, M. R., and Lenski, R. E. (2011). Second-order selection for evolvability in a large *Escherichia coli* population. *Science*, 331(6023):1433–1436.
- Wund, M. A. (2012). Assessing the Impacts of Phenotypic Plasticity on Evolution. *Integrative and Comparative Biology*, 52(1):5–15.
- Xie, Y. (2020). *Bookdown: Authoring Books and Technical Documents with R Markdown*.
- Yedid, G., Ofria, C. A., and Lenski, R. E. (2008). Historical and contingent factors affect re-evolution of a complex feature lost during mass extinction in communities of digital organisms. *Journal of evolutionary biology*, 21(5):1335–1357.
- Zheng, J., Payne, J. L., and Wagner, A. (2019). Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science (New York, N.Y.)*, 365(6451):347–353.
- Zimmer, C. and Emlen, D. J. (2013). *Evolution: Making Sense of Life*. Roberts and Company Publishers, Greenwood Village, CO.

## APPENDIX

### Plots of potentiation over time

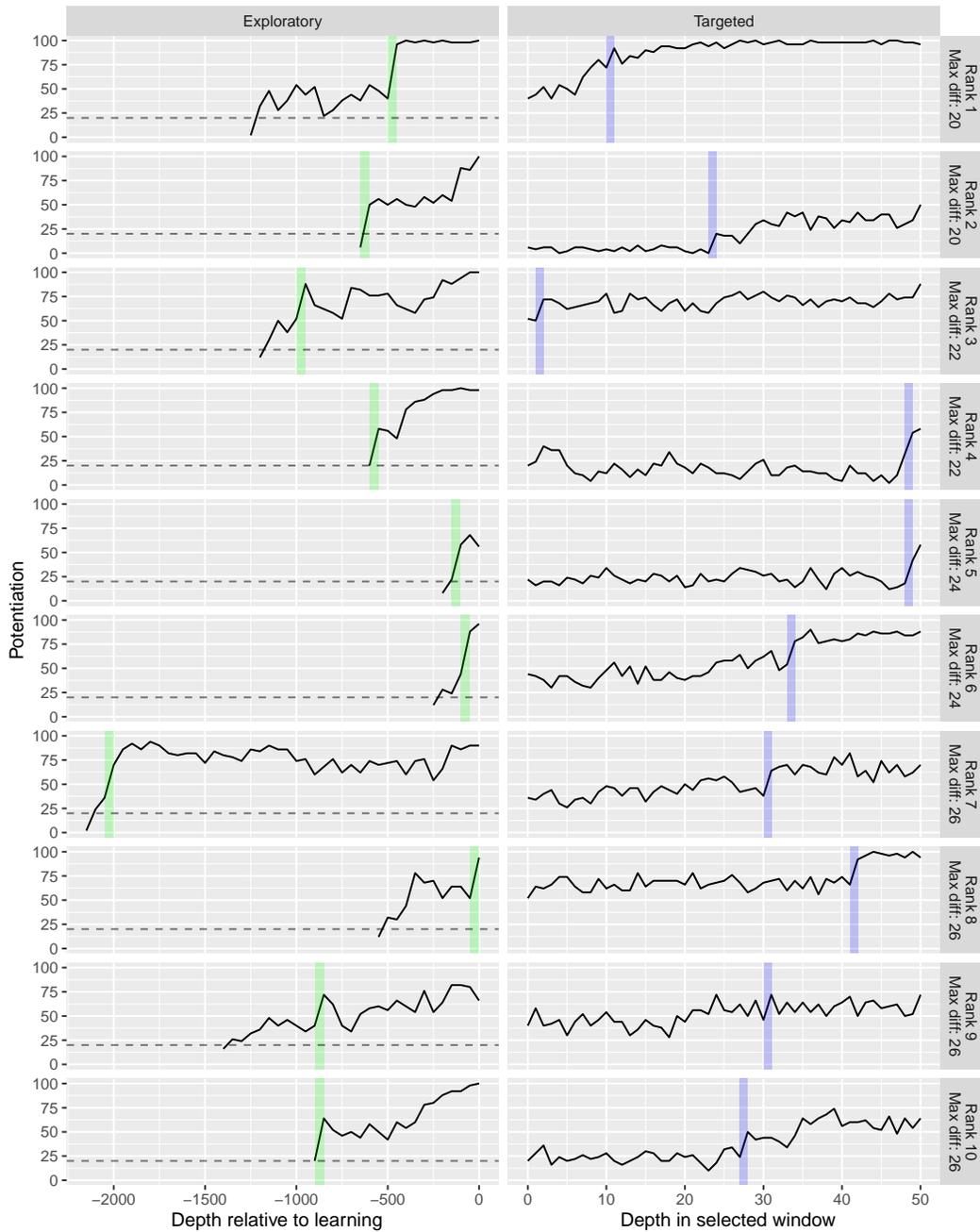


Figure 4: Potentiation over time plots from Chapter 4. Left panes show exploratory replays, with a green rectangle indicating the 50-step window with the largest increase in potentiation. Right panes show targeted replays, with the blue rectangle highlighting the largest single-step potentiation increase. Plots are ordered by the maximum single-step potentiation difference.

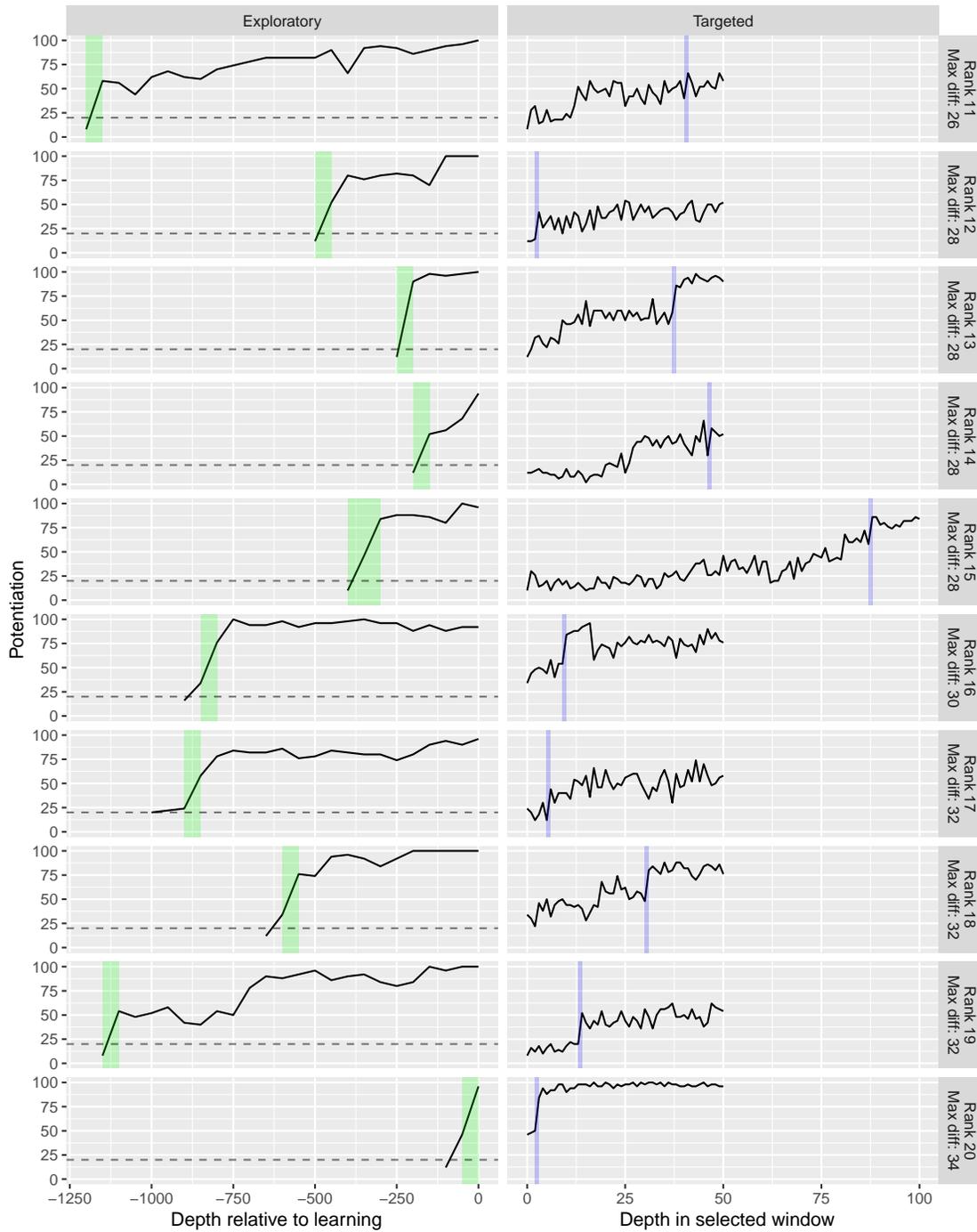


Figure 5: Potentiation over time plots from Chapter 4. Left panes show exploratory replays, with a green rectangle indicating the 50-step window with the largest increase in potentiation. Right panes show targeted replays, with the blue rectangle highlighting the largest single-step potentiation increase. Plots are ordered by the maximum single-step potentiation difference.

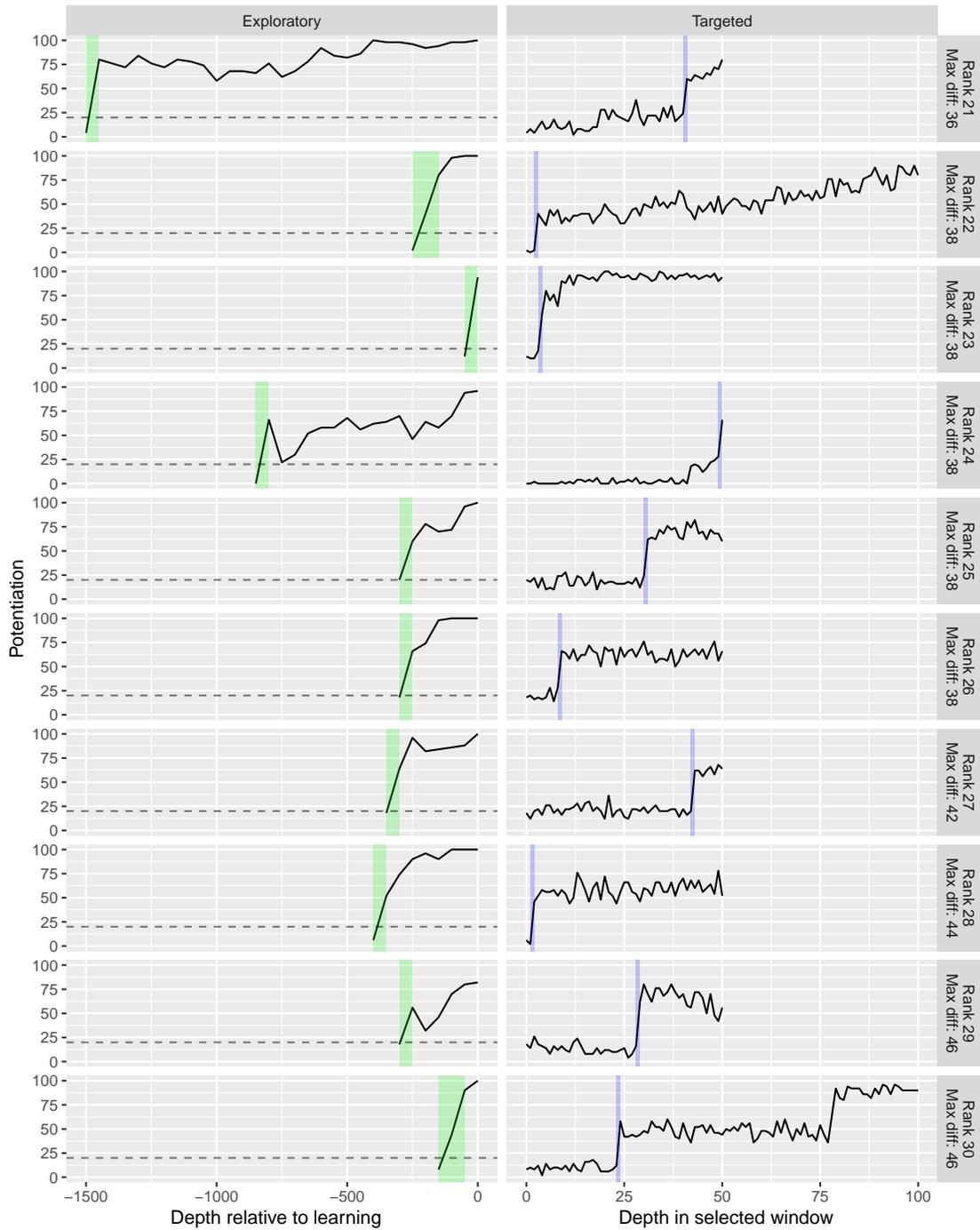


Figure 6: Potentiation over time plots from Chapter 4. Left panes show exploratory replays, with a green rectangle indicating the 50-step window with the largest increase in potentiation. Right panes show targeted replays, with the blue rectangle highlighting the largest single-step potentiation increase. Plots are ordered by the maximum single-step potentiation difference.

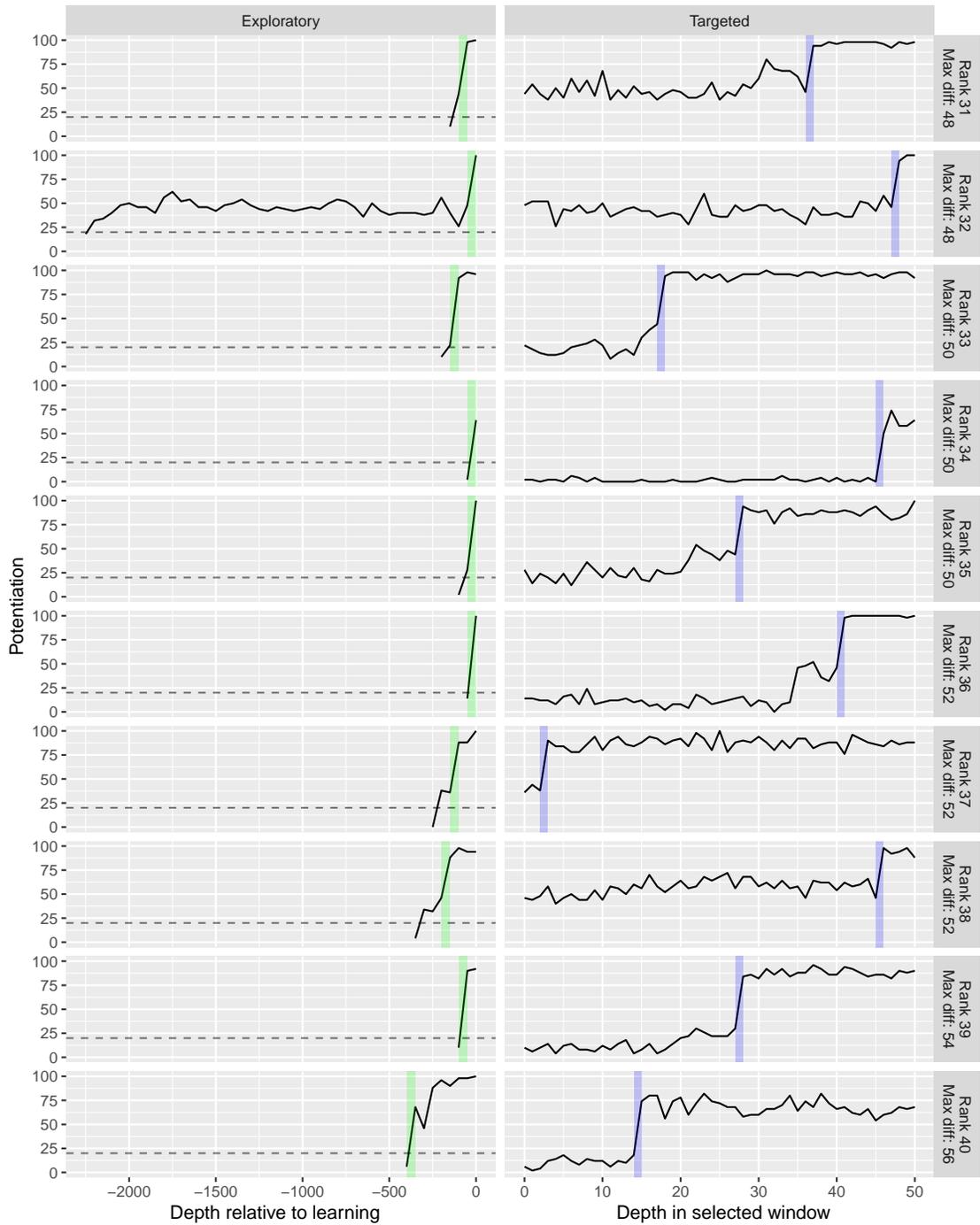


Figure 7: Potentiation over time plots from Chapter 4. Left panes show exploratory replays, with a green rectangle indicating the 50-step window with the largest increase in potentiation. Right panes show targeted replays, with the blue rectangle highlighting the largest single-step potentiation increase. Plots are ordered by the maximum single-step potentiation difference.

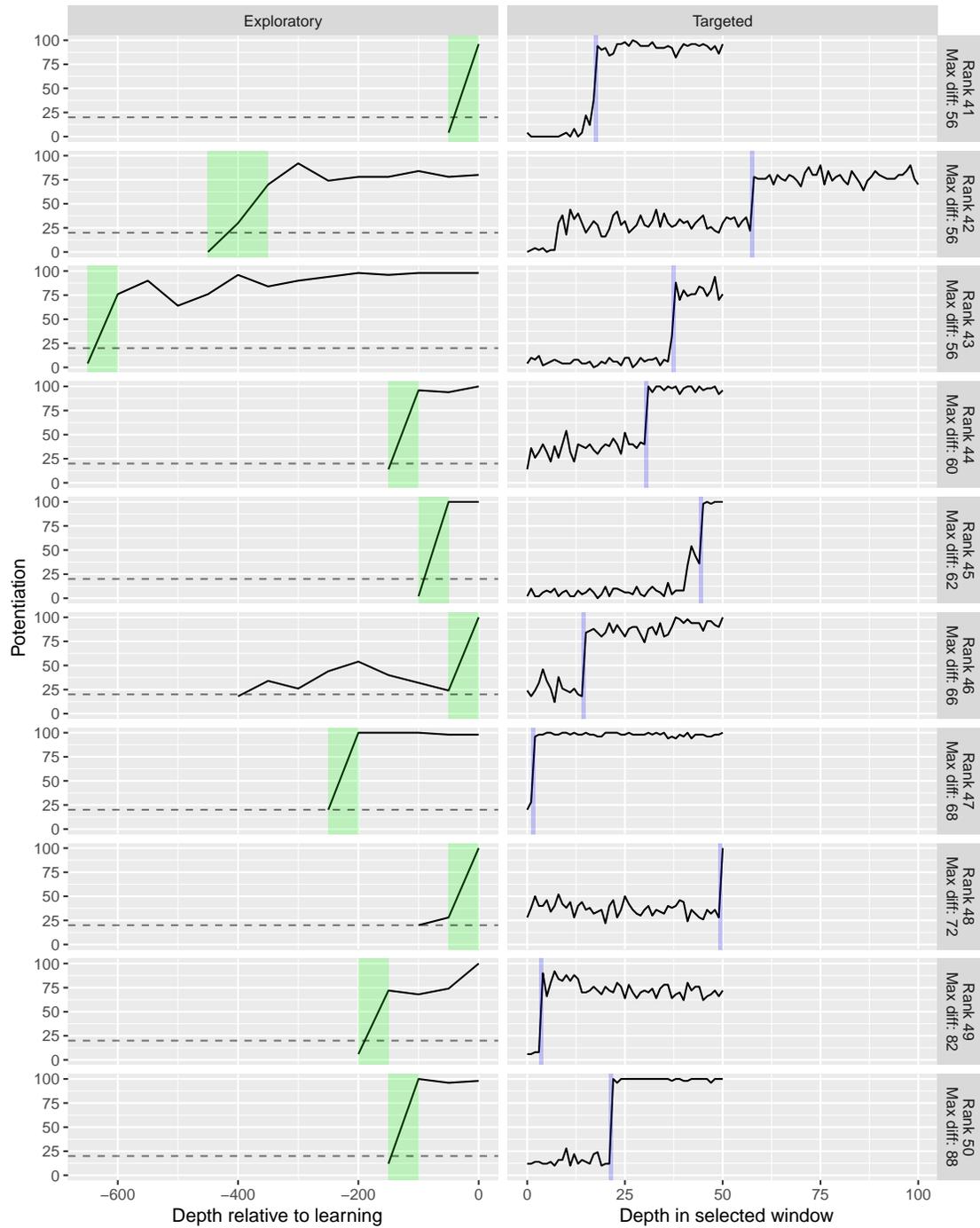


Figure 8: Potentiation over time plots from Chapter 4. Left panes show exploratory replays, with a green rectangle indicating the 50-step window with the largest increase in potentiation. Right panes show targeted replays, with the blue rectangle highlighting the largest single-step potentiation increase. Plots are ordered by the maximum single-step potentiation difference.

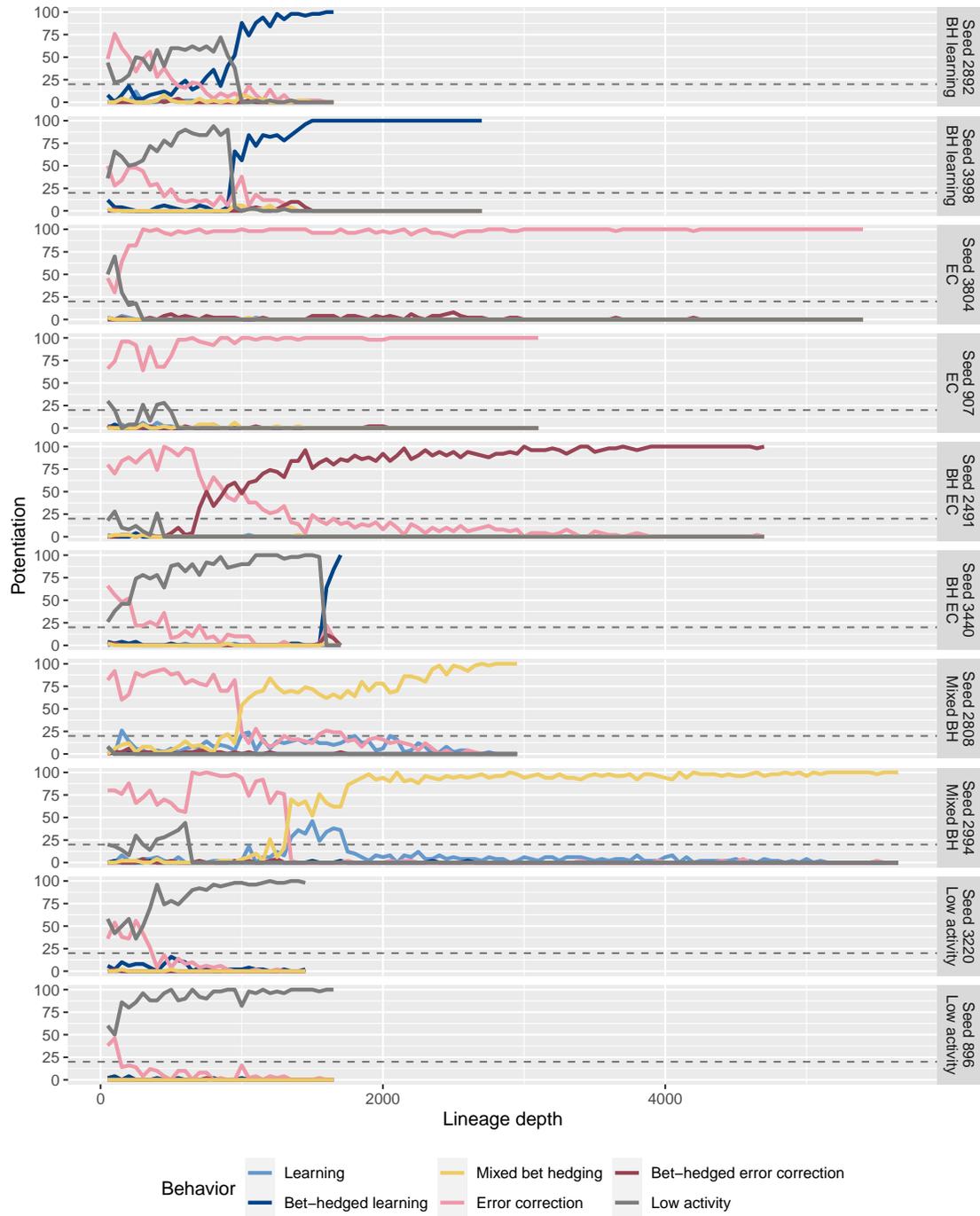


Figure 9: Potentiation over time plots from Chapter 4 for replayed replicates that did not originally evolve learning. Lines show the potentiation of each possible behavior. One replicate is shown per row, and rows are ordered by final evolved behavior. Acronyms: BH: bet-hedged / bet hedging; EC: error correction. Note that seed 3,440 evolved bet-hedged error correction, even though it was a very low probability at the latest replay point (0 of 50 replay replicates).