Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany

for the degree of
Doctor of Natural Sciences

Put forward by
M.Sc. Harald Mack
born in Schwäbisch Hall

Oral examination: October 18, 2023

Evolution from the ground up with Amee – From basic concepts to explorative modeling

Referees:

Prof. Dr. Kurt Roth
Prof. Dr. Michael Hausmann

Abstract

Evolutionary theory has been the foundation of biological research for about a century now, yet over the past few decades, new discoveries and theoretical advances have rapidly transformed our understanding of the evolutionary process. Foremost among them are evolutionary developmental biology, epigenetic inheritance, and various forms of evolutionarily relevant phenotypic plasticity, as well as cultural evolution, which ultimately led to the conceptualization of an extended evolutionary synthesis. Starting from abstract principles rooted in complexity theory, this thesis aims to provide a unified conceptual understanding of any kind of evolution, biological or otherwise. This is used in the second part to develop *Amee*, an agent-based model that unifies development, niche construction, and phenotypic plasticity with natural selection based on a simulated ecology. *Amee* is implemented in *Utopia*, which allows performant, integrated implementation and simulation of arbitrary agent-based models. A phenomenological overview over *Amee*'s capabilities is provided, ranging from the evolution of ecospecies down to the evolution of metabolic networks and up to beyond-species-level biological organization, all of which emerges autonomously from the basic dynamics. The interaction of development, plasticity, and niche construction has been investigated, and it has been shown that while expected natural phenomena can, in principle, arise, the accessible simulation time and system size are too small to produce natural evo-devo phenomena and –structures. *Amee* has thus can be used to simulate the evolution of a wide variety of processes.

Zusammenfassung

Die Evolutionstheorie bildet seit etwa einem Jahrhundert die Grundlage der biologischen Forschung, doch in den letzten Jahrzehnten haben neue Entdeckungen und theoretische Fortschritte unser Verständnis des evolutionären Prozesses rasch verändert. Dazu gehören vor allem die evolutionäre Entwicklungsbiologie, die epigenetische Vererbung und verschiedene Formen von phänotypischer Plastizität, was zur Konzeption einer erweiterten evolutionären Synthese geführt hat. Ausgehend von abstrakten Prinzipien, beginnend mit Konzepten aus der Komplexitätstheorie, versucht diese Arbeit ein einheitliches konzeptionelles Verständnis jeder Art von Evolution, ob biologisch oder anderweitig zu erlangen. Dieses wird im zweiten Teil genutzt, um *Amee* zu entwickeln, ein agentenbasiertes Modell, das entwicklungsbiologische Prinzipien, Nischenkonstruktion und phänotypische Plastizität mit natürlicher Selektion auf der Grundlage einer simulierten Ökologie vereint. *Amee* wurde dabei im *Utopia* framework implementiert, das eine performante, integrierte Simulation beliebiger agentenbasierter Modelle ermöglicht. Der hier präsentierte phänomenologische Überblick über die Fähigkeiten von *Amee* reicht von der Evolution von Ökospezies über die Evolution von Stoffwechselnetzwerken bis zu einer biologischen Organisation jenseits der Spezies-Ebene, wobei sich diese Phänomene autonom aus der grundlegenden Dynamik entwickelen. Das Zusammenspiel von Entwicklung, Plastizität und Nischenkonstruktion wurde untersucht, und es wurde gezeigt, dass zwar die erwarteten natürlichen Phänomene prinzipiell auftreten können, aber die zugängliche Simulationszeit und Systemgröße zu klein sind, um naturähnliche Evo-Devo-Phänomene und -Strukturen zu erzeugen. *Amee* stellt damit ein umfassendes Werkzeug zur Simulation von Evolutionsprozessen auf einer Vielzahl von Ebenen dar.

Contents

List of Figures

List of Tables

1. Motivation

The concept of biological evolution, the autonomous change of organism traits over time, has arguably been one of the most influential ideas in science [Mayr, 2000a], and it has inspired scientific fields ranging from cosmology [Smolin, 1992] to economy [Dosi and Nelson, 1994; Boulding, 1991; Ma and Nakamori, 2005]. Within biology, it has attained a field-defining place - as Dobzhansky famously expressed “Nothing in Biology makes sense except in the light of evolution” - and its progress has been remarkable: Important milestones include the unification of Darwin’s theory of natural selection with Mendel’s particular inheritance [Mendel, 1866] and statistical population genetics into the “modern synthesis” [J. Huxley, 1942], the discovery of the DNA molecule [J. D. Watson and F. H. C. Crick, 1953] and the discovery of the foundations of developmental biology [Goodman and Coughlin, 2000; Raff, 2000]. Over the same time, humanity has ushered in an unprecedented age of innovation and growth, such that we now live in a world of technological possibilities that were utterly unthinkable to our ancestors in Darwin’s time. With this came an unprecedented set of problems, largely arising from an ever tighter interconnection between artificial and natural systems: Climate change, biodiversity loss and ecosystem degradation are arguably consequences of the dynamics of globalizing markets in a growing global society. Thus, the natural ecological and evolutionary processes are increasingly exposed to anthropogenic forcing.

Recent studies show that human influence on natural processes often has reached the same or greater magnitude as natural drivers. This has reached an extent that warrants the definition of a new geological age, the Anthropocene [Steffen et al., 2007]. While the main argument for such a definition is that human forcing has at least reached parity with many natural subprocesses of system earth, the flip side is that we increasingly become an essential element of it, on which the rest of the natural world comes to depend. For viruses or any other parasite that infests primates or ungulates, humans and their livestock are probably the greatest natural resource the planet has to offer. Humans, therefore, have become a keystone species [Paine, 1969] in the global ecology. In other words, we create an entangled web of anthropic and natural processes in which interacting self-replicating entities of many kinds, subject to evolution, play a crucial part. This is obvious when discussing ecosystem degradation, invasive species or zoonoses like SARS-CoV-2 or Influenza, but it is also apparent in the technosphere, where the generation of innovations can be regarded as an evolutionary process [Ma and Nakamori, 2005]. Even the emergence of culture itself has been interpreted as following evolutionary principles [Mesoudi, 2011; Mesoudi, 2015; Whiten, 2017], unfolding in a continuous feedback loop with biological evolution. Thus, a fundamental, and mechanistic, understanding of evolutionary processes and their interplay with other parts of the earth system is of fundamental importance.

At the same time, evolutionary processes still present us with many challenges. On the fundamental side, the question of the origin of life [Damer and Deamer, 2020; Frenkel-

Pinter et al., 2020], the puzzle of the origin and maintenance of hierarchically integrated biological systems like multicellular organisms, eukaryotic cells or organism societies [Rainey, 2023; Rafiqi et al., 2020; Carmel, 2023; R. A. Watson, M. Levin, et al., 2022; Black et al., 2020], or the principles behind ontogenesis and development [Sultan, 2021; Gilbert, 2019] are just a few of the plethora of active research areas in the field. Indeed, many studies over the last decades call into question the classical "gene-eyed view" of evolution, which focuses on genetic mechanisms and understands evolution as a collection of processes that change gene frequencies in a population [K. Laland, Uller, et al., 2014], governed by population genetics. Foremost among them are results from evolutionary developmental biology [Müller, 2007] but other concepts like niche construction [Odlin-Smee et al., 2003] also contributed to this emerging "extended evolutionary synthesis" [Pigliucci, 2009; K. Laland, Matthews, et al., 2016; Müller, 2017a].

This extended synthesis represents not just an extension of the prevalent ideas in the field with new findings but changes the way we interpret self-reproducing entities. The two most important research directions are probably eco-evolutionary dynamics, or "eco-evo", which emphasizes the interplay between ecological and evolutionary processes on the same spatiotemporal scale [Pelletier et al., 2009; Bailey et al., 2009; Schoener, 2011], and evolutionary developmental biology, or "evo-devo", which investigates how phenotypic variation emerges from heritable information. This field has a strong connection to modern genetic research and elucidates mechanistic underpinnings of phenotypic plasticity and ontogenesis. In doing so, it has provided us with a novel understanding of the organization of the genome, the nature of evolutionary change, and the interplay between environmental signals and phenotypic variation. Therefore, it connects the ecological scale to the genetic and cellular interaction networks that make up individual phenotypes.

Niche construction is a related concept that has significant overlap with both. It views adaptive evolution as a two-way process in which organisms adapt to, but also change their surrounding ecosystem, and thus the ecological niches they find themselves in. It thus explicitly considers the interplay between ecological interactions and micro- and macroevolutionary processes that emerge from them. Niche construction theory stresses the fact that these biogenic influences persist across generations such that genetic inheritance is augmented by a general "ecological inheritance", of which the cultural inheritance mentioned before is a special case.

The idea of the extended evolutionary synthesis and Niche construction theory has sparked much debate [Scott-Phillips et al., 2014; K. Laland, Uller, et al., 2014; Baedke, 2020], with critiques pointing out that proponents assume an outdated view of evolutionary theory, or that the proposed extensions are overextending the concepts behind evolutionary thought [Dawkins, 2004]. Regardless of this back and forth, it is clear that evolutionary theory has entered a phase of reorganization, during which previously disparate fields of study are integrated into a larger whole to create a better understanding of natural evolutionary phenomena than ever before.

A very important tool in this regard is computer modeling, which allows the analysis of large quantities of data and the detailed simulation of individual biological processes. Today, it also allows for the large-scale simulation of integrated multiscale systems, starting from fundamental building blocks like individuals and genes and their abstracted interactions. Tools like individual-based modeling have seen strong development over the last two decades or so, and by now allow the simulation of many natural processes from the ground up. Since a concise analytical theory of evolution is out of reach, and probably not practical, outside of special cases largely based on population genetics, this approach makes progress for highly complex situations possible where experiments are too hard or not even feasible in principle.

Indeed, self-organization, emergence and the autonomous formation of structures that adapt and react to their environment are some of the hallmarks of life, and these are the classical properties that define complex (adaptive) systems. Complex systems are normally thought of as being made up of many interacting elements which show emergent behavior that cannot be predicted from the properties of their elements alone. Rather, these properties arise as part of the patterns and nature of interactions in the system –"More is different" [Anderson, 1972]. Complex adaptive systems comprise the subclass of complex systems in which their self-organization reacts to external signals, usually by adapting the interactions and properties of their constituents [Holland, 1992; S. A. Levin, 2002]. Complexity science has made vital contributions to the study of social [Eidelson, 1997], economic [Farmer et al., 2012; Anderson, 2018], and physical systems [Holovatch et al., 2017], and since complexity, adaptiveness and self-organization pervade the domain of life [Camazine et al., 2020; Isaeva, 2012], evolutionary theory can be thought of as being firmly embedded within complexity science - the elements of biological evolving systems are "just" endowed with a set of special properties that allow us to recognize their evolutionary nature. Again, technological advances in computer modeling have made remarkable progress possible. Much of this is owed to the recognition of complex networks [Strogatz, 2001; Boccaletti et al., 2006] as the fundamental architectural principle underpinning many complex phenomena: Most, if not all, complex adaptive systems can be imagined as self-modifying networks of interacting entities [Holland, 1992; S. A. Levin, 2002].

Given this perspective and the prevalence of complex adaptive systems and evolutionary systems in particular in the Anthropocene, the need for integrated modeling approaches arises. Although steps in this direction have been taken, most studies include only one or two of the many relevant processes on a single level of organization [Silver and Paolo, 2006; Lehmann, 2008; Staps et al., 2019], and although theoretical contributions strive for such an integrative approach [Laubichler and Renn, 2015; K. Laland, Matthews, et al., 2016; Müller, 2017a; R. A. Watson and Szathmáry, 2016; R. A. Watson, M. Levin, et al., 2022] an accompanying, similarly integrated modeling approach appears to be still missing. This thesis tries to contribute to this issue. In the first part, a view of evolutionary theory is presented that is embedded into the concepts of complexity science. In the second part, this view is translated into a simplified individual-based model for biological evolution, which is based on self-organizing and self-modifying networks. It integrates development, plasticity and niche construction within a resource-consumer network, and provides many possibilities for further extension. In this way, the present work strives to provide a starting point for an integrated modeling approach for general evolutionary systems.

Part I

Concepts and theory

2. From complexity to evolution

2.1 What is evolution?

2.1.1 Common definitions of evolution

The term “evolution”, in its general meaning, refers to “a process of change in a certain direction” [*Evolution, the Merriam-Webster dictionary* 2023], “a gradual process of change or development” [*Meaning of evolution in English, Cambridge Dictionary* 2023] or a similar variation of “change over time”. Applied to biology, this can be understood as “changes in the proportions of biological types in a population over time” [Millstein, 2022].

In biology, there are many different, more or less specific versions of the term, for example, the prevalent and highly specific “any change in the frequency of alleles within a population from one generation to the next” [Millstein, 2022] or Darwin’s “descent with modification” [Darwin, 1859] and many versions thereof. Regardless of definition, in a biological context, the term is intended to describe the process of how biological diversity arises and is maintained, with the different definitions referring to different mechanisms envisaged to be involved. Many definitions also explicitly refer to specific timescales. In a well-known textbook on evolution, it is defined as “inherited change in the properties of groups of organisms over the course of multiple generations.” (Futuyma and Kirkpatrick [2007], p. 7), i.e., changes in the properties of groups of organisms that happen during a single generation, e.g., via learning or horizontal gene transfer, are explicitly excluded. Indeed, if we understand evolutionary biology as the science that tries to understand the emergence of natural biological diversity, then “descent with modification” makes several a priori assumptions about the process governing this emergence, namely that change occurs gradually over (possibly many) parent-offspring links over timescales larger than individual lifetimes.

2.1.2 A general definition of evolution

However, other processes which are not bound to the concept of “descent with modification”, in its common meaning, arguably influence the dynamics of biological diversity over time, like the aforementioned phenotypic plasticity, ontogenesis [Gilbert et al., 2015; Pfennig, 2021], or social learning that influences behavior [Gariépy et al., 2014]. In that sense, an insistence on “descent with modification” leads to an artificial distinction between “evolutionary” and “non-evolutionary” modes of change in biology that are conceptually separated but mechanistically tightly interlinked. Indeed, evolutionary ideas have become prevalent in other fields as well, from economy [Witt, 2008; Hammerstein and E. Hagen, 2005] to cosmology [Smolin, 1992]. Therefore, in this chapter we employ a very general notion of biological evolution – “evolution is the change of physical or functional characteristics or -organization or the distribution thereof in a population of entities over time” – which is very close to the one given above from Millstein [2022]. Note that it does not refer to biology explicitly, although this will naturally be our main focus, given that it is the most well-studied such system. It covers the emergence of new kinds of entities in the population, or the loss of old ones, as well as purely a change in the distribution of their properties without changing anything else. Moreover, it does not only refer to changes in observable characteristics of some organisms with a fixed identity but allows also for the change of this identity, the emergence of new ones or loss thereof, i.e., a change in the physical or functional organization of the entities in the population. Therefore, it includes the “change in allele frequency” or “descent with modification” definitions but goes beyond them into speciation, innovation and the emergence of altogether new kinds of entities or conglomerates thereof.

Aside from its generality, another reason for adopting this definition is that it is purely descriptive, and does not allude to processes that underly the observed change. Also, taking such a more general approach allows us to place Darwinian evolution and other, similar concepts that arose over time into a larger context, thereby creating a deeper understanding of the requirements for and properties of biological evolution. The same holds for non-biological systems which show similar phenomena to biological evolution, but for which the entities in question are fundamentally different.

2.1.3 Characterization of biological evolution

Before going on, it makes sense to consider the phenomenology of biological evolution in some more depth. This serves to better understand the character of the phenomena it produces, before investigating in more depth. Three of its properties are especially worth noting. First, empirically we find that biological evolution tends to continuously create new adaptations which make organisms “better suited” to the ecosystem they find themselves in, which appears to be the case even in constant environments [Wiser et al., 2013; Lenski, Wiser, et al., 2015]. In contrast to physics, in which the concept of an equilibrium state towards which the system develops is often central, biological evolution behaves differently and it continues to generate new traits, lineages, species and functions – a process that, we can conjecture, is limited only by the fundamental physical interactions that the constituents of organisms are able to engage in.

The complex, ordered structures that evolve in this manner are not, however, indefinitely stable in themselves. This is in contrast to natural pattern formation, e.g., in sand dunes or -ripples, where the emerging patterns represent a natural attractor of the system that remains stable unless disturbed. The second important property of biological evolution is therefore that biological organisms are inherently open, non-equilibrium, dissipative systems [Kondepudi et al., 2020] that take up energy and matter from the environment and expel waste products into it. As a consequence, not only does the population of evolving organisms change over time, but the wider environment it is embedded into does as well. Biological evolution therefore invariably entails a “co-evolution” between environment and population. More than that, some form of environment into which an evolving population is embedded is always necessary - biological evolution without interaction with some environment is unthinkable.

Thirdly, biological evolution tends to create highly complex, ordered, functionally specialized structures, which came about by iterative modification of simpler precursors, some of which are compounds of other such structures [Goldenfeld and C. Woese, 2011]. Eyes, brains, livers, the wings of birds, bats or pterosaurs or termite mounts have no analogy in the non-biological world and structures of equivalent complexity and functionality are not formed by abiotic processes [Wolf et al., 2018]. Most importantly, employing the language of dynamical systems, the ability to create such complex structures allows for biological evolution to change the dimensionality and structure of the state space it unfolds in. This is invariably connected to a change in the dynamical laws that govern the system as a whole. For example, over the course of evolutionary history, the emergence of burrowing, itself brought about by novelties in body plans, quite radically changed the structure and function of the ecosystem [Bottjer, 2010]. Such changes are usually brought about by feedback processes that are affected by novelties in the system.

With these three properties – open-ended adaptivity and production of novelty, co-evolution between population and environment, iterative production of functionally complex, specialized structures and the ability to change its own dynamics – we can see biological evolution as a prototype for a class of systems that is phenomenologically distinct from everything we normally encounter in physics, and which I aim to characterize further throughout this chapter.

2.2 Elements of abstract evolutionary theory

2.2.1 About populations, individuals and traits

The entities making up the population we discuss here are most easily envisaged as animals, bacteria, or other life forms, although a priori, this does not need to be the case. For example, envisaging robots as individuals in the population would be consistent with the definition of evolution given above as well, and indeed, the field of Evolutionary Robotics employs such approaches [Alattas et al., 2018; Doncieux et al., 2015]. Elements of biological organisms, like genes or cells, would do as just as well as cultural artifacts like ideas or traditions, as would, taking the definition word by word, atoms or molecules. In turn, individual entities in the population also need not be individuals in the physical sense but could be aggregates of other entities, i.e., they could be populations themselves.

The collection of traits describing each individual can be of arbitrary composition in principle. In the simplest case, it can just be a simple scalar like body weight, fur color, number of flower petals, charge, or the concentration of metabolic enzymes in a cell, which can be continuous or discrete. Often, evolutionary biology considers indeed such simple cases. However, not all phenomena in such systems can be described by simple scalars. Other types of entities might have more complex traits, for instance, behavioral patterns, socially learned traditions, mate- or prey preference, and many other things.

We, therefore, describe an entity as a tuple of K elements g_k, where each describes an observable trait, i.e., an entity γ is described by $\gamma = (g_k), k = 0 \dots K$, without making further assumptions about the character of the traits in question. These variables can represent physiological characteristics, certain behavioral functions or other things, indeed, their choice is a practical one: If we want to describe a natural population, we always use a finite number of variables per entity based on the question we want to investigate. When such entities interact with their environment, it is by virtue of these traits, i.e., their overall state change over time will be a function of their traits and the traits of the environment they interact with. [1] The entire population Γ then is described as a set of such entities $\Gamma = \{\gamma_1, \dots, \gamma_N\}$.

2.2.2 General processes leading to changes in trait distribution

The Master Equation Asking how the distribution of properties in a population changes over time is the classical problem for stochastic dynamics, and represents the most simple class of changes inherent in the definition of evolution given above. Hence, we can ask what is the probability p_k to find the population in state $\Gamma_i(t) = \{\gamma_0, \gamma_1, \dots, \gamma_k, \dots \gamma_N\}$ at time t? Under the assumption that the system is governed by a Markov process and given a discrete and finite set of states Γ of the population, this time development is described by the Master equation [van Kampen, 2007]:

$$\frac{\mathrm{d}p_i}{\mathrm{dt}} = \sum_l A_{ji} p_j \Leftrightarrow \frac{\mathrm{d}\mathbf{p}}{\mathrm{dt}} = \mathsf{A}\mathbf{p} \tag{2.2.1}$$

The matrix A thereby describes the transition rates of the system from one state to another, and can in principle depend on the current state of the system i and time t. The Markov property thereby assures that no dependency on earlier points in time before t exists, which amounts to saying that all variables that influence p_k are aggregated into the state of the system Γ_i at the current time. The mechanics of the process is represented by the functional form of the transition rates A_{ij}. Trajectories of a system described by Equation 2.2.1 can be simulated using the Gillespie-algorithm [Gillespie, 1976]. Often, the mechanisms behind the transition rates in A are envisaged to be interactions between particles – we might think about collisions or, more general, reactions between them, mechanical, chemical, or otherwise.

Simple processes changing trait distributions Assuming the set of observable traits of a single organism to be $\gamma \subset \mathbb{R}^n$ such that the population can be described as a set of points in $\mathbb{R}^n$, a simple evolutionary process could be a random walk in trait space, well studied in physics, where each entity randomly changes its state by some amount l into an arbitrary direction. For biological organisms, this would amount to spontaneous transmutation, which, with a little imagination, we can envisage as arising by interaction with some environment. Such a process results in diffusion in trait space [van Kampen, 2007]. This process changes the distribution of traits in the system over time and thus constitutes indeed an evolutionary process. In fact, diffusion-like processes play a role in evolutionary theory in the form of neutral evolution (compare section 3.7). It's worth noting that this process allows for an, albeit slow, exploration of the trait space by the population, i.e., it increases the variance of the distribution of traits over time (see Figure 2.1, line **A** for a 1D example). Consequently, diffusion processes indeed can produce diversity in the observable traits of entities in a population from non-diverse initial conditions. More complex behavior like diffusion with drift or anisotropy in higher dimensions can be included readily via biased transition probabilities in Equation 2.2.1, see Figure 2.1, A and B. Other processes which we can think of include random or biased replacement or removal of entities (shown in Figure 2.1, C in a simple variation), which are indeed discussed in the literature [Godfrey-Smith, 2007; Papale, 2020]. In Figure 2.1, D, a blending of traits between two randomly selected partners has been included, which leads over time to a loss of diversity in the population. Random or biased addition or removal could, for instance,

---

[1] We will discuss the concept of "environment" in more detail in subsection 2.2.3.

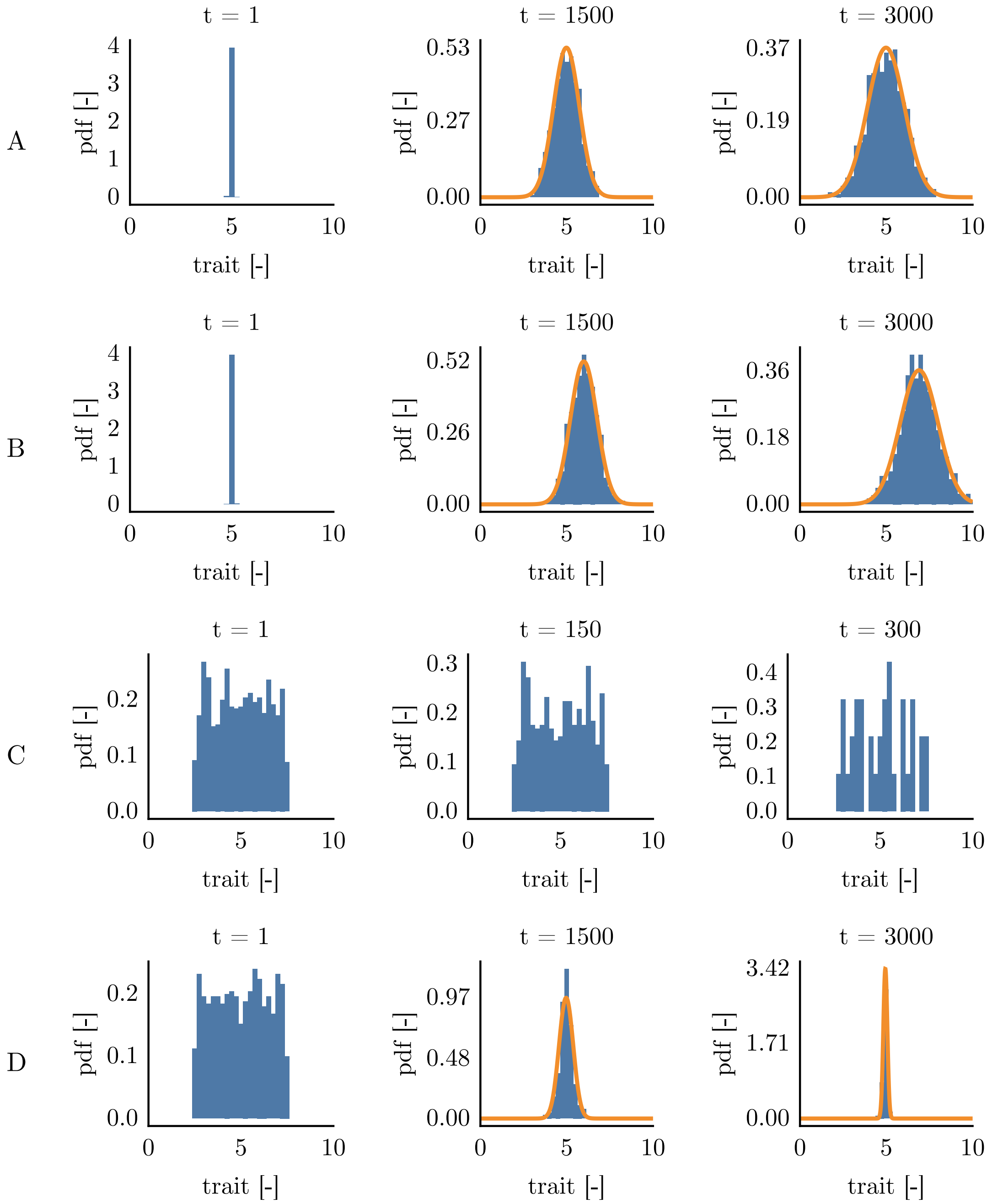


Figure 2.1: Individual trajectories for some simple evolutionary processes, obtained by simulating Equation 2.2.1 for a population of 1000 entities defined by a single scalar trait $a \in \mathbb{R}$, using the Gillespie-algorithm [Gillespie, 1976]. **A**: Standard diffusion with equiprobable steps in either direction of length 0.25, leading to trait-space exploration evident in the increasing variance over time. **B**: Same as **A**, but with biased step probabilities $p_- = 0.3, p_+ = 0.7$, leading to biased trait space exploration towards larger traits. **C**: Random removal of organisms starting from a uniform distribution of traits in $[2.5, 7.5]$. In contrast to **A** or **B**, this will eventually lead to the system's collapse. **D**: A blending process, also starting from a uniformly distributed sample. Two entities are chosen at random and the first one's trait is replaced by $a_1 = 1/2\,(a_1 + a_2)$. The process results in a narrowing Gaussian trait distribution over time by virtue of the central limit theorem. This process results in a loss of variation in the population.

be understood as the migration of individuals to and from the system, or as differential birth and death which are dependent on the traits of the entities. This latter combination is, of course, the one that forms the basis of biological evolution, but we will postpone a deeper discussion of this to later and keep the general, and abstract considerations for the moment. We also notice that both of these are only one of many possible evolutionary processes.

Interactions between environment and individuals Together with the earlier discussion on diffusion, we consequently identify interactions, with some environment the population is embedded in or between its individuals, as mechanisms that can bring about a change in the observable traits, number or identity in the population, in accordance with the definition of evolution given in section 2.1. These interactions are, for simplicity, often imagined as pairwise interactions, but this does not need to be the case, and higher-order interactions may also involve entities from the focal population and the environment. In summary, the processes examined so far only serve as phenomenological descriptions of how the distribution of existing traits in a population can change over time, with references to explicit mechanisms only entering later. The most interesting questions evolutionary theory sets out to answer have been left entirely untouched until now, however. Foremost among them is the question of the apparent organism-environment fit we observe in nature, i.e., how do apparent adaptations of organisms to their environment come about in an ever-changing world, and how does it produce complex biological compounds, like ourselves?

2.2.3 Dependency on external environment

In order to get closer to these questions, we focus on the transition rates between system states given by A and introduce an explicit dependency on an external environment. Environmental influences can include ambient temperature in chemistry or light and nutrient availability for a population of growing plants. These apply generally to the entire population and are often wrapped into boundary conditions of the system. Some environmental influences can preclude a subset of transitions entirely, thus making parts of the state space of the system inaccessible, or do the opposite and facilitate transitions into unexplored areas of the state space. While often wrapped into boundary conditions or system parameters for descriptional simplicity, such influences become explicit While these influences leave a signature in the trajectory of the system, it is hard to imagine how adaptations could arise from them alone. From the point of view of biology, this is more akin to abiotic influences like ice-age- or supercontinent cycles, or local temperature or precipitation. All these influences shape the trajectory of an evolving population but do not yield adaptations or complex structures. Indeed, closer consideration reveals that for such adaptations to be generated, which we readily understand are highly unlikely states to arise at random, more complicated interactions are not enough – rather, a new kind of process is needed.

2.2.4 Complexity – dependency on the population state

Self-organization in populations of interacting entities We now focus on individual entities in the population and the variation of traits at time t and introduce a dependency of the transition rates A on the state of the population at time t. As introduced above, this state can be interpreted as the distribution $p_\gamma(t)$ of individual trait tuples γ over the population. With this modification, the state change of a single entity over time depends on the state of other individuals in the population. This enables local feedback processes between the states of interacting individuals. Importantly, this opens up the realm of self-organization, pattern formation, and emergent properties in the population.

We can interpret such a system as a time-dependent interaction network, in which the individual organisms, particles or other entities interact with a set of other entities and in which interaction intensity and the set of interactors change over time. A reaction-diffusion system giving rise to Turing patterns [Turing, 1952] or Belusov-Zhapodinski patterns [Belousov, 1959], for example, can be seen as a time-dependent interaction network between particles that interact by chemical reactions and momentum exchange. Alternatively, and much easier to imagine, the interactions between the chemical species involved can also be described as a network.

Self-organization and emergence can lead to remarkably complicated global system states not reducible to the structure of individual entities in the population, i.e., these properties emerge as a consequence of the topological structure of the interaction network and the functional character of the interactions.

For natural examples outside of biology, we can think of the formation of sand-dunes [Herrmann, 2006], self-organized criticality in granular matter [Bak et al., 1987], convection patterns in fluids [Getling, 1998] or pattern forming, non-equilibrium chemical reactions [Belousov, 1959]. Because of their complexity, such systems are often studied algorithmically rather than analytically, with cellular automata [Wolfram, 2002] being a widespread tool that is in principle representable via Equation 2.2.1, although while they can be stochastic [Louis and Nardi, 2018], cellular automata where originally envisaged to unfold deterministically. Such algorithmic approaches can rapidly attain a complexity that is not tractable analytically and are therefore able to represent more realistic scenarios at the price of a loss of generality.

Self-organization and emergence as a building block of adaptation The main insight here is that interaction between individual entities in a population can yield complex, regular collective states that arise from unordered, individual interactions. Thus, it provides a mechanism for the autonomous emergence of dynamic complexity and consequently a principle, or at least starting point, for how the complex bodies of biological organisms can be created from simple building blocks.

Moreover, of all fields of science, regular patterns are probably most apparent in biology, so it's not surprising that Turing's seminal paper on pattern formation in reaction-diffusion systems [Turing, 1952] was inspired by biological development. Therefore, the properties of natural organisms can be seen as emergent phenomena that arise from the interactions of their own respective building blocks. Consequently, natural complex self-organization can only come about when the focal entities that make up the population referenced in section 2.1 are collectives of interacting elements, like populations of plants, animals, cells or molecules. Consequently, while we always talk of individuals when talking about (biological) evolution, these cannot be "atoms" in the true sense of the word but are conglomerates of other interacting entities themselves. It must be noted that self-organization and emergence always occur in systems that interact with their environment, i.e., they cannot occur in isolation [Haken, 2008]. This requirement is naturally fulfilled in biological populations, but also in many other systems in physics, chemistry or sociology, to name a few [Haken, 2008]. Moreover, the formation of such global phenomena in a population of interacting entities represents an attractor of the global dynamics, and thus the emerging population-level states are often quite stable against variations in initial conditions, even if the exact trajectory leading to them is not [Strogatz, 2018]. Such phenomena also usually appear on a certain spatiotemporal scale that is usually distinct from the scale on which individual interactions happen. The dynamics of the emergent phenomenon, therefore, can be typically described with fewer degrees of freedom than a full description of the population it arises in. In other words, the properties of individual elements of the system are no longer relevant to the behavior of the system as a whole. Conceptually, however, there is no reference to any particular spatiotemporal scale. In the sense discussed here, emergent phenomena can appear on any level of physical organization from the molecular [Haken, 2008] to the planetary scale and beyond [Stumpo et al., 2021].

Primitive adaptation in populations of complex systems The central corollary of this insight is that complex systems can be nested – complex systems can form other complex systems, i.e., the individuals in a population can be complex themselves, and should be expected to be so if they show adaptative state changes over time. Indeed, if we accept biological cells to be complex systems in the sense that they exhibit emergent phenomena and are made up of many interacting parts, we find that biological organization from molecular dynamics over cells to multicellular organisms and whole ecosystems is “just” a big collection of nested complex self-organizing systems.

Such systems can exhibit multiple different attractors which manifest as different emergent phenomena depending on initial conditions, boundary conditions, external forcing and system parameters [Strogatz, 2018]. Forcing from the outside on the state or parameters can lead to changes in the number and character of the attractors via bifurcations or the switch from one to another. In that sense, self-organizing many-particle, i.e., “complex” systems, can be said to adapt to environmental conditions and support diversity, which both are hallmarks of biological systems. In nested complex systems, this forcing can be exerted by other complex systems in the hierarchy and their self-organization, leading to feedback between layers of organization. Such reaction to external forcing can be interpreted as a rudimentary form of adaptation, although this does not yet allow for the kind of complex adaptation found in natural biological systems. It does, however, provide a starting point for further investigation and rules out simple linear systems as underlying evolving biological organisms. Rather, organisms undergoing complex adaptations will invariably be made up of other interacting entities themselves.

2.3 Complex adaptive systems

2.3.1 The concept of Complex Adaptive Systems

Simple complex systems compared to Complex Adaptive Systems The systems described above have a range of possibilities to adapt to environmental conditions in the sense that they react to environmental input in a specific way and produce a, often quite stunning, range of emergent organization: When environmental conditions change, they reorganize accordingly and eventually reach a new attractive set of individual states on which they remain until another change happens. The individual elements of the complex systems we discussed so far often are exceedingly simple, in particular, they need not show adaptation, but often follow simple behavioral rules or physical laws. This results in the range of emergent properties these systems can exhibit being strictly limited – determined by the properties of their constituents and the interactions they can engage in. In other words, while such “simple” complex systems can exhibit very complicated self-organization over wide ranges of parameters and boundary conditions, their reactivity to these conditions is generally limited to the, typically few, attractive sets within their state space and reorganizations of these state space features by bifurcation or crisis [Strogatz, 2018] is only brought about by some outside forcing on system parameters. The novelty so characteristic for biological evolution is thus not found in such systems.

A subclass of complex, self-organizing systems, however, shows a more adaptive behavior which, at first glance, can give the appearance of goal-oriented action informed by incoming signals – adaptation in the full sense of the word. Colloquially, this is much closer to what we understand as “adaptation”. Societies, ecosystems, cells and other natural systems are often grouped under this header [S. A. Levin, 2002]. Consequently, these systems are often called “Complex Adaptive Systems” [S. A. Levin, 2002; Holland, 1992]. We again hold on to the model of a dynamic interaction network to guide our imagination. We readily find that ecosystems, social systems in humans and animals, (artificial) neural networks, or economic systems also have network structures.

Properties of individual entities in Complex Adaptive Systems A distinguishing property of Complex Adaptive Systems arguably is that their individual constituents can adaptively react to conditions in their surroundings. Thus, they have at least a “simple complex” structure. In particular, they actively influence the connections they have to other entities and how they interact with them, while in simpler systems, this is typically fixed and intrinsic to the individuals. In that sense, the individuals in a population that forms a Complex Adaptive Systems exhibit some form of agency of their own – they actively process signals from the outside and modify the sources of these signals, i.e., their connections to the rest of the network. This in turn influences how their own state changes over time. Therefore, they react to environmental input by altering their interaction patterns, which can result in fundamental changes in the state space architecture and the phenomenology that the population as a whole exhibits. When compared to simple complex systems, we see that this changes the basis of how emergent properties come about. Therefore, complex adaptive systems show a wider range of phenomena and adaptations than simple complex systems. For example, learning processes in neural networks are associated with the formation and reinforcement of links between individual neurons and collections thereof [Tovar-Moll and Lent, 2016].

Individual adaptations in CAS as a two-fold process This change of interaction pattern can be seen as a two-fold process: First, individual agents can change with which other agents they interact based on their properties. Second, individual entities in the network can change their dynamics as a reaction to outside influences. In such a setting, the action of individual agents can often take the form of conditional activity – the agent does something under condition x and something else under condition y [Holland, 2006]. The action potential in excitable cells can serve as a simple example here, where ion channels in the cell membrane open and close at a precisely defined threshold potential [Hodgkin and A. F. Huxley, 1952]. Often the topology changes in the system are conditional too, for example, links in social networks are often formed or cut when an evaluation of the connected node by an agent yields a certain result, a behavior especially relevant in opinion dynamics [Peralta et al. [2022] and references therein]. In this case, the individuals are often described by a set of rules they use instead an integrated analytical representation of their full dynamics, i.e., we can identify the set of rules with the set of traits γ.

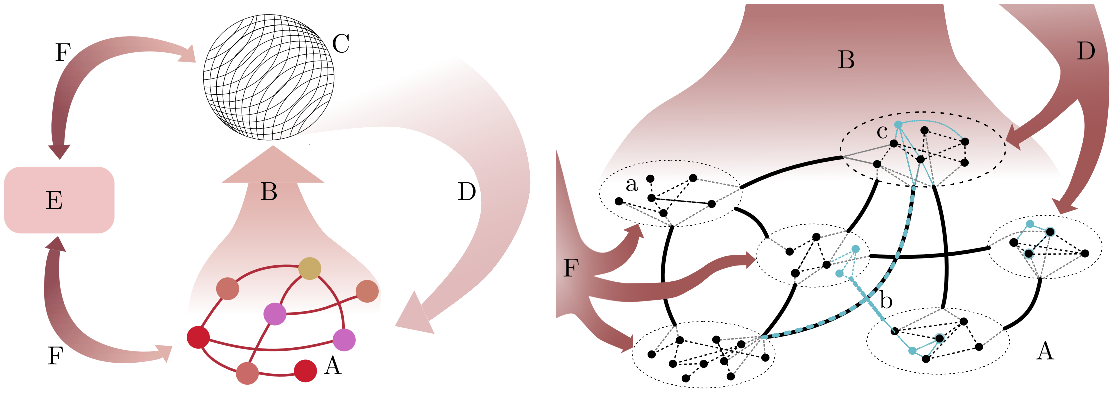


Figure 2.2: **Left**: A schematic representation of Complex Adaptive Systems. The individual elements of a CAS[2] are organized into an interaction network (A). Often, the individuals are heterogeneous in their properties. Their interactions are nonlinear and often conditional in nature. Additionally, there is a "background" environment E with which the system is in exchange (F, lower). This exchange constitutes a feedback loop in which entities change their local interaction environment and in turn, adapt to it. Through this interaction, collective behavior can emerge that cannot be isolated or reduced to individual elements (C), e.g., flocking or herding behavior in ecology, institutions in human societies and many more. Often, this takes the form of a new kind of interaction network (not shown). The emergent level of organization interacts with the environment as well (F, upper), creating a second feedback loop. Finally, the emergent structures can feed back to the individual behavior, yielding changes in behavior and possibly in the emergent phenomenon (D). (modified from NikNaks [2014].) **Right**: Zoom into the lower half of the left figure. As discussed in the text, CAS can usually be seen as nested complex networks (a), in which adaptation comes about by changes to the network structure in reaction to input from their local environment (c, light blue), which might be other nodes in the network, elements of the background environment (F) or a feedback from the emerging global pattern (D). New entities can emerge as network motifs and form connections with others by connecting internal nodes with internal nodes of other entities(grey lines in "a"). As a consequence of internal reorganization of individuals, these connections can change (b, light blue) and result in changes to global system behavior and the arising emergent phenomena.

Complex adaptive systems are, therefore, often nested complex systems, which we can envisage as interaction networks of interaction networks. Indeed, Complex Adaptive Systems can form Complex Adaptive Systems as part of their unfolding, a fact that originally yielded some heterogeneity in how different authors used the term. Think for example of a social system as an interaction network in which each node is made up of a network of neurons, or an ecosystem in which each species is made up of an interaction network of individuals which in turn are an interaction network of cells, chemical species and so on. These properties lead to Complex Adaptive Systems usually being modeled with agent-based computer simulations [Railsback and Grimm, 2019] instead of the more familiar differential equations.

2.3.2 Hierarchical architecture and environment

CAS have a hierarchical organizational structure The insight that Complex Adaptive Systems are usually nested complex systems, and thus “interaction networks of interaction networks”, introduces an organizational hierarchy into the concept. Indeed, many natural and artificial complex systems have been shown to feature a hierarchical organization – from neural networks over gene-regulatory networks and ecosystems to organizations and societies or the internet [Mengistu et al. [2016] and references therein]. Adaptations of the entire system arise as an emergent phenomenon from adaptations at the individual scale, which in turn comes about by a reorganization of the elements these are made up of and so on. Adaptation in the sense of a non-random, specific state change in reaction to environmental conditions, which includes learning and, by extension, “intelligence”, therefore is ultimately always a collective phenomenon.

CAS are self-organized on multiple levels of organization through feedbacks Because of the nested-network architecture of CAS, we in general observe self-organization on at least two, generally multiple levels (see Figure 2.2, right). The first level is naturally the level of individual network nodes which adapt their dynamics and linkage to the inputs they receive from other nodes or some environment process. This process entails self-organization of their internal elements and their interactions. From this local adaptation emerges self-organization and possibly adaptation on the higher network level. This, in turn, feeds back to the individuals, manifesting itself in a change in their interaction pattern. Such cross-level feedback is therefore identified as another hallmark of complex systems and particularly complex adaptive systems and is a direct consequence of their architecture (compare Figure 2.2, left). In CAS, these feedbacks are especially important because they are made up of more complex agents than “simple complex systems” and thus, their phenomenology is much richer. In particular, we can interpret this feedback as a restriction of the dynamics of the lower-level agents by the emergent properties at the higher organizational level, which, especially in biology, is important to stabilize the system’s structure. For example, multicellular organisms are made up of, a priori, independent entities (individual cells) which are cooperating to create the multicellular body [Michod and Roze, 2001]. Individual cells are thereby restricted to specific functions via the developmental process, itself a product of natural selection over long periods of time [Niklas, 2013]. Failure of this restriction essentially leads to a loss of multicellularity [H. Chen et al., 2015], i.e., cancer, and thus the destruction of the system as a whole.

CAS as embedded systems Additionally, the concept of an environment is always a part of CAS by definition. This is, firstly, because they are defined by their ability to adapt to changing conditions, i.e., the aforementioned background process is by definition assumed to exist and be outside of the system, and second because each element of them is an open system that exists in exchange with other such elements and their local background environment. The environment of each node is, therefore, a combination of other nodes it interacts with and the “global” environment the population is embedded in. This also means that the environment of each node is not only subject to change over time but subject to change effected by the individual entity itself, either by changing its connections to other entities or environmental states or by changing the environment itself, e.g., when organisms change their environments through their activities [K. Laland, Matthews, et al., 2016]. Therefore, the concept of environment, or, embedding into some other system, is inherent to the idea of Complex Adaptive Systems. Thus, CAS fulfills two properties of the three properties that characterize biological evolution - adaptation to environmental conditions and elements that are open systems.

Because of their capability to change their dynamics as a reaction to environmental conditions, which naturally encompasses biological adaptation, we can think of Complex Adaptive Systems theory as a physical background theory for biological evolutionary theory [Kauffman, 1992], because it provides the conceptual underpinning and a theoretical framework with which we can interpret and interlink evolutionary phenomena with the wider physical world.

A glimps into possible mechanistic underpinnings It has been shown that modularity [Clune et al., 2013] and hierarchical organization [Mengistu et al., 2016] in interaction networks can evolve when there is a performance penalty for network connectivity, based on the empirical observation of such costs mostly in neuronal networks. Additionally, it was shown that such an architecture increases the adaptability of the networks and that the emerging sub-networks solve sub-problems of the tasks employed to gauge network performance, although these tasks are known to have an a priori modular structure. Following the discussion before, we can interpret these “sub-processors” as emergent properties of the network and regard the network of these modules as a higher level of organization.

Given that costs for connectivity is a plausible phenomenon in many CAS (e.g., energy cost to maintain a complex metabolic network made up of many enzymatic reactions), the suggested mechanisms may thus be seen as important catalysts for the unfolding of CAS into “hierarchical layers” of emergent dynamics, each of which represents a complex (adaptive) system in itself. However, in Mengistu et al. [2016], the authors note that there are likely other, still unknown, mechanisms involved in the evolution of hierarchy in such systems, with connection costs being just one.

2.4 Complex Constructive Systems – CAS with complex adaptations

2.4.1 The concept of Complex Constructive Systems

CAS can only explain parts of biological phenomenology Complex adaptive systems as discussed above can yield a correlation between the individual traits and behavior in the population and their environment, and thus a form of adaptation. More than simple complex systems, their elements can change their own pattern of interaction, i.e., how they are connected to their environment and process external signals. This yields a more adaptive phenomenology on the population level because the interactions that give rise to emergent properties are subject to change themselves, i.e., adaptation on the system scale arises as a consequence of adaptation at the individual scale.

This alone, however, does not explain how complex and highly specialized traits of biological or cultural individuals arise, and why these often seem to be so uncannily well “designed” to allow these entities to survive and thrive in their environment. Moreover, it does not explain how continuous novelty can come about that allows them to exist in the wide variety of conditions we find in nature, or expand into new ones. The question, therefore, is how cumulative complexity and open-ended adaptation of individuals can arise in Complex Adaptive Systems, and what an “algorithm” for the generation of such complexity might look like.

Description of an algorithm for constructive CAS Given the examples above, this algorithm must be largely independent of the detailed structure of the entities, and arguably can only rely on some broad properties they all share. The observation that, in the examples described throughout this chapter, more complex functions and structures come about from simple ones may serve as a useful guiding principle. Consequently, an iterative approach appears to be a sensible generator for such structures. This could take the following general form:

- Select an entity to modify based on some criteria.
- Modify the entity based on its current state or system memory.
- Return the modified entity to the population.
- Repeat iteratively.

This process does not require reproduction, death or any other biological process normally associated with evolution. Indeed, the population in question can be a population of biological entities, but it could also be a population of ideas in a human's mind, a population of cultural concepts or economic strategies, or similar things.

Two general processes must be fleshed out – the selection of entities to modify and the modification process itself. Third, and perhaps most importantly, this algorithm can be seen as operating on some form of "memory" or "trait reservoir", i.e., a pool of existing traits that it continuously modifies, expands, or contracts. In the simplest form, this is the set of traits the organism has, giving rise to a population-level reservoir via set union. However, individual entities can also have some form of memory, such that both can interact in modification, e.g., when organisms reproduce sexually or when individual humans or organizations modify their behavior based on observed behavior in others in combination with their own assessment of a situation.

In summary, we identify this algorithm as one possibility through which changes in interaction patterns associated with individuals in CAS can come about. The main new element is the inclusion of some form of iteratively changed trait memory, which was absent before. In biology, this memory is associated with the inheritance of, e.g., genetic material, while in human culture, we identify it as the set of traditions or behavioral traits that are transmitted from one individual to another, and in individual humans or AI, the memory is built up from past experiences and modified according to comparison with the behaviors of others or some preconceived expectation.

Complex Constructive Systems as a subclass of CAS We introduce the term "Complex Constructive System" for the subclass of CAS that are capable of these iterative, memory-based modifications of entities and the formation of cumulative complexity over time. This highlights their inclusion into the concept of complexity. They form a subset of Complex Adaptive Systems in that the above algorithm allows for adaptations of individuals to external signals, thus fulfilling the basic requirement of CAS. Specifically, in the algorithm above, the selection of an entity or modification of its traits can depend on external signals, bringing about adaptive behavior.

For contrasting examples, we consider that, e.g., animals can exhibit culture, which is arguably Complex Adaptive System, but their societies and technological artifacts lack the cumulative nature and increasing complexity of human culture[3] [Dean et al.,

---

[3] This point is debated in the literature [Whiten, 2019]. However, cumulative culture in animals, where it is present, does not reach the generality of human culture but remains confined to individual traditions, e.g., bird songs [H. Williams and Lachlan, 2021].

2013], which we can identify as belonging to the set of Complex Constructive Systems. Similarly, traffic networks, power grids [Oughton et al., 2018] or ant colonies [Bonabeau et al., 1998] can be considered complex adaptive systems, but they lack the memory-based modification mechanism that learning systems or evolving systems exhibit. Indeed, a main example of CCS[4] is complex neuronal networks that can learn and adapt their behavior in an open-ended fashion and keep track of the performance of previous behavior and circumstances, in which, as a consequence, complex cognitive functions can emerge. These can be biological in origin, but the modern deep-learning-based AI models would arguably fall into this category, too. Therefore, we can see complex constructive systems as a conceptual unification of learning systems and evolving systems [R. A. Watson and Szathmáry, 2016].

Complex Constructive Systems can emerge in populations of simple complex systems, like Complex Adaptive Systems can, or in populations of Complex Adaptive Systems, thus creating nested CAS. Compare, e.g., prebiotic evolution on organic molecules with the coupled biological-socio-technological evolution we see in human societies [Fitzhugh et al., 2019]. This includes the possibility of nested CCS, i.e., multiple complex constructive processes can be active at the same time in the same population, or one can emerge as part of the dynamics of another. These additionally can be linked, e.g., by a shared process that selects entities for modification. An example is arguably the case with human cumulative culture, where the CCS of biological and cultural evolution unfold concurrently. While the principle algorithm given above is also applicable to simple systems that do not show self-organization or emergence, this will not yield open-ended adaptation or adaptive, cumulative complexity, because their range of dynamics is too limited, i.e., the potential extent of the associated trait memory is small. Rather, a change in the distribution of properties over the population over time is all that can be expected, i.e., the situation described in Figure 2.1.

2.4.2 Trait performance and trait discovery

Emergent performance evaluation and individual-based trait discovery Holland [2006] discusses that to show adaptation, in the ways described above, two processes must exist in a CAS – performance evaluation and rule (trait) discovery – and he assigns these properties to the agents themselves which are interpreted as classifier systems [5] Furthermore, Holland [2006] describes the trait discovery process as a recombination process of building blocks of traits that have worked in past interactions. Ignoring this latter part for the moment, the assumption that the entities themselves are the ones that implement rule discovery and performance evaluation appears to not be a necessary one. While the trait discovery process will necessarily be associated with individuals, performance evaluation can be an emergent process in the population, brought about by, e.g., competition or exchange of information. Finally, because of the open nature of CAS and consequently CCS, we have to interpret such statements stochastically, e.g., performance-based selection does not yield a deterministic selection process but rather can be imagined to bias the selection probability for modification towards entities with higher performance.

Performance evaluation Focusing on the performance evaluation process, we can think of it as some function that maps the current state of the entity to some orderable quantity

---

[4] Complex Constructive System

[5] Holland does not distinguish CAS from CCS, however.

c, scalar in the simplest case: $P(\gamma) = c$. Consequently, it will induce some form of ordering in the population. In the algorithm laid out in subsection 2.4.1, performance evaluation is associated with the selection of agents to be modified. If associated with individual agents, this process can take the form of some agent-internal model of the things they interact with and the consequences of this interaction. This automatically places complexity constraints on the agent - in particular, an individual capable of such information processing will be a CAS itself. For example, individuals can use cues produced by the behavior of others to assess the quality of different behavioral alternatives [E. Danchin, Giraldeau, et al., 2004]. A simpler example would be the conditional relocation of organisms, for example, seasonal or generally resource-based migration of animals. In such cases, organisms will evaluate their current state (spatial position) as disadvantageous and move to other places until they found a more suitable habitat – again an individual-based evaluation process of the current trait "location".

On the other hand, performance evaluation can also be an emergent property of the system. A simple example can be derived from Figure 2.1, C, where entities are randomly removed from the population. We identified the random removal process as representing some local environmental influence, and if this depends on the state of the organism instead of being purely random, performance can be associated with, e.g., survival time. This is just a restatement of the very natural observation of differential survival in natural populations [Hall and Hallgrímsson, 2013; Ridley, 2004]. Another purely emergent example is social norms, which determine what kinds of behavior, i.e., social traits, are acceptable in a group of individuals and against which individuals evaluate their own actions and those of others [Young, 2015]. Similarly, the number of sales or profit a product generates in a free market can result in the product being modified or removed from production if they are too low.

Such examples also reveal that, in fact, performance evaluation can happen for the complete entity or for individual traits it possesses, i.e., P can be thought to act on individual traits instead of on the entire entity at once in some cases: $P\left(g_k\right) = c_k$. The exact form of this depends on the structure of the evaluation process and the individuals. In particular, if interactions with the local environment are a consequence of a complex nonlinear interaction of individual traits, performance will most likely pertain to the individual as a whole, because the interaction function can, in such cases, not be inverted to unravel which trait yields which impact on the outcome of interactions. Such is the case in many biological settings.

On the other hand, if performance is strongly determined by one or only a small number of traits, the system can be simple enough to reduce it to the evaluation of trait performance. Arguably, such trait-based performance evaluation is most prevalent where the function P is a property of the individuals themselves, e.g., in learning by trial and error, and thus in cultural evolution.

As a corollary, we note, again, that the distinction between global and local performance evaluation functions depends on the point of view. If we consider individuals as being conglomerates of interacting elements themselves, an individual-based performance evaluation can be considered global *on the individual scale* because the individual itself would, in that case, be the complex (adaptive) system under consideration and the traits would be the elements that are evaluated. Due to the generally nested nature of CAS, what is considered an individual is often a pragmatic choice rather than a natural one. In biology, this observation gave rise to the concept of multi-level selection [Okasha, 2006; Lewontin, 1970].

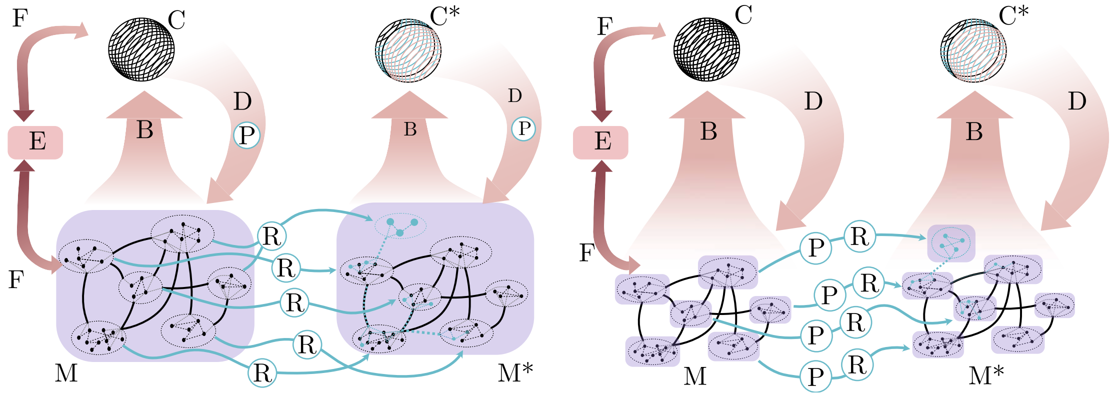


Figure 2.3: Schematic representation of Complex Constructive Systems in two different modes. **Left**: In the first case, the performance evaluation is an emergent property of the system at the population level and thus is part of the feedback of emergent properties D onto individual entities. As examples, we could think of differential survival or reproduction in ecology, or social norms in societies. Individuals (or their individual traits, depending on system details) change then by an individual-based trait discovery process R that makes use of the outcome of P, e.g., by genetic mutation or individual behavioral innovation or recombination of existing traits (light blue arrows, and light blue elements in the individual networks, compare Figure 2.2). Performance evaluation can change over time with the evolution of the emergent properties from C to C*. **Right**: The same system, but with performance evaluation as part of individual dynamics that acts in conjunction with the trait discovery process R, e.g., learning by trial-and-error.
Trait memory is a priori associated with the individuals but can be associated with the population if exchange between individuals exists. Trait memory evolves with the population in all cases, which yields iterative adaptation and the potential for complex structures beyond what non-CCS-systems are capable of. The trait discovery process can be random, i.e., independent of the effects of modification. More complex cases are possible when individuals are complex enough to maintain some form of predictive model based on experience, like in our species.

Trait discovery processes By necessity, trait discovery is associated with individuals, as it represents the modification step in the algorithm laid out in subsection 2.4.1. In general, we can understand it as a function that takes a set of traits (which can be the complete individual) and maps them onto a new set of traits: $R(\{g_k\}) = \{g'_k\}$. Which agent is selected for this procedure is thereby dependent on the performance evaluation process, but not explicitly represented here.

In Holland [2006], a genetic algorithm is suggested to create new traits from old ones by “effectively cross-breeding them”. This assumes that traits have a modular substructure that can be recombined to produce new ones, thus introducing another layer of organization. Indeed, pairwise exchange of discrete traits or recombination from an individual-based reservoir are examples of individual-based trait discovery processes, e.g., sexual reproduction or horizontal gene transfer in biology. Social learning or learning by trial-and-error with subsequent modification of behavior according to the outcome could be also seen as a trait discovery process. Biology implements trait discovery in a multitude of ways, from random mutations introduced during DNA copying and DNA recombination during sexual reproduction over epigenetic inheritance to social- and individual learning. Lastly, the background environment can influence trait discovery, too, e.g., ionizing radiation or the presence of some chemical elements which introduce errors into the DNA of biological organisms or modify their development. Therefore, the trait reservoir of each entity and the population itself can be considered open, i.e., subject to in- and outflux of traits.

Another feature of the trait-discovery process is that the newly created traits can be correlated with their expected performance, or not. Contrast, for example, random genetic mutations with conscious learning processes in humans and other mammals in which performance evaluation is not only applied to the trait or skill or behavioral rule that is to be modified but also to the one that potentially replaces it. This requires some form of prediction of the consequences of a trait change, and hence a high level of agent complexity. While common in human behavior, empirically we know that no such pre-evaluation takes place in DNA mutations (discussed more deeply in chapter 4), i.e., there is no correlation between the effect a specific mutation has on the carrier and its occurrence [Stearns and Hoekstra, 2005]. Complex Constructive Systems thus support a plethora of phenomena that implement the algorithm presented in section 2.4.

2.4.3 Memory, trait reservoirs and inheritance

The role of memory in the trait discovery process All the processes for trait discovery discussed so far make, in one form or another, use of a reservoir of traits from which they draw to adjust the current state of the individuals. This reservoir can exist on the individual or population level. Any individual-level trait reservoir can give rise to a population-level trait reservoir by means of set union if the trait discovery process allows for the exchange of traits between individuals. Otherwise, the trait reservoir will be bound to lineages of consecutively modified individuals (compare section 3.2). Additionally, the background environment can serve as a source of trait change, as mentioned before. In nature, these processes are mixed, e.g., genetic recombination and random mutation happen concurrently, and we can imagine similar processes to take place in human cultural evolution via the recombination of ideas learned from different sources and their change in the learner's brain through interpretation.

Trait exchange must be particulate We note immediately that, if trait discovery has a component based on exchange between individuals, a blending process of traits from different individuals will in the long run result in a degeneration of the trait reservoir's diversity and therefore a loss of adaptability, as Figure 2.1, line **C** shows. While this can be counteracted to some degree by an external influx of traits, an exchange or recombination process that is discrete and particulate, such that the recombined elements themselves remain intact, will maintain the variation in the trait reservoir and thus allow for higher adaptability. This is the basis for sexual reproduction in eukaryotes [Stearns and Hoekstra, 2005], compare Figure 3.3.3.

CCS generalize descent with modification Irrespective of its implementation, however, the trait discovery process as laid out above will make use of the current state of the organism and modify it, which we understand as a form of "inheritance". Consequently, the process yields a generalized form of "descent with modification", albeit one that does not make explicit reference to reproduction or destruction and recreation. Thereby, the concept of Complex Constructive Systems generalizes biological evolution. It does allow for the continuous transfiguration of entities over time (compare the Lamarckian view of evolution, see subsection 3.2.1 or cultural evolution, see subsection 4.3.2) as well as for imperfect reproduction in the biological sense. Other examples include the continuous modification of products by designers in a free market economy, based on sales or profits made.

This last example is of course an extension of cumulative culture, but does highlight that the concept of CCS goes beyond biological evolution. Therefore, performance-based death or reproduction, the classical account of evolution associated with Darwin's concept of "descent with modification", is only one way of creating a Complex Constructive System.

It is important to note that the trait memory changes over time in such an architecture, which allows, on the one hand, for the continuous addition and improvement of existing structures via iteration, but on the other hand also implies that parts of it will get lost over time, i.e., those which get modified or replaced by the trait-discovery process. In nature, this can manifest, for example, that the gene pool of a natural population changes over the generations, that cultural traditions evolve with time, available products or technologies in an economy get lost or are invented and so on. We note in passing that, because the trait memory is part of the system's state, its existence does not automatically invalidate the Markov property of some stochastic model of it.

On the nature of the memory In considering the various examples introduced until now – ecosystems or social networks and the individual entities they are made up of – species, individuals, traditions or cultural content – we understand that the trait reservoir can exist as the physical structure of the entities themselves, i.e., adaptation proceeds using existing traits (or their building blocks) as templates which are then modified or reproduced with modifications. Indeed, this is the general idea of everything described above.

Alternatively, they could also be encoded in some special structure that is read, subsequently decoded and the respective traits reproduced [Rocha, 2001; von Neumann, 1966; Barbieri, 2008]. Both possibilities exist in biology as well as cultural and technological systems. For example, learning by imitation can be seen as template-based rule discovery, while the genetic code or transmission of culture via human language represents code-based rule discovery. We will discuss the differences between these two principles and their consequences in more depth in subsection 4.1.7. Here, we simply note that template-based rule discovery is far simpler than code-based rule discovery because the latter requires specialized reading and decoding structures, and thus can be expected to be far more prevalent.

On the nature of the building blocks for traits and individuals

On the character of the memory's constituents Finally, we can give some rough characterization of the nature of the memory's elements, by considering the range of dynamical possibilities individuals must be able to attain to create continuous adaptation and complex functions. Because long-term conditions cannot be reliably predicted as a consequence of nonlinear dynamics, the possible adaptations that can be created by the trait discovery processes must be as broad as possible. A range that's too narrow will probably lead to a loss of performance across the population at some time, thus leading to a loss of further adaptation. In nature, this results in the extinction of the population, the disappearance of a product, product line or even entire branch in the economy or a decay or transformation of a society.

When discussing the expressiveness of programming languages, we have the hard notion of Turing completeness that tells us under what conditions a programming language or computational machine is capable of representing any algorithm [Rich, 2008]. In the abstract situation we are discussing here, and given that we are discussing the dynamical capabilities of physical entities, such a notion does not exist.

Indeed, there is no such condition for dynamical systems and in particular Complex Adaptive Systems, like a developing body, the brain or a social network. More to the point, there is no universal theory that could a priori predict what kind of structures could emerge over time in a Complex Adaptive System. As in all nonlinear dynamical systems, there exists a third realm that complements the known and the unknown – the unknowable [S. A. Levin, 2002].

Considering individuals as sets of traits that together bring about the interactions between individuals and the environment, we find that the possible set of traits from which individual functions can be "built" should be such that the greatest possible bandwidth of individual functions can be generated from them. This assures adaptability to the widest range of conditions on the one hand, and, in turn, the emergence of systems with a wide variety of emergent properties. In other words, simple systems incapable of forming complex, self-organizing systems will not result in CCS.

Limitations and organizational ways around them Assuming system autonomy, we find as a corollary that unless the individuals in question are Complex Adaptive Systems themselves which are capable of internal model formation, there will be no entity that could decide on what building blocks are suitable and which are not. Even if they are, such decision-making is only possible for very limited timescales because of the generally present nonlinearities. Rather, any CAS with complex adaptivity will "have to make do with what it got", and this may or may not allow continuous adaptation and complex structure formation under given environmental conditions. These limits are generally given by physical properties, e.g., stability of molecular bonds or structural stability under a change of acidity.

In many cases, the combination of different elements yields progressively new functions, however, such that a hierarchy of progressively combined building blocks, themselves forming the building blocks for new levels of organization, can rapidly expand the spectrum of possible interactions, and therefore the spectrum of adaptations to environmental signals. In this nested system, lower-level laws, do still apply and restrict what can emerge on higher levels, e.g., physical or chemical laws restrict possible biological emergent structures.

Considering this more closely, we find that, first, such a process can rapidly, probably after only a few levels, produce a system of building blocks that can produce more individual configurations than ever can be realized, by virtue of combinatorics. This assures adaptability in principle. Secondly, a trait discovery process on one level of the organizational hierarchy gets all the capabilities of the level below it for free, i.e., it does not have to find them itself but merely has to combine them into different forms to yield different functionality. This can indeed yield leaps in properties, as the interactions that create each level of organization are invariably nonlinear.

For example, individual amino acids can form protein chains which in turn can act as structural elements in cell walls or provide functions like adhesion and reaction catalysis, thereby far exceeding the physical capabilities of individual amino acids. Inspired by this example, we identify self-organization and emergence in addition to combinatorics as fundamental ingredients for the expanding functional capabilities in CCS. With this, the existence of physical building blocks that allow for such a scheme to function becomes a lot less wonderous, because it is traced back to the nature of nonlinear interaction in many-particle systems.

2.4.4 Principles of population dynamics in Complex Constructive Systems

Novel traits must yield higher performance or will vanish We will now consider the trait discovery process and performance evaluation process together and explore how they bring about complex functionality in a concrete population, as it is found in biology. We start out by observing that in order to stay in the population, a new trait, e.g., a new behavior must be evaluated to have high performance by the function P. This is because, depending on the system details, if a new trait has low performance, it will more likely continue to get modified or will be filtered out – in a biological population, for instance, its carrier will more likely die than others, or a behavioral trait will be discarded or modified by a learning individual in human society, while a product that doesn't yield enough profit will be phased out, often in favor of a modified successor. Traits that have high performance will therefore have higher lifetime within the population because they will stay unmodified for longer times. This, in turn, will lead to a growth of their frequency in the population over time (see Figure 2.4).

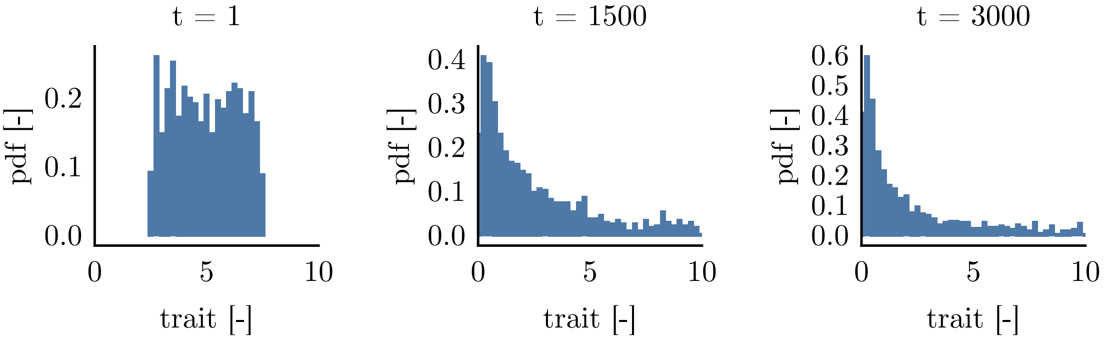


Figure 2.4: Illustration of a simple performance-evaluation-trait-discovery interaction in a population of simplistic, non-interacting individuals consisting of a scalar trait $\gamma = a$. Performance evaluation is given by $P(\gamma) = 1/a$. The probability to be subject to trait discovery is proportional to $P(a)$, such that smaller a are chosen less frequently. If chosen, the trait a is replaced with a random number a' drawn from a uniform distribution between 0 and 10. Lower values of a accumulate over time. Note that this system as such does not constitute a CCS, because the underlying individuals are too simple.

Traits enabling emergent structures must yield higher performance Because interactions between individuals or with the environment are mediated by the individual's traits, and these interactions are generally nonlinear and take place in a dynamic complex network, changed traits can enable new kinds of interactions that give rise to new emergent structures.

As mentioned before, this is the key process for the formation of complex structures with specialized functions. From subsection 2.4.4 and our understanding that feedback processes are common to self-organizing systems, it can be deduced that newly introduced traits that result in new emergent structures or properties must be such that they confer higher or equal performance to their carriers, i.e., the emergent structure must not be such that the performance of their constituents is harmed relative to the rest of the population where this emergent structure does not exist. Otherwise, the respective trait will be subject to trait discovery again, which likely will result in a loss of the emergent property, too. If that is the case, this trait can spread in the population over time, thereby increasing the prevalence of the new emergent structure. Elements of it will in turn come under trait discovery again, possibly changing the emergent's structures properties. The interaction between these emergent structures, in turn, can, as mentioned before, yield a new system that can be complex adaptive or -constructive in itself (compare Figure 2.5).

Iteration can yield complex functional specialization

Higher-level adaptation can emerge from lower-level CCS When a trait that enables the emergence of a novel structure spreads through the population, the number of these emergent structures will grow as well (compare the example in Mengistu et al. [2016] and others discussed throughout this chapter). In turn, it is more likely that some individuals that form these structures will become subject to the trait discovery process eventually, thus generating variation in the emergent structures by changing their elements. This, in turn, will result in changed feedback processes between the two organizational layers, which can change the performance evaluation process acting on the individual level. Depending on the effect this feedback has on the performance evaluation of constituents, some variant individuals which form a variant emergent structure may have higher performance, which results in a spread of the variant emergent phenomenon. These emergent phenomena do not need to be CCS themselves.

For example, we can imagine herds or swarms emerging from behavioral changes in individual animals that increase the tendency to associate with conspecifics. This results in higher performance for the individual because it was better protected from predators or could find sustenance more easily [Parrish and Edelstein-Keshet, 1999]. Further specialization of behavior, now within a new social environment, then led to differentiation of social structure or division of labor and so on via the iterative performance-based trait discovery introduced before. Similarly, the emergence of agriculture as a cultural trait in humans led to the emergence of large societies and ultimately civilization [Childe, 1950], which in turn allowed for a plethora of further cultural innovations.

The emergence of new CCS on existing ones While not all emergent structures in a CCS have to form CCS themselves, the ones that do are arguably some of the most striking phenomena in nature. Emergent CCS are, by definition, accompanied by new performance evaluation and trait-discovery processes that emerge from individual interactions. In such a system, internal specialization of elements is often the consequence. In nature, the higher level of organization often entails a modification of the constructive processes on the lower level that leads to further specialization and adaptation on the higher level. For example, multicellular organisms derived from more or less loose associations of individual cells, itself the consequence of new traits on the cellular level [Staps et al., 2019]. This, in turn, lead to a changed environment for each cell, such that their performance was evaluated differently from an independent cell. As a consequence, the trait discovery regime changed as well, because new cellular traits also had to benefit the multicellular association, or they would disappear again, resulting in the emergent multicellular structure being lost again over time.

Ultimately, this led to division of labor and the emergence of specialized organs like brains, eyes, hyphens or a digestive tract which each contribute to the performance of the whole structure by solving some subproblem, e.g., providing information about the surrounding, processing this information or acquiring the necessary energy to run these functions.

Sorting and constructiveness on different levels Consequently, if the dynamics of a CCS gives rise to an emergent structure or function, it also induces a form of sorting on these by virtue of the performance advantages or disadvantages it confers to its constituents [Vrba and Eldredge, 1984]. Note that this does not require a process of performance evaluation or trait discovery that operates explicitly on the higher organizational level. For example, swarming behavior may spread through a population and consequently increase the prevalence of swarms, but this is a consequence of the performance increase that traits allowing for swarms to emerge confer to individuals, not because there is a performance evaluation or trait discovery for swarms as a whole.

On the other hand, if this emergence gives rise to a new constructive process that has its own performance evaluation and trait discovery, this can lead to entanglement between the two levels and cooption of the lower-level constructive process by the higher-order emergent process, as it happened a few times in biological organisms [Michod, 2003]. Figure 2.5 shows a schematic overview of early stages of both of these processes.

Finally, with the emergence of a higher-level CCS, by definition, some form of new trait reservoir comes about, because the traits of higher-order entities are generally not reducible to the ones of their constituents.

Concurrent CCS and emergent conflict Two interesting possibilities present themselves when a new trait on the individual level emerges that increases individual performance and, at the same time, influences the performance of the higher-level structure. Either, it can increase the latter or be at least neutral, which may lead to the path of iterative modification and specialization as discussed before, or it may decrease the performance of the emergent structure, such that the individual's performance can be said to increase at the expense of the performance of the emergent structure. The classical examples for the latter case are cancer in multicellular organisms or parasitism in ecology, but we can more generally call individuals with such traits "selfish elements".

We can identify this as a major bottleneck in the emergence of higher-level CCS, because such selfish individuals may proliferate and thus increase in frequency in the population faster than their non-selfish alternatives, thus precluding the proliferation of emergent structures.

This problem can only be resolved if the performance evaluation processes on all levels of organization are aligned with each other, which in biology is often achieved by canalizing reproduction of the collective. In such a case, only the whole can reproduce, e.g., the entire multicellular organism and not its individual cells [Howe et al., 2022]. Other possibilities encountered in nature are policing, e.g., by the immune system or an institution in a society, or the enforcement of social norms [Kandori, 1992]. The problem of the emergence and stability of the individual-level cooperation that is required for the long-term stability of selfish entities is mostly studied with game theory and its relevance, in particular for biology, is discussed elsewhere in more depth [Bourke, 2011; Herdeanu, 2021].

2.4.5 Hierarchy revisited – A hypothesis on the unfolding of Complex Constructive Systems

With this, we arrive at a basic hypothesis on the unfolding of a Complex Constructive System endowed with building blocks as described in subsection 2.4.3 and a performance

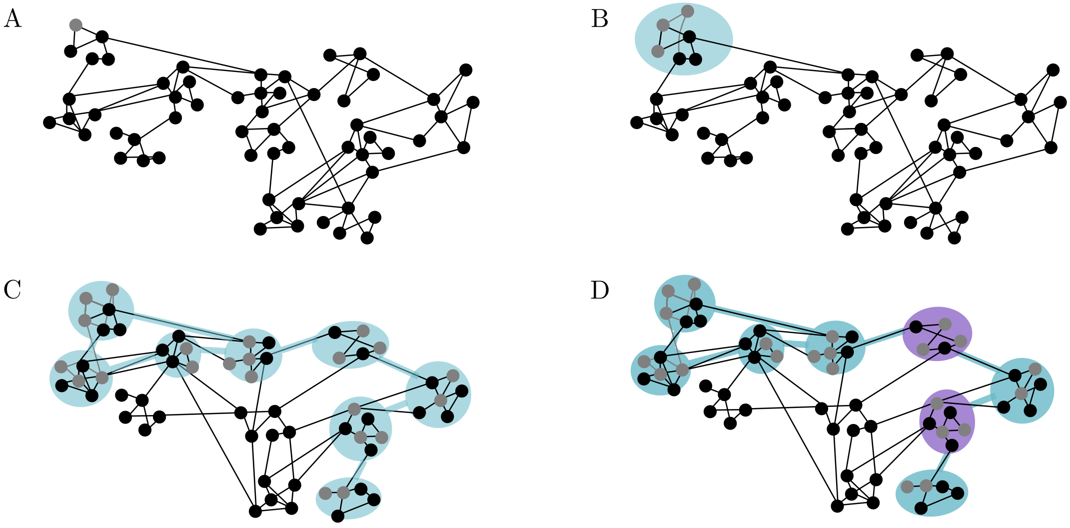


Figure 2.5: Schematic representation of innovation in CCS. **A:** Performance-based trait discovery leads to a change in individual traits (red) and, consequently, behavior. The new trait becomes an element in the trait reservoir of the system. **B:** If the performance of the changed individual allows it, the trait can spread in the population and give rise to new emergent structures and functions (light blue), which are formed by aggregates of individuals. **C:** The feedback between the emergent structure and the individuals can further enhance their performance, which can lead to a spread of the emergent structure or function in the population. These emergent entities often form again an interaction network on a higher organizational level. These interactions, however, are still mediated by individuals, although the behavior they give rise to generally is not reducible to individual properties. **D:** Further changes to individuals can give rise to changes in the emergent structure which in turn may give variation in traits in their population. Feedback between layers of organization changes the performance evaluation and trait discovery processes individuals are subject to, which in turn can lead to internal differentiation of the higher organizational structure if this leads to higher performance of the individuals.

evaluation process and trait discovery processes as described in subsection 2.4.2. This hypothesis has been discussed in the research group for chaotic,complex and evolving systems at the institute of environmental physics at the university of Heidelberg in the context of biological evolution, for which it has been fleshed out further in another work [Sevinchan, 2021], and variations can be found in the literature [Fields and M. Levin, 2020b].

The hypothesis has two parts. First, a system as described above has the capacity for producing hierarchies of organization, which progressively expand the functionality that the system as a whole can realize with each new layer. This appears as a natural consequence of the properties of its fundamental building blocks, namely the capacity for self-organization and emergence. This expansion of capability can, in conjunction with the repeated action of some performance evaluation- and trait discovery process, lead to the emergence of continuous adaptation to changing environments and the emergence of complex, specialized structures. A key element in the process is that this process makes use of some form of trait reservoir that is continuously modified over time, thus creating cumulative changes. Some of the structures arising in this system may be complex- or

Complex Adaptive Systems themselves, e.g., neuronal networks that arose in biological evolution.

As a caveat, this process can stall at any point, for arbitrary times, and there is no obvious, a priori sufficient condition that guarantees the emergence of a higher level of organization within any given timeframe. Rather, we might, from a particular trajectory, a posteriori identify the elements and phenomena which allowed for each generation of hierarchy to emerge, given these signals are not erased by later dynamics or organizational degeneracy. Thus, like in all nonlinear dynamical systems, our prediction capabilities are fundamentally limited, and the intellectual spheres of the "known" and the "unknown" are joined by a third, the "unknowable" [S. A. Levin, 2002].

Finally, this hierarchical structure allows for performance evaluation and trait discovery processes to emerge as part of the unfolding of an existing constructive complex system – much like cultural evolution by social learning and innovation emerged in biologically evolving populations. Complex Constructive Systems, like any other CAS, can be nested. In biology, this phenomenon has been recognized as "major evolutionary transitions" [Szathmáry and Maynard Smith, 1995; Szathmáry, 2015].

In the network picture, this hierarchical integration of structures into building blocks for new structures would manifest itself in modularity and the emergence of recurring motifs in the network, as discussed before in subsection 2.3.2. At least when a complex *adaptive* system emerged at some point, these motifs would be recognizable not only by their repetition of functionality but also, and perhaps primarily, by the existence of interfaces between their "inside" and "outside", which is inherent in the idea that these entities receive and process "signals" from their individual environment.

As a final remark, we note that the problem of the emergence of higher levels of organization, is still an open problem, in particular in biology, where its relevance is most apparent.

2.5 Summary

As our fundamental system of interest, we imagine a population of "agents", each characterized by a set, otherwise unspecified, observable traits $\gamma = (g_k)\,, k = 1, \dots, K$. Further, a network of interacting such agents provides a useful abstraction, because this concept is arguably capable of representing all patterns of interactions that can arise in such systems, including dynamic topology, multiple layers of organization and more, while remaining accessible to computational methods and being supported by a substantial body of theory [Boccaletti et al., 2006; Domenico et al., 2013]. While normally restricted to networks of pairwise interactions for simplicity, the concept as such is not limited in that way, and higher order interactions are an active area of research [Battiston et al., 2020].

Such networks are, in nature, always embedded into some "environment" which we, mostly for reasons of simplicity, do not represent explicitly, but rather as some form of the simplified background process, as boundary conditions or system parameters. We keep in mind, however, that the environment itself is generally made up of other populations of entities undergoing their own unfolding and evolution concurrently to the focal one, and so the interaction network extends beyond the boundaries any model will invariably place on it. We define the process of evolution in such a system as "the change of physical or functional characteristics or -organization or the distribution thereof in a population of entities over time" (see section 2.1).

The simplest form of change in such a system is a change in the distribution of observable characteristics. While such a system in principle can be represented by a Master equation under broad conditions, often, computer simulations are more practical than analytical solutions, at least when the systems are driven by more complex interactions or are high-dimensional. Interaction of an individual with its neighborhood in the network results in a state change and can therefore generally be seen as a function that depends on the current state of the focal entity, the state of its neighbors in the population, and the state of background environmental factors with which it interacts. As a simple example, we can recall the way the states of cells in a cellular automaton are updated.

The following discussion is guided by the question of how complex adaptations in natural biological populations came about. Additionally, we assumed that the complex, highly ordered and specifically interacting states we observe in natural populations of molecules, cells and individuals are the consequence of the autonomous unfolding of the system.

Consequently, in subsection 2.2.4, we identified self-organization and emergence in interaction networks as a keystone process that allows for complex, organized states on the population level to arise from unordered, and unguided, interactions at the individual level [Kauffman, 1992; Thurner et al., 2018]. Indeed, it seems natural to place such unguided and unordered interactions at the basis of the unfolding of the system, well known as they are from fundamental physics. Such systems are capable of a form of novelty, i.e., emergent phenomena – the observed ordered states are not encoded in the elements, but in their ordering and patterns of interactions. These emergent phenomena therefore are unique to the population level of organization and do not correspond to lower-level processes or quantities directly.

Furthermore, being nonlinear dynamical systems, complex systems generally feature multiple attractors which correspond to different emergent phenomena, and initial conditions or outside forcing can influence which self-organized state is reached as well as their number and character [Strogatz, 2001]. In that sense, these systems are capable of a limited form of adaptation to environmental conditions. Furthermore, they are characterized by feedback across scales, in which emergent properties act on the individual entities and vice versa. This, however, is limited by the form of interactions of which their constituents are capable, and this nature is static and does not change over time. Thus self-organization and emergence are fundamental building blocks for complex biological adaptations, but are not sufficient to explain the diversity and complexity observed, and, most importantly, the capacity of biological, social or cultural systems to continuously create new adaptations hitherto unseen in the history of the system.

As a second keystone, we consequently identified dynamical changes in the interaction patterns themselves – which on the one hand includes the network connections, but more importantly, changes to the functions individual entities use to interact with each other and the embedding environment. Owing to their greater capacity for adaptation and emergence, such systems are called “Complex Adaptive Systems” or CAS for short, see subsection 2.3.1. Complex adaptive systems are a subset of complex systems in that self-organization and emergence are central phenomena for them as well. However, they exceed the adaptive capabilities of “simple” complex systems, in that their constituents can change how they interact with their environment, i.e., their dynamics includes changes to the function through which they interact with their environment and active modification of the set of entities with which they interact. Think, e.g., about a biological organism learning to avoid certain plants or predators over the course of its lifetime, about humans (or other animals) forming or cutting ties with others in a social network or other such phenomena. In order to allow for active modification of interaction patterns, Holland [2006] postulates that there often is a conditional structure to it, i.e., the entities making up a complex adaptive network react in one way under a certain set of conditions and in another way under a different set of conditions and so on.

It must be noted that such behavior implies a more complex structure of the individual entities in the population than we normally know from simple complex systems. In CAS, we might suspect, these entities are often complex systems themselves, with their own emergent properties. Indeed, bifurcations in nonlinear dynamical systems can happen suddenly and in a "conditional" manner if the parameters of the underlying function are changed, thus showing that these entities at the very least are nonlinear dynamical systems. Thus, in subsection 2.3.2 we found that CASs can be conceived as nested complex networks, with feedback between layers being an important functional element for the ensuing adaptation. Which "layers" of this network are explicitly represented in a model of CAS and which are not is thus a modeling choice in general. As a corollary, CASs are, by definition, always embedded into some form of environment, which is distinct from the focal population but with which its individuals interact, and which we generally understand to be made up of networks of interacting entities itself, as mentioned before.

Finally, building on Holland's concept, we introduced the concept of "Complex Constructive Systems" as a subset of CAS which are endowed with a process P that assesses the "performance" of individuals or traits, i.e., maps them to some orderable quantity – called performance evaluation – and a "trait discovery process" R which produces new traits *from existing ones.* The latter part allows for the iterative modification of functionality that ultimately can lead to the highly specialized adaptations we find in nature and the emergence of the complex functional systems they are based on, e.g., the different specialized organs of multicellular organisms.

The system is consequently envisaged to use the outcome of the performance evaluation to, typically stochastically, determine which traits or entities are modified by the trait discovery process. We found that these processes can emerge from the interactions within a population (i.e., a Complex Constructive System emerges as a result of the unfolding of a Complex (Adaptive) System) or they can be inherent to the entities' dynamics (one Complex Constructive System emerges on top of another one), although in general, they always exist as part of and generate cross-layer feedbacks, as is natural for complex systems.

The deciding factor is that such systems, via their trait discovery process, draw on a form of "memory" of traits from which to build new ones, which, depending on the discovery process' details, can be the set of traits available in the entire population (e.g., when discussing horizontal gene transfer or communication), but also some form of individualized memory (e.g., in individual learning from experience or clonal reproduction).

Thus, we are faced with a generalized form of Darwin's "descent with modification" which, however, is not bound to self-reproduction. Accounts of biological evolution that are similar to the Complex Constructive System concept have been put forward in the literature – mainly with a focus on getting rid of the reproduction of organisms as a necessary condition [Papale [2020] and references therein].

The iterative mechanism of performance-based trait discovery was also discussed in some more detail in subsection 2.4.4, in which the process of how complex adaptation can come about as the iterative specialization of emergent phenomena was fleshed out some more. Noteworthily, nested CCSs are not trivially stable, because they can produce conflicting performance evaluation on different levels, which in biology is called "evolutionary conflict" [Stearns and Hoekstra, 2005].

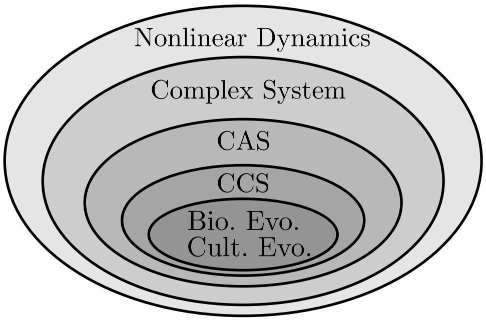


Figure 2.6: Venn Diagram of how different theoretical concepts used in this chapter are related to each other. We recover Complex Constructive Systems as being Complex Adaptive Systems endowed with additional processes, which in turn are a special case of complex systems. This concept in turn is based on the study of nonlinear dynamical systems Phenomena like biological or cultural evolution are recovered to be instances of CCS.

Therefore, we end up not with a one-sentence definition of evolution, but with a functional definition, which asserts that a system that allows for Complex (Adaptive) Systems to emerge as a consequence of the interaction of its elements and in which some iterative performance evaluation and rules discovery process are active, has principally the same dynamic capabilities as biological evolution.

We shortly discussed the nature of the building blocks involved in such a system, thereby arriving at the notion of a continuous scaffolding process in which one Complex (Adaptive) System yields building blocks which in turn form a second Complex (Adaptive) System and so on, limited by the physical conditions they are subject to, e.g., energetic demands for certain chemical reactions or stability against temperature (subsection 2.4.3). Finally, in building a general hypothesis on the unfolding of Complex Constructive Systems, we found that while they can create progressively more complex layers of self-organized phenomena, there is no obvious way to predict when and how fast one layer emerges from another one. In keeping with their nonlinear nature, such systems thus are only predictable to a limited degree, and investigating them always contains "unknowable" elements. Moreover, such systems generally are a conglomerate of nested or interacting CAS, some of which can have constructive qualities, while many will not.

Biological evolution, consequently, is identified as a special case of Complex Constructive Systems which, as we shall see in the next few chapters, has created a plethora of nested complex constructive- and Complex Adaptive Systems.

3. Natural selection, neutral evolution and biological Complex Constructive Systems

We now move on from the abstract concepts of complex adaptive- and complex constructive systems presented in chapter 2 and focus on biological evolution, which is arguably the most well-studied and most complex CCS. The focus of this chapter is how performance-evaluation and trait-discovery processes look like in biology and how these relate to older concepts like Lamarckian and Darwinian evolution or neutral selection. The identification of the performance evaluation and trait-discovery functions in different proposed evolutionary processes will thereby highlight how specific evolutionary processes emerge from ecology.

In subsection 2.4.2, we introduced performance-based trait-discovery as the fundamental process underlying complex constructive systems and motivated in the following sections how this allows for the emergence and continuous improvement of complex, specialized functionality. The performance-evaluation process P and the associated trait-discovery process R are kept very general in that definition, such that a variety of natural processes can play these roles. Of special interest, more so than the individual processes that fulfill these roles, is the interaction between them and the character of the trait reservoir they build upon.

In discussing them, we restrict ourselves to a description of these effects and the reasoning behind them. While much mathematical theory has been produced especially in the field of population genetics, this is arguably more necessary for empirical application than for the fundamental understanding of the architecture of natural CCSs.

Reproduction as the hallmark of biology A hallmark of biology in contrast to other complex constructive systems like social or technological evolution or learning systems is that biological individuals engage in physical reproduction, i.e., they produce copies of themselves (asexual reproduction) or combine traits with other individuals to produce offspring (sexual reproduction). Additionally, biological organisms are not permanent. We are familiar with variable but finite lifespans in multicellular organisms like animals and plants, but evidence shows that even bacteria, long thought to escape aging by dividing into two identical daughter cells, are subject to senescence and a correspondingly increasing mortality rate with age [Steiner, 2021]. Even if this was not the case, interactions with the surrounding ecosystem bring about a certain probability of damage and death for all organisms that's inescapable, as does failure to acquire enough energy and sustenance to maintain their own structure. With reproduction comes inheritance as a natural consequence, such that trait memory is associated with individuals and lineages in the asexual case, plus the local population in the sexual case.

Hierarchies and the complex constructive nature of biological evolution Additionally, we understand that the biosphere creates and maintains diversity of function and form over time and that biological populations tend to expand into unoccupied landscapes and find ways to use unexploited resources for growth and reproduction. The biosphere is thus capable of producing innovation, adaptation, and reorganization of existing elements.

This is most apparent in the fossil record. Fossil biological organisms, while at times bewilderingly different from contemporary ones, nevertheless show features that identify them as being related to modern animals [Stearns and Hoekstra, 2005; Ridley, 2004]. For example, many theropod dinosaurs have a very similar skeletal structure to birds with a great number being largely feathered as well [Unwin, 1998; Norell and Xu, 2005]. Similarly, the ancestors of modern cetaceans can be found in a group of terrestrial mammals closely related to even-toed ungulates [Shimamura et al., 1997; Chebii et al., 2021]. Over larger timescales, it becomes apparent how different organs and body plans evolved from less specialized forms. For example, while aforementioned theropod dinosaurs had recognizable feathers, early versions of them lacked lift-generating adaptations that true flight feathers have [Dimond et al., 2011; Sumida and Brochu, 2000], which only appeared when flying or gliding behavior became a part of the organism's behavioral spectrum and so lift-generating properties of body parts came under selection.

Another reason why the unfolding of the biosphere can be seen as a complex constructive system is that it generated, starting from organic chemistry [Luisi, 2019], a plethora of complex adaptive and complex constructive systems which became hierarchically nested into each other. For example, simple cells established the genetic coding system that all organisms on the planet employ to this day, with some slight variations [Koonin and Novozhilov, 2017]. Their evolution over time gave rise to eukaryotes, a new form of Complex Adaptive System that came about via symbiosis of multiple prokaryotic cells [López-García et al., 2017]. From these, multicellular organisms eventually evolved, yet another form of Complex Adaptive System. And these, finally, evolved to become internally heterogeneous with specialized organs with different functions. Some of these structures, e.g., the immune system or neuronal networks, can be seen as complex adaptive systems in their own right, with the latter eventually giving rise to culture. While these "evolutionary transitions" [Szathmáry and Maynard Smith, 1995] arguably are of paramount importance for the understanding of the biosphere but also CCS in general, in this chapter we are concerned with the other end of the spectrum, namely how different performance-based trait-discovery processes operate on individual entities.

In chapter 2, we discussed all these phenomena as defining features of complex constructive systems, and so biological evolution and the unfolding of the biosphere can be seen as the prototype for the concept of complex constructive systems, endowed with a special form of performance-based trait-discovery that generally is based on reproduction.

3.1 Elements of ecology

As a basis for further discussion, we again envisage a population of individuals, each defined by a collection of traits $\gamma = \{g_k\}$, which interact with each other and with some elements of a background environment e. This population is therefore embedded into some larger system and is extended in space and time. Furthermore, because we discuss biological entities, these individuals reproduce sexually or asexually. For the sake of simplicity, we represent the emerging ecological network, again, as a network of pairwise interactions, although we understand that, in general, a hypergraph (i.e., a graph that can have edges containing many nodes [Battiston et al., 2020], thus formally allowing for interactions of multiple entities at once), would be more appropriate [Majhi et al., 2022]. For example, a single organism could interact, over some time interval Δt (e.g., the average lifetime of a group of organisms), with various predators, prey, symbionts, or conspecifics on which it feeds, with which it mates or whom it fights with, or many other things. On the abiotic side, organisms contribute to a change in the chemical composition of soil, water, or the atmosphere, they burrow or bulldoze or build complicated structures like coral or sponge skeletons and more. Therefore, both biotic and abiotic elements of the ecosystem have to be included in its dynamics.

Ecological niches of individuals and groups and ecological networks The links of a node in the interaction network describe the pattern of interaction an organism engages in with its environment. We call the set of links the *ecological niche of the individual.* The literature knows of multiple concepts of ecological niches [Leibold, 1995; Pocheville, 2014]. The concept applied here is close to the one described by Elton [1927], which can loosely be summarized as "the role an organism plays in its environment".

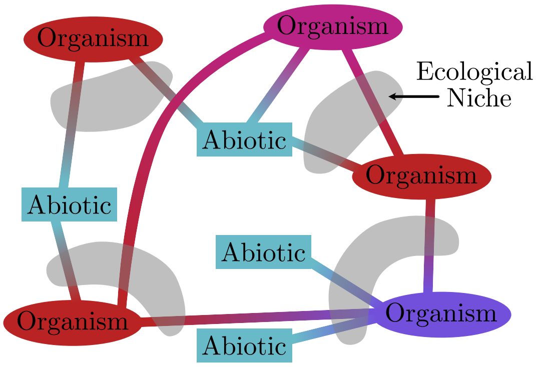


Figure 3.1: Agents and abiotic entities can be seen as nodes in an interaction network, for simplicity reduced to pairwise interactions here. A single agent engages in a multitude of interactions with other agents and abiotic entities, which are influenced by its traits, with differences in traits being indicated by color. The set of interactions of an agent with its ecosystem (indicated by the grey shaded areas) we define as "ecological niche". Note that the network can change over time, and thus ecological niches are dynamic as well. The individualistic concept of a niche can be extended to a definition of groups by summarizing the links each individual in a group has by means of a set union, e.g., for all the red organisms in the adjoining illustration. Note that this group definition does not entail that each individual of a group engages in all interactions identified for the group - there can be individual variations based on local conditions.

The niche concept is normally associated with groups We notice in passing that the concept of an ecological niche is typically applied to a group of organisms instead of individuals, normally species [Pocheville, 2014], where individuals are only considered representatives of an equivalence group of organisms, normally a species. However, a definition on individuals as given above is readily extended to a definition on groups by a set union on the links of each individual in the group. Moreover, the niche concept is not uniquely defined in the literature, and different concepts are applied depending on the problem at hand. For example, the niche concept of Hutchinson [Hutchinson, 1957] focuses on the properties of the environment that allow for a species to survive and reproduce, while the Eltonian niche mentioned above emphasizes the role it plays in the ecosystem.

It must be noted that an ecological network as presented here is always a temporal aggregate of the myriad of interactions organisms engage in at every given moment. Therefore, when extending the niche concept to groups, we extend it to groups that have similar individual niches, i.e., have similar phenotypes. Individual variation in phenotypes or environment will thus be integrated as noise.

The network picture above also reveals readily that the ecological niche depends on the environment an organism exists in. Thus, in general, all the different biomes or ecosystems a group of organisms occupies must be considered to get a picture of the true ecological niche a species is capable of filling. Therefore, there is not one ecological niche, but in fact, a realized ecological niche for each organism, given by its interactions in its current environment, and a fundamental ecological niche which is comprised of the union of all possible such realized niches, and which is generally unknowable, and can only be estimated on the basis of currently realized ecosystems. The concept of realized- and fundamental niche originated with Hutchinson's niche concept [Hutchinson, 1965], but readily carries over into the network-based concept.

Niches are dynamic entities In general, the niche as defined here is dependent on the time interval Δt over which the network has been aggregated. The niche may vary, for instance with the seasons from one year to another, but may also change due to evolutionary adaptation, plasticity, or changing ecosystem composition. A consistent definition of an ecological niche, therefore, relies on a timescale separation between the ecological timescale Δt and the timescale on which the niches change due to these factors. In biological terms, this implies a separation between ecological and evolutionary timescales, which more modern research has shown to not be guaranteed [Schoener, 2011]. Rather, these timescales often overlap, such that evolutionary and ecological processes interact, i.e., evolutionary processes become ecological agents [Pelletier et al., 2009].

Environment and individuals as open systems Because reproduction is a fundamental element of any biological system, biological individuals must be open systems. More than that, they not only exchange material and energy with their environment but actively acquire resources for the purpose of reproduction. This, however, does change the surrounding ecosystem by changing the availability of these resources in the future for all other organisms that use them. At the same time, they excrete waste material, spoil some resources or change their environment in other ways through their activity [K. Laland, Matthews, et al., 2016]. As introduced in general in chapter 2, ecology and evolution are therefore not conceptually separable from some environment into which they are embedded. Rather, the exchange of material and energy and the change of their flow through the ecosystem by virtue of organism action are central to our understanding of such systems, as will be discussed in more detail below and in chapter 6.

A short note on the concept of species We mentioned species a few times along the way, which warrants a closer look at the concept. It must be noted that, like with ecological niches, there are quite a few species concepts that are applied in different situations [Rosselló-Mora, 2001; Mayr, 1976; J. W. Taylor et al., 2000; Mayr, 2000b]. A species can be considered a group of organisms that can be considered equivalent according to some criterium, such that they can be grouped together. For example, the biological species concept defines a species as a "group of interbreeding natural populations that are reproductively isolated from other such groups" [Mayr, 2000b]. This, however, does not apply to any asexual organism, because these do not interbreed. Likewise, a physiological species concept, by necessity used in paleontology, uses morphological differences to define SpeciesNote, while the ecological species concept defines a species as a group of organisms that are adapted to a set of environmental resources or conditions [Ridley, 2004]. Each one of those, for the same group of organisms, may yield somewhat different groupings. Many, are, however, connected to each other by the fact that they are all the product of the evolutionary trajectory of a group of organisms.

For example, the morphology of an organism reflects its ecological niche, i.e., its activity patterns and exploited resources. Likewise, a group of interbreeding individuals shares an evolutionary history, thus showing similar morphology and so on. It must be noted, however, that this similarity hinges on the process of inheritance, or, more generally, how the trait-discovery process of the underlying CCS functions and interacts with the memory of the system. For example, some behavioral aspects of organisms are determined genetically, but others are cultural artifacts that must be learned, thus being disconnected from the genetic pool of the lineage or population. A species defined purely via social traits and interactions might consequently not reflect well, e.g., morphological differences. Species are therefore always pragmatically introduced classifications, with any particular species concept being more useful in some situations and less so in others. Because CCSs are fundamentally based on individual changes, however, the concept will largely be ignored in this chapter. Where it is used anyway, it simply designates a class of organisms with similar properties.

Ecological interactions Organisms interact with their environment in qualitatively different ways. The most well-known kind of interaction is probably competition for resources, which naturally arises from the fact that reproducing organisms have to acquire the means for reproduction individually. Competition is also arguably the most well-studied such interaction, with phrases like “struggle for survival” that are used to describe evolution hinting at its importance. A well-established phenomenon in this context is the “competitive exclusion principle”, i.e., the assertion that no two species can have the exact same ecological niche, i.e., exploit the same resources at the same time at the same place, or one will eventually disappear [Hardin, 1960]. In nature, this manifests in even closely related species interacting with their environment in slightly different ways, i.e., in niche differentiation [Begon et al., 2005]. It must be noted that in modern ecology, the principles that underly the coexistence of species have been found to be more complex than simple interspecific competition, leading to a more nuanced view of how different types of organisms coexist in nature, with new research being done continuously [J. M. Levine et al. [2017] and Gravel et al. [2011] and many more].

Aside from competition, there are other classes of ecological interaction, e.g., mutualism, commensalism, Amenalsim, or Antagonism, representing different effects on the interaction partners. These represent a classification scheme for pairwise interactions by whether the reproduction rate and survival of one or both interaction partners increase, decrease, or experience no effect on average.

Thus, the ecosystem network we built upon is generally composed of many different classes of interactions between biotic elements, often mediated by an organism’s or species’ effect on abiotic elements or by other organisms, and each individual engages in many different interactions concurrently. Because the environment for each organism invariably contains many other organisms, natural biological evolution thus is almost always *coevolution* between two or more groups of organisms, and not just adaptation to a resource-carrying background environment.

3.2 Lamarck and Darwin through the lens of complex constructive systems

Different processes have been proposed that can constitute a biological CCS The observation that populations of biological organisms change over time is quite old, and precedes the work of Darwin, Wallace, or Lamarck, by generations [Allen, 2014].

Indeed, the empirical observation from fossils, but also in the wild, that organisms continuously adapt to their environment precipitated the question about mechanisms for such change. Historically, quite a few of these have been proposed that can result in performance-based trait discovery. The scientific process has thereby not been linear, with many such proposals having been favored within the scientific community, discarded and then revived in another context[1]. Among the most well-known, aside from the Darwinian concept of evolution by descent with modification, are Lamarck's evolution by inheritance of acquired characteristics [Lamarck, 1830], Baldwin's evolution by learned behavior [Baldwin, 1896] or Waddington's genetic assimilation [Waddingtion, 1942]. Indeed, while all of these processes have at various times been discarded in favor of others, all have been revived in a modified form, thus warranting their inclusion. Moreover, they represent different, but interacting, trait-discovery processes that can be active simultaneously in a population and form their evolutionary trajectory, compare Figure 2.4.4. All explanations presented here start out with some form of change in the environment of an organism. In biological terms, this results in a change to the ecological niche the organism exists in, which may result in it becoming less able to survive or reproduce. In the CCS language introduced in subsection 2.4.1, this entails a change to the performance of the individual. As a result, the organism may more likely become subject to the trait discovery process and thus, evolution.

3.2.1 The Lamarckian concept of evolution

Use and disuse of organs and increase of complexity The Lamarckian explanation for evolutionary change rests on the idea that a change in environmental conditions leads to changes in the use and disuse of organs by an organism. This, in turn, leads to changes in the form or function of this trait [Lamarck [1830] according to Handel and Ramagopalan [2010]]. Textbook examples include the lengthening of a giraffe's neck because it always stretches it upwards in order to reach leaves on high trees, or a blacksmith acquiring strong muscles through their work. On the other hand, the disuse of organs would lead to degeneration and eventual loss. Therefore, according to Lamarck, traits of organisms change via the preferred usage patterns, which arise from the exposure of organisms to different environmental conditions. In the network picture shown in Figure 3.1, this amounts to changes in the strength or assortment of interactions of organisms with their surroundings over time. Variation in phenotypes, therefore, is created in response to usage patterns, i.e., it can be said to be directed.

Secondly, and in modern times of lesser importance, Lamarck posited that organisms become more complex over time by a "prime cause" [Handel and Ramagopalan, 2010], which appears as a reflection of the contemporary thinking that evolution was directional from "lower" to "higher" animals and finally to humans. While incorrect in this form, it is mentioned here for completeness but is largely ignored otherwise.

---

[1] See, e.g.,Gissis et al. [2015] for a discussion of different aspects of Lamarckism

Inheritance of acquired characteristics The concept most closely associated with Lamarck is the inheritance of acquired characteristics his view professes [Burkhardt, 2013; Handel and Ramagopalan, 2010]. In Lamarck's view of biological evolution, the characters formed by the use and disuse of organs are inherited by offspring. Therefore, a blacksmith's children would inherit the strong muscles acquired through his work, and a giraffe that acquired a long neck because it lived among tall trees would have offspring with naturally longer necks. Consequently, there is no branching in a lineage of Lamarckian organisms that would be associated with different characteristics. Rather, a single lineage would change over time without a priori restrictions to what changes can happen and what cannot. Therefore, Lamarckian inheritance would not give rise to a tree of life in the sense we know it. The inheritance of acquired characteristics in the Lamarckian sense has been called "soft inheritance" [Jablonka, 2017; Dickins and Rahman, 2012].

More deeply, behavior, learning, and their interplay with environmental conditions play the primary causal role in the Lamarckian concept of evolution. An explanation for the existence of morphological species would consequently rely on different usage patterns in a given environment. Such a scheme in principle would allow for one species to change into another over time, and for organisms actively shaping their evolution by changing their behavior. Speciation and diversification of organisms, e.g., of Darwin finches on the Galapagos islands, would therefore rely on environmental differences alone. Moreover, diversity would be reversible if the environmental conditions would be homogenized. Diversity in the Lamarckian view would be consequently very malleable.

Lamarck and contemporary discussions Lamarck's ideas have been largely reduced to the concept of inheritance of acquired characteristics although this close association is mostly a fallacy [Burkhardt, 2013; Gould, 1992].

For a long time, Lamarck has been seen as conclusively refuted [Mayr, 1972]. The argument thereby relies on the separation of germline and soma in multicellular organisms and the experimentally verified (mostly) Mendelian character of inheritance in sexually reproducing organisms and by the later discovery of the central dogma of molecular biology [F. Crick, 1970]. In particular, it has long been thought that there is no connection between an organism's germline cells and its somatic cells that would allow for the inheritance of acquired characteristics.

However, several discoveries in the field of epigenetics established that, to a limited degree, inheritance of such characteristics is possible (see subsection 4.3.1) and can indeed play a role in the inheritance of characters relevant for survival and reproduction [Lind and Spagopoulou, 2018]. In prokaryotes, horizontal gene transfer is an important mechanism that yields novelty in traits but relies on the uptake of DNA from the surrounding environment, e.g., in the emergence of antibiotic resistance [Sun et al., 2019]. These and other findings have been variously interpreted as constituting Lamarckian inheritance [É. Danchin et al., 2011; Jablonka and Lamb, 2008].

Similarly, it has been discussed if cultural evolution proceeds via Lamarckian inheritance, when we use social learning as its fundamental trait discovery process [Kronfeldner, 2006]. Therefore, the concept of soft inheritance has garnered new interest in modern evolutionary thought (compare chapter 4), leading to a revival of Lamarckian ideas in a modern context.

3.2.2 Lamarck and CCS

Performance evaluation Performance evaluation as discussed in subsection 2.4.2 in the Lamarckian view is therefore associated with individual traits and organs, and performance is assigned based on a discrepancy between the way an organ is used and its current structure, in other words, how this organ fulfills the requirements the organism's actions place upon it. When this performance is low, it will come under increased trait discovery and will change over time. When there is no usage for an organ, we can assign a low performance, too, but because there is no requirement associated with it, trait discovery will just lead to its degeneration over time, culminating in its eventual loss. It must be noted that this process is continuously active during the lifetime of the organism, and is not "evaluated" at a specific stage of its lifecycle.

Trait discovery The trait discovery process, in turn, is envisaged to be based on use and disuse, i.e., organs and structures which are used more because of a particular behavior of the organism will develop over time to be better suited for this task, i.e., they will become more specialized. New organs, therefore, arise from new requirements, which in turn come about by environmental changes. It appears that new organs must be, in some way, based on existing structures which by virtue of the above mechanism, are transmuted into another form. Inherent to this view is the idea that evolutionary change is generally positive for the organisms, because, through use and disuse, organs and traits will change to serve their function better than before. Again, trait discovery is active continuously during the lifetime of an organism and is not tied to specific life-history events. It shares this property with learning systems and thus it appears indeed to be important for cultural evolution.

Trait reservoir in Lamarckian evolution In Lamarck's concept of inheritance of acquired characteristics, the trait reservoir on which the trait-discovery process operates is associated with individual organisms and is continuously modified according to the needs of the organisms during their lifetime. This modification is therefore dependent on the performance of an organism's trait (compare Figure 2.4.2) and does not act blindly. Lamarck did, however, not provide a mechanistic explanation of how inheritance worked, given that experimental data was virtually absent during his time [Burkhardt, 2013]. In this regard, his theories were never fully fleshed out. A discussion of possible trait memories associated with Lamarckian modes of evolution is consequently postponed until we discuss inheritance in more detail in chapter 4.

All in all, it's plausible that a Lamarckian system, if augmented with a trait memory system will form a CCS, and given the importance of at least some aspects of Lamarck's reasoning, it remains an important body of theory with active research potential.

3.3 The Darwinian concept of evolution - Natural Selection

3.3.1 Principles of natural selection

Variation in traits leads to variation in interactions Starting from the network picture shown in Figure 3.1, empirical investigation shows that organisms vary significantly in their observable traits within a population [Stearns and Hoekstra [2005] and references therein]. This manifests in a distribution of traits over the population, ranging from body size, age, color, and many physical properties to behavioral traits like inherent curiosity or learned hunting behavior. This variation is larger in large populations than in smaller ones, because unlikely trait and trait combinations are more probable to be realized in the latter. Because these traits mediate the interaction with other entities in the ecosystem network, biological and abiological, the interactions they engage in will vary from individual to individual as well. This does pertain both to the intensity of interaction, e.g., the proportion a single prey species makes up in the diet of a predatory individual, and to the quality of interaction, i.e., with what an individual interacts at all and in what way. For example, an ape could use a plant both as food and nesting material or only for one of those.

Variation in reproduction and survival Natural organisms must acquire material and energy from the outside world in order to survive and reproduce. How they do this depends on their physiological and behavioral traits. For example, predatory individuals with longer legs or a higher shin-to-thigh ratio are generally able to run faster than others and thus could catch more or different prey. Similarly, a prey organism with some coat color matching environmental patterns better could blend in more than others and avoid predators better, and organisms with a more active immune system are more resilient against parasites and pathogens and more likely to survive an injury. Therefore, variation in traits of an organism leads to variation in interaction with its environment, i.e., other elements of the ecosystem, which, in turn, leads to variation in the means it has available to maintain itself, survive, and reproduce.

Note that this does not mean that every variation in interactions needs to result in variation in reproduction - different types of the same interaction or even different interactions with the environment may lead to very similar means for reproduction and self-maintenance. Therefore, not all differences in interactions are relevant for evolution [Stearns and Hoekstra, 2005]. In particular, interactions between the organism and ecosystem are not fully determined by the organism's traits but depend on a plethora of influences that are not part of the individual. Because these influences cannot all be modeled explicitly, ecosystem models always have a stochastic component, unless large, well-mixed populations are assumed that allow for the application of differential equations or the law of mass action. A classic example is provided by the Lotka-Volterra equations, which describe a two-species predator-prey ecosystem [Lotka, 1920; Volterra, 1927].

Natural selection arises from the correlation between the distribution of traits and the distribution of reproduction and survival in a population The logic of natural selection now builds on a correlation between variation in traits and variation in survival and reproductive success across generations. Given some variation in traits that leads to variation in ecological interactions, some organisms will be more successful in acquiring the means for reproduction and self-maintenance than others, i.e., they survive for longer or reproduce more often. These traits can be inherited by their offspring. If a similar pattern of ecological interaction holds for them, too, their survival and reproductive success will be biased in the same way, such that over time, organisms with advantageous traits will increase in number, while others will decrease and vanish. Thusly, iterative modification of existing structures and ongoing adaptation to changing external conditions is enabled.

Natural selection, therefore, relies on variation in organismal traits that mediate variation in ecological interaction. These manifest in variation in survival and reproductive success such that the two distributions are correlated [Stearns and Hoekstra, 2005].

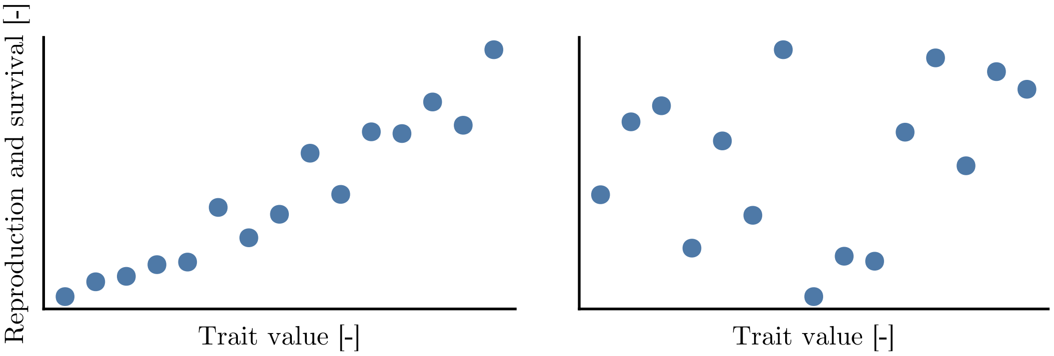


Figure 3.2: Natural selection arises through correlation of variation in organismal traits with variation in reproduction and survival (left). A lack of this correlation produces no natural selection (right). This alone does not entail causality, such that pseudo-selection can come about by spurious correlation, especially in small populations, due to stochastic effects.

It must be noted, however, that this is purely a statistical principle that does naturally not automatically imply a causal connection between variation in traits and variation in reproductive success or survival. Rather, these causal connections must be inferred separately, which can pose a significant challenge in natural populations, given that many ecological interactions influence reproduction and survival simultaneously in addition to environmental events.

3.3.2 Inheritance in the context natural selection

In nature, parents and offspring resemble each other, i.e., the traits of the offspring organism are similar to those of the parent organism (or parent organisms in the case of sexual reproduction). Remarkably, Darwin, as well as Lamarck, proposed their theories of evolution before the mechanisms of inheritance were known. Darwin, in particular, proposed the mechanism of pangenesis, in which heritable particles are produced from all the cells of a body and the trait of the offspring is a blend between the trait of the parents. Such a mechanism, however, leads to a loss of variation in traits over time (compare Figure 2.1, D) when reproduction uses recombination, as it does for many organisms in nature. Consequently, variation in interactions will be lost over time which also yields a loss of variation in reproduction and survival. Therefore, such a mechanism will, over time, preclude natural selection because variation in traits will be lost. Moreover, a rare, useful trait will likely be lost or diluted in the next generation by blending with another, less useful trait, which is especially problematic when the new trait is an innovation carried by a single organism. An inheritance mechanism based on blending inheritance can therefore not support natural selection.

In contrast, Mendel discovered that inheritance is particulate [Mendel, 1866], i.e., traits are inherited by offspring in discrete units. It can be shown that such an inheritance system can preserve variation if other influences are neglected [Hardy, 1908; Weinberg, 1908]. Individual units of inheritance (chromosomes, see chapter 4) that produce different variants of traits can be assorted in variable combinations in different organisms, thus producing new variations of phenotypes. In asexual populations that inherit their traits from only a single parent, the blending-recombination problem does not exist and thus inheritance can be seen as a more flexible process in such populations.

Inheritance is, however, a vital part of evolution by natural selection because it generates the trait memory required for its functioning as a CCS.

Inheritance is imperfect Inheritance is not perfect in natural populations. While the fidelity of inheritance is often high [Ridley [2004] and Stearns and Hoekstra [2005] and references therein], in a population of organisms subject to birth and death over multiple generations, variations in traits will invariably appear through inaccuracies in the inheritance of traits during reproduction. How variation is generated is discussed in chapter 4. Here, it is important to note that inheritance in nature not only generates the trait memory system but at the same time provides the core of the trait-discovery process, i.e., it introduces new traits.

While for the logic of natural selection to hold, it suffices, in principle, that variation exists, the constant generation of new variants enables it to go on without necessarily reaching a homogeneous state, which otherwise would be inevitable because of the continuous removal of variation. Therefore, evolution by Darwinian selection requires a source of new variation in order to continuously produce adaptations to changing environments. Inheritance and mutation are therefore fundamental ingredients to the Darwinian scheme of evolution.

While the term “mutation” is mostly applied to genetic changes, here we use it in a more general context that denotes all heritable changes in traits in an offspring organism in which it differs from its parent. As will be discussed more deeply in chapter 4 and chapter 5, the process that brings them about, however, can be far more complicated than the “copying errors” that are often cited, and how variation in traits is generated is a topic all of its own.

Inheritance and mutations belong together It must be noted that in the current concept of evolution by natural selection, mutations are generally bound to reproduction. In contrast to Lamarck, where variation is generated over the course of the lifetime of an organism and is inherited by offspring, it has been verified that while there are mechanisms for the inheritance of acquired characteristics, these are generally not stable over long periods of time, but last $\mathcal{O}(10)$ generations (compare subsection 4.3.1) in animals, while horizontal gene transfer in bacteria behaves differently and can create long-term stable innovations because of the different nature of DNA (compare chapter 4). Mutations that are heritable over longer times generally are located in the DNA of the organism, such that their occurrence is bound to reproductive events.

Moreover, Lamarck posits a correlation between the behavior of an organism and the traits that change, like in the famous giraffe example. Again, empirical investigations show this to generally not be the case [Stearns and Hoekstra, 2005; Mayr, 1972]. Rather, mutations are generally assumed to be independent of the effect they have on their carrier [Stearns and Hoekstra, 2005; Ridley, 2004; Mayr, 1972].

However, while there is no known way in which organisms can bring about an exact genetic mutation based on its effect on the phenotype, mutation *rate* as such is frequently tuned through intergenerational selection or during the organism’s lifetimes, which results in changed adaptability [Denamur and Matic, 2006]. This process, however, mostly occurs in bacteria.

All in all, not only is the trait-discovery process associated with natural selection (mostly) bound to reproduction events, but there is a trade-off between adaptability, i.e., the speed with which new traits can be discovered, and the maintenance of existing traits, i.e., heritability.

3.3.3 Properties of natural selection

The strength of natural selection depends on available variation and correlation strength Because of its stochastic nature, evolution by natural selection depends on the amount of variation of traits present in a population. Starting with a homogeneous population and assuming a constant probability of mutations of traits and constant probabilities of death and reproduction per individual, a larger population will, after some time, also exhibit a larger pool of variation. Moreover, rare mutations which might confer uniquely advantageous ways of interacting with the environment and allow for more effective acquisition of resources will be more likely to appear in a large population within a given timeframe. A similar effect appears if, everything else being equal, one population has a higher probability of reproduction and death than the other. Because mutation is generally bound to reproduction events, a higher overturning rate in the population also results in more mutation events and consequently will generate variants more quickly. Therefore, if everything else is held equal, larger populations or those with a shorter generation time will in general evolve more quickly [Stearns and Hoekstra, 2005] (compare Figure 3.3, D) than those with long generation times, smaller populations, or slower reproductive periods.

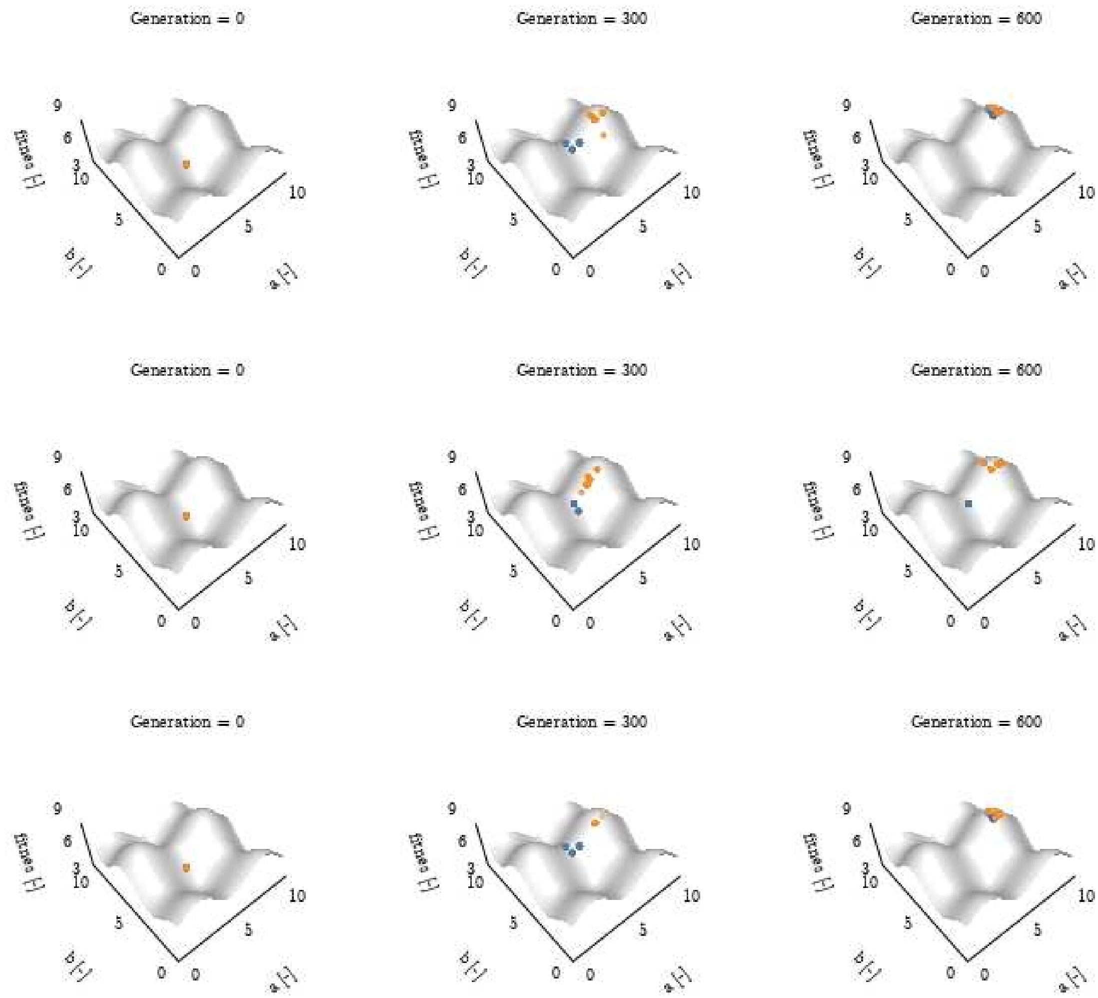
Secondly, a stronger correlation between the distribution of traits and the distribution of reproductive success and deaths per time will yield faster adaptive evolution than a weaker one, because the probability of an organism that carries deleterious traits being removed from the population is higher, or the probability of having offspring is lower. Therefore, the next generation will contain more offspring of organisms carrying advantageous traits when compared to a population under weaker selection (compare Figure 3.3, A).

Stron selection events are often associated with a change in environmental conditions and therefore a more or less sharp reduction of reproductive populationsize. This can have adverse effects in that it quickly reduces the available variation in a population and therefore reduce its future evolvability, i.e., create a bottleneck in the trait memory of the system. Additionally, a smaller population is more vulnerable to stochastic effects (see section 3.7), which additionally can reduce evolvability.

All in all, evolution by natural selection can be fast when there is a lot of variation in traits in a population that correlates strongly with survival and reproductive success [Stearns and Hoekstra, 2005]. It must be noted, however, that in nature, evolution is generally a mixture of many effects, acting on multiple traits at the same time, and with different strengths and interdependencies. Additionally, the way a phenotype is produced from heritable information can skew and change the effect of natural selection (compare chapter 4 and chapter 5).

Secondly, natural evolutionary processes are in many cases better modeled stochastically than deterministically because of the plethora of interactions natural organisms engage in with their surroundings and because mutations are singular events that occur in individual organisms. Therefore, random events, like the occurrence of a specific mutation at a specific time, can significantly influence individual realizations of any natural selection process.

Natural selection removes trait variation from the population Biased survival and reproduction that natural selection leads to a larger proportion of the next generation


Generation = 0

Generation = 300

Generation = 600

Generation = 0

Generation = 300

Generation = 600

Generation = 0

Generation = 300

Generation = 600

Figure 3.3: The effect of natural selection strength in an agent-based model of natural selection. Each agent has two traits $a \in [0, 10] \subset \mathbb{R}$ and $b \in [0, 10] \subset \mathbb{R}$. Per timestep, five percent of the population randomly dies and gets replaced unless stated otherwise. The probability of death and reproduction is proportional to the value of the function $f(a, b) = 5 + \sin(a\,\pi\;3) + \cos(b\,\pi\;3)$ at the point given by an organisms parameters. The population is homogeneously initialized with $(a, b) = (4.5, 4.5)$ for all individuals, and populationsize is held constant at 1000 unless stated otherwise. The function f is shown as a grey surface. An organism mutates by randomly adding a number $n \in [0, 1]$ to either trait a or b, with the mutation rate being fixed at 0.05. **Upper row**: Higher rate of overturning ($p_{death} = 0.15$, orange) vs lower rate ($p_{death} = 0.05$, blue). This yields more reproduction events per timestep, consequently more mutations and faster explorations of state space, which translates to faster evolution. **Middle row**: Two populations of size 200 (blue) and 2000 (orange): A larger populationsize yields more reproduction events per time, which leads to proportionally more mutations and consequently faster evolution by natural selection. This effect is roughly equivalent to a higher overturn in the population. **Lower row**: For the orange population, the values of f are squared before death and reproduction, leading to a larger difference in probabilities for being selected or removal and reproduction. Advantageous mutations consequently spread faster through the population, leading to faster evolution.

being made up of offspring of comparatively few individuals, as mentioned before. Because of inheritance, this means that more individuals in the population will carry similar traits. Natural selection, therefore, reduces variation in the population over time (compare Figure 3.3, A). Very strong selection consequently can leave the population with an impoverished trait pool and thus potentially vulnerable to future changes in the ecosystem and to stochastic effects. The main effect is that large portions of the system's memory are effectively erased. Note that, as a first approximation, because the process of natural selection hinges on reproduction, only the organisms which reproduce are relevant, i.e., a population of size N in which only $n \ll N$ organisms reproduce behaves effectively the same as a population of size n in which every organism reproduces.

The interplay between selection and mutation gives rise to adaptation Natural selection and mutation can together create a ratchet effect in which advantageous mutations spread through a population via biased survival and reproduction, before another, more advantageous mutation appears which then spreads and supplants the first one, and so on. This is contingent on heritability being high enough to allow for the preservation of advantageous traits and on the mutation rate being so low that the majority of the organisms in the population are replaced before the next advantageous mutation appears. We can also imagine this scheme to hold in the presence of higher mutation rates, even when advantageous mutations are rare. Because the trait will eventually be spread through large parts of the population, it is unlikely to get lost again by chance. After a few rounds of this ratchet effect, the population can be comprised of organisms with traits that would be exceedingly unlikely to appear at random. This scheme is the base assumption of adaptive dynamics, which is an important line of mathematical modeling in contemporary evolutionary theory [Brännström et al., 2013].

Where this ratcheting process acts on complex (adaptive) individuals, it is easy to imagine how it can result in the complex functions of natural biological organisms. At the same time, the process also can act on much simpler systems, and is, in that sense, general enough to apply to all imperfectly reproducing systems [Lewontin, 1970] (and possibly beyond, see Papale [2020] and references therein).

Mutation-selection balance This tendency to remove variation from a population is balanced by mutation, and the actual distribution of a trait in a population consequently depends on the balance of these two processes. This leads to the idea of mutation-selection-balance, which is especially important for asexual species because, without recombination, mutation is the only process that introduces new variants into the population. The balance between mutation and selection, therefore, limits adaptation in the system, because any trait or combination thereof that yields advantageous mutations will, with a certain probability, undergo mutation upon reproduction. Because there are typically many ways in which a trait can mutate, there is a disequilibrium between mutations "towards" focal trait and all the possibilities for the latter to mutate into some other trait. Consequently, an advantageous trait, once lost, is highly unlikely to be regained through mutation in the heritable information.

It must be noted that while this concept holds in general, developmental processes (see chapter 5) will modify this phenomenon. This mutation-selection balancing process also constitutes the basis for Muller's ratchet [Muller, 1964], which states that asexual populations cannot get rid of (slightly) deleterious mutations, and thus a population state dominated by highly adapted individuals will always deteriorate to some degree until the mutation-selection balance is attained again. Thus, asexual populations will have a limit on how well-adapted they can become.

While this, in general, does not pose a big problem in large, well-mixed populations, for small populations, stochastic effects are more pronounced. This leads to the effect of mutational meltdown [Gabriel et al., 1993], in which adaptations are consecutively lost by mutation and not regained due to the stronger influence of stochasticity and weaker selection (compare Figure 3.4, Figure 3.3 and section 3.7). In nature, this can result in a further decline in populationsize, increasing the meltdown effect. This can, via positive feedback, eventually lead to the extinction of the population.

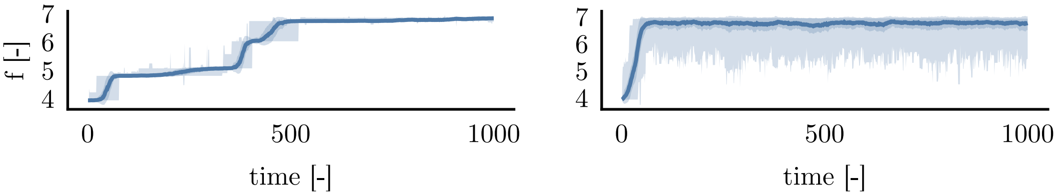


Figure 3.4: Simulation of mutation-selection balance with the computer model of natural selection introduced in Figure 3.3, approaching the same peak. Mutation rates are 5×10^{-3} (left) and 5×10^{-1} (right). Shown are the mean value of f (solid line) and the respective standard deviation (shaded area). We find the lower mutation rate to evolve slower and more in line with the assumptions of adaptive dynamics, while for the higher mutation rate, advantageous mutations occur more or less concurrently. The lower mutation rate produces a higher mean value for f with much lower variance than the higher mutation rate, although both maintain the adaptive peak.

While well appreciated for asexual populations [Lynch, Bürger, et al., 1993], mutational meltdown can play a role in small sexual populations as well under some conditions [Lynch, Conery, et al., 1995], a fact that can play a role in conservation biology [Hedrick and Kalinowski, 2000].

In general, the higher the mutation rate, the lower heritability and the more likely it is for advantageous traits to get lost again over generations. Consequently, life has found ways to control or reduce mutation rates, see chapter 4, or to buffer against variability.

Sexual reproduction can overcome mutation-selection balance Sexually reproducing organisms, on the other hand, recombine their traits upon reproduction, such that, independently from the overall mutation rate, advantageous traits can combine within the same organism without them having to arise sequentially within a lineage of asexual organisms. Likewise, deleterious traits can be removed more efficiently. Because recombination produces variation in the assortment of traits in the population and, in the absence of other effects, can also maintain genetic variation in a population (this effect is called the Hardy-Weinberg law, [Hardy, 1908; Weinberg, 1908]) this allows for the evolution and maintenance of more adaptive states than could be maintained in an asexual population, compare Figure 3.5.

The price is the fact that two individuals are necessary for reproduction, which might not be available at all times. Moreover, sexual reproduction entails that only one sex is actually producing offspring, and only half of its offspring will do so in turn. By comparison, any

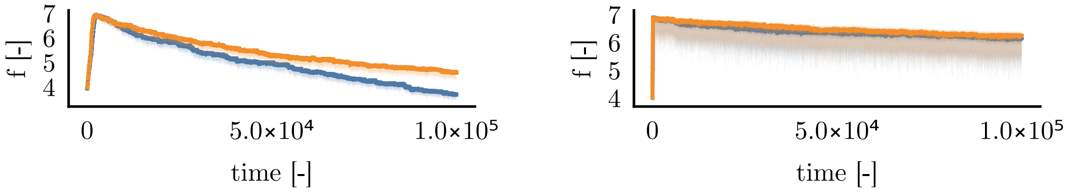


Figure 3.5: The model from Figure 3.3 and Figure 3.4, with a mutation rate of 0.3 and a population of 1000 organisms. Blue: asexual population. Yellow: Sexually reproducing population, where each parent contributes trait a or b with a fifty percent chance to the offspring. **Left**: Mutation width per mutation event is $n \in [0, 0.1]$, i.e., mutations have small effects. **Right**: Mutation width per mutation event is $n \in [0, 0.5]$, i.e., mutations have larger effects and a single step away from a peak in f can produce significant loss of survival and reproduction. We find that mutational meltdown affects asexual populations more as their mean fitness (solid blue) decreases stronger the longer the simulation runs, although both populations adapt quickly in the beginning. The effect is lessened when mutations have stronger effects, as can be seen in the lower panel.

offspring of an asexual organism can reproduce again, such that, simply by numbers, an asexual population should outcompete a sexual population, at least in the early stages of their interaction. Because of the twofold cost of sex in natural evolution [Maynard Smith, 1978; Gibson et al., 2017], the advantages recombination of traits brings in terms of adaptiveness must outweigh its cost in order for it to be maintained, as explored a little in Figure 3.5, and multiple models for its maintenance and origin have been put forward and have been investigated empirically (e.g., Maraun et al. [2019]).

Inheritance records the history of past selection Because the continued existence of a trait in the population is bound to inheritance, only those traits that at least do not hamper reproductive success too much will be left in the population after some time. In other words, only those traits will be left in the population which have been advantageous, or at least not too deleterious, in the past. The distribution of traits in the population is, therefore, the product of a (possibly) long history of selection over many generations [Stearns and Hoekstra, 2005]. Consequently, organisms are a product of different past selection regimes. This, in turn, applies to the trait memory of the evolving population, too, in line with the concept of a CCS which requires continuous modification of elements in some memory structure.

3.3.4 On the nature of evolution by natural selection

Evolutionary relativity It's worth emphasizing that selection can only act on the variation that is present in the current population because organisms can only interact with the ecosystem as it currently – or over the time interval of interest – exists. Therefore, differential survival and reproduction will lead to an increase in some traits if they are advantageous relative to the others it interacts with.

In other words, there is no absolute sense of optimality towards which a Darwinian system develops. Rather, while it is capable of rapidly producing unlikely states, it will only reach local optima, i.e., reach states in which the population is dominated by a collection of traits for which any (or at least most) alternatives will be worse with respect to survival and reproduction. In the language of evolutionary game theory, such a configuration is called an evolutionary stable strategy [Nowak, 2006]. It must be noted, however, that the notion of an evolutionary stable strategy is mostly bound to a static ecosystem with respect to which a population can evolve [Grunert et al., 2021]. Thus, as a theoretical concept, it is mostly applicable to microevolution. In natural systems, the fact that organisms and many abiotic systems are CAS and thus are open systems, as well as possible evolutionary innovations that lead to a reorganization of local interactions, can change the adaptive value of certain traits over time, such that what once was a well-adapted set of traits can be maladaptive later.

Therefore, if we frame Darwinian evolution as the process of finding configurations of traits that are effective at survival and reproduction for a given ecosystem, we find that Darwinian evolution does not produce optimal solutions in an engineering sense. Rather, it produces configurations over time that are more effective than what already exists within the set of interacting entities relative to the system's state at any given time. Therefore, there is no natural sense of "progress" in the concept of natural selection as such.

3.3.5 Evolutionary continuity, exaptations and restrictions

Following the train of thought started in subsection 3.3.4 further, we find that not only does Darwinian evolution produce only local adaptations, but the evolutionary trajectory of a population can only pass through reproductively viable states. Thus, a Darwinian evolution trajectory that links two trait configurations via a lineage, e.g., from aquatic to terrestrial habitats or from terrestrial locomotion to flight, must occur by successive changes to configurations that can effectively survive and reproduce. This process can be likened to rebuilding a car into a submarine while the car is going at eighty miles per hour. As one can imagine, the likely outcome of such an operation will be suboptimal when compared to some from-the-ground-up engineering solution, involving structural or functional compromises and generally higher complexity than would be deemed necessary, or prudent, by a human designer. For example, birds evolved from small, bipedal, theropod dinosaurs, which were dependent on their running and walking ability to survive and successfully produce offspring. Consequently, their legs must have kept this function while they were slowly adapting to gliders and eventually fliers. The consequence is that birds use their legs essentially for providing a jump start when starting to fly, but otherwise just carry them around, their use as grasping tools notwithstanding. Bats and pterosaurs, starting from a quadrupedal initial condition, use all four limbs for starting, which allows for more efficient morphology. A similar argument can be made for feathers, which are rather heavy when compared to the flight membrane of bats or pterosaurs. All these compromises result in an arguably somewhat awkward "design" of birds for flight when compared to other animals or human systems.

Moreover, a number of structures will be identifiable as having originally arisen in some function that they no longer fulfill, having been repurposed in another context. Some of these traits may have not served a purpose at all in the past, but are now fulfilling vital roles within the physiology of the organism. These repurposed properties of an organism are called "exaptations" [Gould and Vrba, 1982]. Flight feathers are an example of this phenomenon because their precursor structures probably arose as insulation but now play a role in flight and mate finding.

Other traits may have been originally useful but have lost their function in a new ecological context without acquiring a new one, and are just carried along in the population without doing any harm (or at least not much), acquiring mutations over time and losing the structure that allowed them to function in the original capacity. Such traits are often characterized by comparatively high variation within the population because no selection acts on them that will remove some variants over time. Some traits may be vital for reproduction and survival, and so any newly arising trait will evolve "around" them, adding a new function or changing an existing one without changing such keystone traits. The example of bird's legs shows this phenomenon. This leads to the entrenchment of features across evolutionary lineages, such as the general layout of the body plan (compare chapter 5 for a more detailed discussion).

The continuity condition that arises as a corollary from the reproduction-inheritance-based progression of natural selection, therefore, can lead to restrictions of what can evolve at any given time, can force evolution into objectively suboptimal paths, and create a higher complexity than would be strictly necessary as judged by an intelligent engineer.

Natural organisms are therefore a mosaic of traits that have arisen in a given context but that can have been repurposed towards a different function, can have continued in their original function, or can have lost it over time. This paragraph, therefore, supplements the idea of evolutionary relativity presented in subsection 3.3.4, because all the phenomena discussed here would not be possible if there was some global optimization going on in evolution.

3.3.6 The primacy of reproductive success over survival and the concept of fitness

We mentioned that natural selection is based on a correlation between variation in traits in a population and variation in survival and reproductive success. These latter two, however, are not symmetric. A trait that confers better survival probability per time but does hinder reproduction will rise slower in frequency in the population than another trait that does reduce average survival time but confers a high degree of reproductive success within that time frame. While reproductive success is both a necessary and sufficient condition for the continued existence of a lineage of organisms that carry a specific trait, survival is just a necessary one. Indeed, in nature, the life cycle of many organisms is such that it prioritizes reproduction over individual survival, e.g., males of some species of spiders get often eaten by their mate, salmon wander up a river to lay eggs only to die immediately after that and many more organisms behave similarly [Stearns and Hoekstra, 2005]. We can therefore conclude that the relevant variable for evolution by natural selection as presented herein is lifetime reproductive success, not survival in itself.

Closely related to this discussion are the concepts of "fitness" and "survival of the fittest". Fitness is readily invoked in discussions of evolution, although it is surprisingly difficult to come up with a consistent definition [Ariew and Lewontin, 2004; Doebeli et al., 2017; Barker, 2009]. For now, we use the term merely to refer to lifetime reproductive success, with more differentiated meanings being introduced as needed.

3.3.7 The evolutionary tree as a consequence of Darwinian evolution

Consider a population of imperfectly reproducing entities with a mutation rate that is low enough to allow for adaptive dynamics. Furthermore, we consider asexually reproducing populations for simplicity. Because mutation is tied to reproduction events, the offspring of a mutant individual M carrying a trait A' will likely carry the mutated trait as well. At the same time, any non-mutated siblings of M carrying trait A will also have offspring that probably will lack the mutation. Therefore, the two lineages exist side by side. We can see this as a branching event in the history of the population, where a single lineage of A individuals branches into two lineages - one carrying A and the other carrying A'. New branches may persist for long times, giving rise to their own daughter branches, or disappear again such that only the original remains. Because the differential survival and reproduction of such lineages depend on the ecosystem they interact with, these branching events can be brought about by differences in ecology, e.g., by spatial isolation. In sexually reproducing populations, reproductive isolation, i.e., the inability to produce fertile offspring (or offspring at all) is a deciding factor for a lineage split usually associated with the advent of a different species (see Figure 3.1, however). We note in passing that reproductive isolation does not necessarily entail some kind of physiological incompatibility, but can also just be caused by spatial isolation and other effects, as is the case for polar bears [Kelly et al., 2010].

The branching effect of course is a consequence of the inheritance-based trait-discovery process inherent to Darwinian evolution, and would not be present in a Lamarckian framework. Over long timescales, this effect gives rise to the Tree of Life.

3.4 Natural selection in the context of CCS

3.4.1 Performance evaluation

Because evolution by natural selection is associated with survival and reproduction, performance evaluation as defined in subsection 2.4.2 is straightforwardly identified as differential survival and reproduction. Importantly, because reproduction is associated with individuals, and not with *individual traits*, this assigns a performance value (lifetime reproductive success) to individuals and not, a priori, to individual traits. Therefore, the individual's performance is the result of the interplay of its traits over its lifetime. While this can be dominated by single traits (e.g., some genetic defects that preclude survival to reproductive age), this is not the general case. Rather, the survival and reproductive success of an organism is generally the result of the interaction, and cooperation, of the traits of an organism. Secondly, because lifetime reproductive success is the result of the interactions an organism engages in over its lifetime, and these interactions involve other organisms of the focal or other populations and dynamic abiotic factors, this form of the performance evaluation process is strictly an emergent process in a population of reproducing entities and has no analogon on the level of individual dynamics unless the individual is a conglomerate of reproducing entities in itself.

3.4.2 Trait discovery and trait reservoir

Trait discovery Trait discovery can be identified with mutation and recombination. As mentioned, the mutation part is understood to be independent of the traits or actions of the individual, in contrast to Lamarckian evolution. Thus, mutation is normally treated as a background process in which offspring randomly differ, with some probability, from their parents. The situation is more complicated for sexual reproduction, however, because a mate is required, which often involves active choice and assessment. This latter phenomenon gives rise to sexual selection [Futuyma and Kirkpatrick, 2007], which can drive the morphological evolution of organisms to a significant degree. In all cases, however, trait discovery is associated with reproduction events, again in contrast to Lamarckian evolution, where trait discovery was spread out over the entire lifetime of an organism. It must be noted that while mutation is strictly associated with individual organisms, sexual recombination can only take place in a population, and so many natural populations that undergo Darwinian natural selection rely on a mixture of trait discovery processes that depend on different underlying mechanisms.

Trait reservoir Because the system relies on the imperfect reproduction of individuals for trait discovery, individuals also act as the loci for the trait reservoir of the system. The dynamics of this trait reservoir is largely determined by the presence or absence of recombination, which can produce new combinations of traits more rapidly than asexual, i.e., sequential within-lineage, evolution could do. Mutation selection-balance is another determining process, which prevents a population from collapsing onto a single trait configuration through natural selection. Recombination consequently explicitly relies on the distributed nature of the trait reservoir and thus mixes traits across a population, while mutation only is concerned with the traits of a single individual at the time of reproduction. In both cases, however, evolution by natural selection is dependent on the existence of a population because it is bound to reproduction, and inheritance of acquired characteristics is largely excluded. Populations evolve, but individuals do not. Moreover, the size of the population and the variation of traits across its individuals are important determining factors of how active natural selection can be in a population. Therefore, natural selection relies explicitly on the size of the trait reservoir. Finally, it's worth reiterating that the trait reservoir of a Darwinian population reflects the history of past selection because only those traits that allow for reproduction can remain in the population over time.

3.5 Modeling evolutionary processes on a Darwinian basis

3.5.1 Birth-Death processes

We follow Doebeli et al. [2017] in proposing stochastic birth-death processes as a fundamental modeling tool for biological evolutionary processes because offspring creation and death are the observable processes in nature that give rise to evolution. Thus, they are the natural choice when attempting to model biological populations. This naturally ties in with the discussion in subsection 2.2.2, where more general stochastic processes have been discussed.

The system to be modeled can be imagined as an ecosystem as visualized in Figure 3.1 that changes its structure and composition over time. Thus, the evolution of the system can be seen as a sequence of such networks made up of nodes that change their links and states over time, die, and are born.

Mathematical representation of Birth-Death processes A birth-death process is a Markov process that operates on a population of size $N(t)$ in which each member of the population is determined by some state a_i at time t. The state a_i is extremely general and can denote and collection of traits of an agent, from complex properties with internal dynamics like its genetic system, proteome or metabolic functions to some simple scalar, like age or body size. This also means that in general, a_i is time-dependent, $a_i(t)$. We encountered this already in subsection 2.2.2 where the Master equation was introduced to treat such systems.

Rate equations For continuous time, we define functions $b(a)$ and $d(a)$ that determine the birth rate and death rate of an individual given its state a, respectively. Additionally, these depend, in general, on the state of the biotic and abiotic environment $e_{bio}(t), e_{abio}(t)$ the focal population is embedded into.

$$\begin{aligned} b_i(t) &= b(a_i(t), e_{bio}(t), e_{abio}(t)) \\ d_i(t) &= d(a_i(t), e_{bio}(t), e_{abio}(t)) \end{aligned} \tag{3.5.1}$$

Noteworthily the biotic environment e_bio includes effects that stem from the current population, but may also include influences from other populations which are not explicitly modeled, or any other biotic influence that comes about through ecological interactions, e.g., competition for resources, symbiosis or parasitism. The variablee_{abio} models the abiotic state of the environment, e.g., temperature or salinity, but can also include climate or geology if the respective timescales are relevant for the question at hand. The functions in Equation 3.5.1 consequently are a model for how lifetime reproductive success arises from ecological interactions.

In contrast to Doebeli et al. [2017], we include an explicit time dependence in all independent variables a, e_{bio}, e_{abio}, because the state of any given agent is in general subject to phenotypic plasticity or development (compare chapter 5), while the environmental variables have their own dynamics as well, e.g., seasonality. Moreover, a_i incorporates the heritable traits of the agent, and in this way the "memory" of the system. In this way, the system only depends on the current state of the population and environment and therefore constitutes a continuous-time Markov process. That being said, any representation of a real system through Equation 3.5.1 would, of course, be exceedingly complex because it contains all influences on birth- and death-rate of a type a_i. Moreover, the lack of time homogeneity makes analytical treatments of such processes exceedingly complicated.

We note that Equation 3.5.1 only implements perfect reproduction, i.e., there is nothing that can generate variation in states in the system beyond what was initially present. As pointed out in Doebeli et al. [2017], this renders the model purely ecological. Such a system can be useful under the assumption that ecological and evolutionary timescales are decoupled, for instance when dealing with ecosystems on the species level. With respect to Figure 3.1, this means that there is a fixed set of organism-, environment- and interaction classes given by the ones currently represented in the system, and this does not change over time.

To make Equation 3.5.1 into a proper evolution model, we need to add a representation of imperfect inheritance, i.e., an equation for the rate of generation of new variants. In Doebeli et al. [2017], this takes the form

$$c_i(t) = c(a_i, e_{bio}(t), e_{abio}(t)) \tag{3.5.2}$$

If Equation 3.5.2 is zero always, we recover Equation 3.5.1 and thus a purely ecological system. The function $ci(t)$ represents the generators of variation in the system and therefore includes, e.g., mutation, but also ontogenesis, plasticity, mate choice, recombination, and other effects, as discussed in subsection 3.3.2. Depending on the structure of the states a_i, it can make sense to conceptually separate the parts of c_i that pertain to inheritance, e.g., mutation or recombination and those that pertain to the physical realizations, like ontogenesis. While both of these process classes can generate novel variants of a_i, their underlying concepts are radically different (see chapter 4 and chapter 5).

For a system unfolding in a finite, countable state space with agent states $a_i = \{a_1, a_2, \dots a_M\}$, which each have a population $N_i(\tau)$, the state of the system at time τ can be described as a vector

$$\mathbf{N}(\tau) = (N_1(\tau), N_2(\tau), \dots N_M(\tau)) \tag{3.5.3}$$

and transitions can be described as

$$\mathbf{N}(\tau) = (N_1, N_2, \dots, N_i, \dots, N_j, \dots, N_M) \rightarrow \tag{3.5.4}$$
$$\mathbf{N}(\tau') = \left(N_1, N_2, \dots, \tilde{N}_i \neq N_i, \dots, \dots, \tilde{N}_i \neq N_j, N_M\right) \tag{3.5.5}$$

The time development of the probability P_{ik} that the system transitions from state i at time s to state j at time t can be described by the Master equation introduced in Equation 2.2.1.

The elements of the transition-rate matrix from one population state to another, A, therefore are a combination of the functions b, d, c, which together represent an ecological system in which organisms produce offspring with modification (via mutation, recombination or ontogenetic influences), i.e., biological evolution in a general form.

It must be emphasized that the level of representation is a modeling choice: We here ignored, e.g., spatial movement, and subsumed all environmental influences into the arguments e_{bio}, e_{abio}.

Formulation of a generic evolutionary birth-death process Note that all natural populations are ultimately finite, though the carrying capacity, i.e., the system size, does not need to be fixed or even be knowable a priori [Erwin, 2008]. Thus, while we can be certain there is a finite maximum populationsize $N_{i,max}$ in any given time interval, its exact value is ultimately a modeling choice. Similarly, the finiteness of the state space $|S| = M$ is a modeling assumption. While we can find some representation of a population that accommodates this assumption by classifying the individuals according to appropriate criteria, the nature of an evolutionary process does not allow us to predict which configurations will arise over time, limiting the representation presented here to situations where such innovation can be ignored. Also, all such classifications that discretize state spaces are ultimately choices that need not be unique and influence the outcome of the analysis, e.g., species.

Finally, we introduce “forbidden” transitions whenever the population for state a_i reaches zero:

$$\mathbf{N}(\tau) = (N_1, N_2, \dots, N_i = 0, \dots, N_M) \not\rightarrow \mathbf{N}(\tau') = (N_1, N_2, \dots, N_i \neq 0, \dots, N_M) \tag{3.5.6}$$

For the system described, excluding forbidden transitions, we find the infinitesimal generator A to have the structure

$$\begin{aligned} \text{nothing: } & A_{ii}(\tau) = 1 - d_i(\tau) - b_i(\tau) \\ \text{death: } & A_{ij}(\tau) = d_i(\tau) \\ \text{faithful inheritance: } & A_{ij}(\tau) = b_i(\tau)\left[1 - (c_i)\right] \\ \text{imperfect inheritance: } & A_{ij}(\tau) = b_i(\tau)(c_i) \end{aligned} \tag{3.5.7}$$

where A_{ij} is the local rate with which the process transitions from state i to state j and the rate of faithful inheritance arises from the requirement of the sum of "faithful inheritance" and "imperfect inheritance" $\overset{!}{=} b_i(\tau)$ because mutation is bound to reproduction. For all states that have forbidden transitions, we also have

$$A_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ else} \end{cases} \tag{3.5.8}$$

Note that all the complexity of the system is subsumed into the rate functions b, d, c. Finally, we explicitly assumed that mutations do not depend on the state the system transitions to, only the current one. This recovers the fundamental assumption that evolution by natural selection is "blind", i.e., it cannot predict the future, only "remember the past" by virtue of heritable information. Modifications to this system would be necessary in order to accommodate cultural evolution or any Lamarckian process.

3.6 A simplistic birth-death process

3.6.1 Description of a simple process dynamics

We use the formal model described above and simplify it down to an analytically tractable system. In doing so, we follow Nowak [2006]. The process is described by a single state variable $i \in 0, .., N \subset \mathbb{N}_0$ and unfolds in discrete generations, i.e., it is a discrete Markov chain. Therefore, we have only one kind of individual in the system, we hence consider, in ecological terms, a population of a single species. With this, Equation 2.2.1 becomes

$$P_i(t+1) = \sum_{j=0}^{N} A_{ji} P_j(t) \tag{3.6.1}$$

In each step, the process goes from a state i to either $i+1$ with probability β_i, which is the birth probability or to a state $i-1$ with probability δ_i representing the death probability. It can also stay the same with probability $1 - \beta_i - \delta_i$. We assume the states 0 and N to be absorbing states, i.e., $\delta_0 = \beta_0 = \delta_N = \beta_N = 0$. We assume asexual reproduction and perfect inheritance, i.e., in the sense of subsection 3.5.1 we have a purely ecological model that assumes that mutations are much rarer than reproduction events. This is sufficient because, for this section, we are only interested in selection. This process can be described by the generator

$$\begin{aligned} A_{0k} = A_{k0} &= \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases} \\ A_{Nk} = A_{kN} &= \begin{cases} 1 & \text{if } k = N \\ 0 & \text{otherwise} \end{cases} \\ A_{ij} &= \begin{cases} \beta_i & \text{if } j = i+1 \\ \delta_i & \text{if } j = i-1 \\ 1 - \beta_i - \delta_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{3.6.2}$$

where the first two equations describe the absorbing states and the last describes the possible transitions in between. We are interested in the probability of reaching state N when starting from state i, and call this probability p_i. The goal is to derive a general expression for this probability in terms of the birth and death probabilities β_i, δ_i. Following the derivation in Nowak [2006], we introduce $\gamma_i = \delta_i/\beta_i$, given that $\beta_i \neq 0 \ \forall \ i$, and $y_i = p_i - p_{i-1}, i \in [1, N] \subset \mathbb{N}$. Using that $p_i = \delta_i p_i + (1 - \beta_i - \delta_i)\, p_i + \beta_i p_{i+1}$ from Equation 3.6.2, and excluding $i = 0, i = N$ because then $p_0 = 0, p_N = 1$, we find that $y_{i+1} = \gamma_i y_i$. This yields

$$y_n = \prod_{k=1}^{n-1} \gamma_k p_1 \tag{3.6.3}$$

if we exclude the absorbing states $0, N$ again. Using $\sum_{k=1}^{i} y_k = \sum_{k=1}^{i} (p_k - p_{k-1}) = p_i$ and consequently $\sum_{k=1}^{N} y_k = p_N - p_0 = 1$ we find

$$\sum_{j=1}^{i} y_j = p_i = \left(\sum_{j=2}^{i} \prod_{k=1}^{j-1} \gamma_k + 1 \right) p_1 \tag{3.6.4}$$

and therefore

$$\begin{aligned} \sum_{j=1}^{N} y_j &= \left(\sum_{j=2}^{N} \prod_{k=1}^{j-1} \gamma_k + 1 \right) p_1 \\ \Leftrightarrow p_1 &= \frac{1}{1 + \sum_{j=2}^{N} \prod_{k=1}^{j-1} \gamma_k} \end{aligned} \tag{3.6.5}$$

Finally, we are able to express the probability to reach state N from state i as a function of the birth and death probabilities only, given $\beta_i > 0 \ \forall$ states $i \in 1, \ldots, N$.

$$p_i = \frac{1 + \sum_{j=2}^{N} \prod_{k=1}^{j-1} \gamma_k}{1 + \sum_{j=2}^{N} \prod_{k=1}^{j-1} \gamma_k} \tag{3.6.6}$$

Note that we made only very general assumptions on the birth and death rates β_i, δ_i, so Equation 3.6.6 holds for any scenario of selection, including neutral drift [Nowak, 2006].

3.6.2 Adaptive dynamics on a simple birth-death process

The question we try to answer here is, given the system presented above, under what conditions a single mutant organism can give rise to a lineage that takes over the entire population at the expense of the currently most common type, i.e., if it can "invade" the population. Similar questions are posed in evolutionary game theory, often with respect to the evolution of cooperation [Nowak, 2006]. Again following Nowak [2006] and starting out from Equation 3.6.6 we consider a population of one individual of type A in a population of $N - 1$ individuals of type B. The probability that A gives rise to a lineage that takes over the population is called the fixation probability and is denoted by p_A. The reverse case, i.e., the probability that one B individual eventually takes over a population of $N - 1$ A individuals, is denoted by p_B. Therefore, with N_A being taken as the state variable, we find

$$\begin{aligned} p_A = p_1 &= \frac{1}{1 + \sum_{j=2}^{N} \prod_{k=1}^{j-1} \gamma_k} \\ p_B = 1 - p_{N-1} &= \frac{\prod_{k=1}^{N-1} \gamma_k}{1 + \sum_{j=2}^{N} \prod_{k=1}^{j-1} \gamma_k} \end{aligned} \tag{3.6.7}$$

which gives us the ratio

$$\frac{p_B}{p_A} = \prod_{k=1}^{N-1} \gamma_k = \prod_{k=1}^{N-1} \frac{\delta_k}{\beta_k} \tag{3.6.8}$$

Therefore, which variant will take over the population at some point depends on the product of the ratios of the birth- and death rates in all non-absorbing states of the process, i.e., on the selective regime for the possible states of the system. Referring back to subsection 3.5.1, this may be a complicated function that depends on the current state of the biotic and abiotic environment. Equation 3.6.8 therefore describes intraspecific evolution.

3.7 Neutral evolution - Evolutionary change without selection

Neutral evolution can arise in multiple ways, e.g., variation in traits does not always lead to variation in fitness, for instance, because the affected trait does not significantly influence the interactions with the current environment. Furthermore, not all variation in traits is equally heritable. In nature, phenotypic variation has a genetic component, which is highly heritable, but also has other components (environmental, epigenetic, and others), which are generally less heritable (see chapter 4). In Equation 3.5.2 this is subsumed into the rates of mutation, although in reality, heritability of a trait is a complicated process.

Evolution without selection The question, therefore, arises if non-heritable variation does influence the evolutionary trajectory of a population. Neutral evolution requires that there is variation in traits which does not affect fitness. Because interaction with the environment determines variation in fitness, which phenotypic traits are neutral and which are not is determined by the ecosystem an organism is a part of, and migration to a different one or a change in ecosystem structure may change the selective regime a trait is subject to. The proportion of the population of organisms that share a certain neutral trait is subject to a random walk, influenced by the particularities of trait transmission. A neutral trait that is shared by most of the population is said to be “fixed”. A simple model to compute the probability of a trait becoming fixed in a population of finite size is the Moran process [Moran, 1958], which is based on the description given in subsection 3.6.2. In the discussion below we once more follow Nowak [2006]. We again consider two types of organisms A and B with populations N_A and N_B with $N = N_A + N_B =$ const as before, with absorbing states $N_A = N, N_B = N$, i.e., mutation is once more excluded.

The significance of the fixation of a trait, i.e., reaching the absorbing states, is that it “deletes evolutionary history”, i.e., any further evolution will invariably start from organisms with trait A, or B respectively, and evolutionary history can only be traced back to the point where this fixation occurs, at least for the one trait under consideration here.

The case where a new variant occurs in a homogeneous population, i.e., $N_A = 1, N_B = N - 1$, is of special interest because mutations that introduce new variants are bound to individuals, and thus start always with very low population.

Using Equation 3.6.7, Nowak [2006] goes on to show that if type A has fitness r and type B has fitness 1 (i.e., we discuss relative fitness), the probability of a single mutant of type A taking over a population of size $N - 1$ of type B is

$$p_A = \frac{1 - 1/r}{1 - 1/r^N} \tag{3.7.1}$$

which under the assumption of $N >> 1, r > 1$ collapses to $p_A \approx 1 - 1/r$.

We understand from this that even in very large populations, there is no guarantee that an advantageous mutation takes over, a fact that arises from the stochasticity of the system.

Finally, Nowak [2006] derives the rate of neutral overturn of the population, i.e., the rate at which a population of N organisms transitions from homogeneously exhibiting trait B to homogeneously exhibiting trait A by introducing a mutation rate μ, which gives us the rate at which mutants of the respective other type are introduced. With $p_A = p$, this yields

$$R = N\mu p \mid p = 1/N \text{ for neutral evolution} \tag{3.7.2}$$
$$\Rightarrow R = \mu \tag{3.7.3}$$

Therefore, the rate at which neutral traits fixate in a population over time is only dependent on the mutation rate, and independent of the populationsize. Important to note is the assumption of perfect neutrality here, which for most natural cases is arguably an idealization.

It's noteworthy that the process of neutral evolution is, from an observational point of view, not distinguishable from directional selection if only the trait distribution over the population is considered. Only a consideration of the fitness distribution over the traits present in the population can reveal if selection is active or not. Neutral evolution has been proposed by Kimura [1968] and Kimura [1983] to be the dominant mode of evolutionary change on a molecular level, i.e., when considering substitutions in amino acids in proteins or base pairs in the DNA. Here, redundancy is relatively high (see section 4.2), and therefore, the system exhibits much potential for neutral evolution. This, however, does not mean that most evolution is neutral because morphological variation almost invariably affects the ways in which organisms interact with their environment. It has, however, been found that the rate at which amino-acids in proteins and base-pair pairs in DNA are **neutrally** substituted is more or less constant in time, albeit specific for the individual protein or genetic sequence [Stearns and Hoekstra, 2005]. This has led to the idea of the molecular clock, which has been developed into a powerful tool for reconstructing evolutionary history, augmenting and complementing traditional paleontology.

3.8 The Neo-Darwinian Synthesis

The rediscovery of Mendel's laws of inheritance around 1900 [Keynes and Cox, 2008], closed an important gap in Darwin's theory of natural selection, namely the problem of the inheritance process. Darwin's original hypothesis of pangenesis [Darwin, 1868] was refuted experimentally [Galton, 1871] and subsequently abandoned[2], which lead to "the eclipse of Darwin" [Bowler, 1983], during which Darwin's ideas were largely ignored and Lamarck and other theories were favored. With the rediscovery of Mendelian inheritance [Bateson, 2009], Darwin's ideas slowly gained favor again. The statistical formulation of evolutionary change via population genetics and the combination with Mendelian inheritance and the concept of natural selection [R. A. Fisher, 1930; Wright, 1931; Haldane, 1932] and the realization that these concepts could explain wider phenomena in biological evolution [Dobzhansky, 1937; Mayr, 1999] eventually led to the "modern synthesis" or "neo-darwinian synthesis" [J. Huxley, 1942], in which population-genetics, Mendelian inheritance, and natural selection were combined into a single theory. This theory was later augmented by the discovery of the DNA molecule [J. D. Watson and F. H. C. Crick, 1953] and the introduction of the neutral theory of molecular evolution [Kimura, 1968]. Other noteworthy developments after its inception include inclusive-fitness theory [Hamilton, 1964] which is relevant for the evolution of cooperation and which later influenced the gene-eyed view of evolution [G. C. Williams and Dawkins, 2018; Dawkins, 1976], in which organisms only play the role of vehicles by which genes propagate copies of themselves into the next generation.

While subject to modification over time, its core tenets – that inheritance of acquired characteristics is principally excluded, that the occurrence of mutations is random and independent of their effects and that natural selection is the only process that can bring about adaptations – still form the backbone of much of evolutionary theory and teaching [Futuyma and Kirkpatrick, 2007].

In particular, its derived variant in the form of the gene-eyed view has received criticism over the past few years, with calls being made for an updated, or extended, evolutionary synthesis in the light of new findings [Pigliucci, 2009; Müller, 2007; K. N. Laland, Uller, et al., 2015; Müller, 2017b]. This has sparked considerable debate over some of its core elements [Scott-Phillips et al., 2014]. In particular, findings pertaining to the mechanics of development (see chapter 5) and the realization that inheritance is neither limited to genes nor fully excludes acquired characteristics (subsection 4.3.1) call into question the view that evolution is, essentially, repeated rounds of population genetics.

3.9 Summary

This chapter introduced two prominent processes that allow a population of biological organisms to constitute a Complex Constructive System, starting from the observation that all biological populations are made up of reproducing organisms. While more have been discussed in the literature [K. Laland, Matthews, et al., 2016], only Darwinian natural selection and Lamarckian evolution have been contrasted here. These have been chosen because they are the most well-known, but also because Lamarck has been frequently invoked over the last decade or so in the context of new discoveries in inheritance and in cultural

---

[2] As is so often the case, modern experiments seem to show exceptions to this refutation [Liu and Q. Chen, 2018].

evolution [Kronfeldner, 2006]. It has been discussed how both, Lamarckian and Darwinian evolution can give rise to CCS, whereby the Lamarckian case requires significantly more complex organisms than the Darwinian one. This is because performance evaluation and trait discovery are both associated with individuals in the Lamarckian case, given that performance evaluation depends on the usefulness of an organ for the organism's behavior in a given environment, while trait discovery depends on use and disuse. Performance evaluation in Lamarck's system is, moreover, associated with individual traits, not with the organism as a whole. Lamarck's theory leads to a continuous transformation of the trait memory of the individual and consequently the population as a whole, and it is not obvious how a usage-based trait discovery process as suggested by Lamarck would give rise to different species.

Darwinian natural selection, on the other hand, strictly proceeds by variation in reproduction and survival which is correlated with variation in individual traits. This correlation is mediated through ecological interactions, which allow some organisms to acquire the means for reproduction and survival more effectively or more efficiently than others. Inheritance thereby allows for the preferential retainment of those variants that confer higher reproductive success or survival than others. It must be noted that such a correlation, a priori, can be spurious, i.e., causal mechanisms in selection have to be demonstrated separately.

We noted that reproduction is a necessary and sufficient condition for the traits of an organism to stay in the population, while survival is only necessary but not sufficient. Consequently, evolution by natural selection generally will increase the lifetime reproductive success of organisms in a population over time, although this effect is, in nature, often intertwined with many others, like migration, or neutral evolution. While Darwin originally suggested a different inheritance model, we now understand that - first, the appearance of a particular variant is independent of its effects and that inheritance is particulate and does, again with some exceptions, exclude acquired characteristics. Therefore, Darwinian natural selection as a theory rests on the three pillars of variation, reproduction, and inheritance [Lewontin, 1970].

Performance evaluation, given by lifetime reproductive success, is, therefore, an emergent property of the system and depends on the pattern of interactions individual organisms are subject to. Consequently, Darwinian evolution is very general, and while it is the sole process that is universally accepted as producing adaptations over time and thus being able to give rise to CCS in biology, it can become active wherever there is variation in reproductive success correlated with some variation in traits, whatever the source or nature of the reproduction in question - chemical, technological, biological or social. This performance evaluation is bound to an individual as a whole because only individuals reproduce.

Trait discovery in Darwinian natural selection is generally driven by mutation and recombination of heritable traits. The latter is, in sexually reproducing organisms, a costly process that must be maintained by significant long-term advantages - the ability to mitigate mutational meltdown and a generally faster concentration of advantageous traits in single individuals probably among them [Stearns and Hoekstra, 2005]. Because these processes are bound to reproduction events and generally exclude acquired traits, this gives rise to discernible lineages with distinct properties over time, which is the ultimate cause for the tree structure of relationships between all living things.

Another consequence is that the trait memory of the system can be associated with individual organisms and lineages, but can also become communal through sexual reproduction or horizontal gene transfer [C. R. Woese, 2002]. Independent of the exact mechanism of inheritance, the traits in the population at any given time only represent the traits that allowed for reproduction in the past, with any past non-reproductive traits, even if they would be useful now, being excluded via reproduction-bound inheritance. Moreover, this creates the restriction that any mutational change to a trait must still allow for reproduction. Thus, natural selection produces structures that are characterized by compromises and suboptimal (by human judgment) or overly complex elements that an intelligent engineer would arguably not be inclined to accept.

Based on the system properties underlying natural selection, we proposed birth-death processes as a fundamental modeling tool for evolving populations and used them to discuss neutral evolution, i.e., the case where there is no correlation between traits and reproductive success. Under simplifying assumptions in asexual populations, the properties of neutral evolution have been discussed along the lines presented in Nowak [2006]. An important result is that the rate at which neutral traits become fixed in a population only depends on the mutation rate and is independent of the populationsize. Fixation of a neutral trait does effectively erase the trait memory of the population for this trait, which can be highly deleterious if this trait comes under selection through changed environmental conditions. This also shows that evolutionary change in a population is possible without natural selection.

4. Inheritance processes and the architecture of evolving entities

4.1 An abstract concept of inheritance and self-reproduction

4.1.1 Inheritance as a dynamical system

In this chapter, we focus on the inheritance process that underlies evolution and first present an abstract view of inheritance and self-reproduction that leads to a general understanding of Self-reproduction entails inheritance by definition because it implies that the produced entity resembles its parent. In general, inheritance is imperfect, although the error rate may be very low. Therefore, any self-reproducing entity is faced with the problem of how to transmit heritable traits to its offspring in a reliable enough manner, which can be seen as a mapping of traits from parent to offspring plus some limited variation. We thus interpret inheritance as a dynamical system I that operates on the set of possible configurations of heritable traits $\mathscr{H}$. We note that while such a system can in principle be formulated, the natural inheritance processes are generally way too complex to be analytically tractable, and so the formulation presented here is only used for illustration purposes and to highlight the necessary properties of self-reproducing systems as they emerge from the underlying assumptions of evolutionary theory.

Initial dynamical systems formualation of inheritance Upon reproduction, a configuration of heritable traits $h \in \mathscr{H}$ is transmitted to an offspring organism in a possibly modified way to yield a new configuration $h' \in \mathscr{H}$. This process, here called I must keep the parent organism functional by assumption. Therefore, a copy must be created:

$$\begin{aligned} &I = (D_I, \mathscr{H} \times \mathscr{H}, T,) \\ &D_I : \mathscr{H} \times \mathscr{H} \times T \to \mathscr{H} \times \mathscr{H} \times T : (h, 0, t) \mapsto (h, h', t'), t' > t \end{aligned} \tag{4.1.1}$$

where D_I denotes the copying process of heritable information[1] h onto another configuration h' and $T \subseteq \mathbb{R}$ denotes time. In the case of discrete generations, these can be used to measure time, $T \subseteq \mathbb{N}$. Here, we used $0 \in \mathscr{H}$ as a placeholder variable for the offspring organism.

Because we did not specify the properties of h and D_I any further, there are no a priori restrictions on what causes transmission errors. Neither is there any restriction as to what constitutes heritable information. Indeed, Darwin formulated his theory of natural selection without knowledge of the material basis of, or the laws underlying inheritance. Consequently, the logic of natural selection (and by extension neutral evolution) applies irrespective of the nature of heritable information, as long as Lewontin's three principles

---

[1] For now, we use the term "information" in its colloquial meaning, but will later introduce it more thoroughly.

can be established for a given system (compare section 3.3). The standard assumption is that transmission errors are causally independent of the structure and content of heritable information that is transmitted as well as independent of the effects they have, i.e., they are considered to occur "randomly", and are therefore modeled by a stochastic process [Ridley, 2004]. This makes Equation 4.1.1 a stochastic dynamical system, a notion not further explored here.

We note that this implies some form of dynamical isolation of the copying process, during which errors are introduced, from the rest of the system, a topic we will return to later. While this assumption is motivated by observations in nature, it is not strictly necessary for the logic of evolution to hold.

4.1.2 Self-replication as a dynamical system

We now try to embed the inheritance process into a dynamical system's view of self-replication, retaining the goal of illustrating causal structures in the system instead of aiming for quantitative modeling. Self-reproduction cannot occur in isolation but requires an organism to secure from its environment whatever is needed to make a physical copy of itself, as mentioned in chapter 3. These interactions are largely defined by heritable information[2] Variation between organisms in these interactions leads to natural selection by conferring variation in lifetime reproductive success and survival.

Consequently, there exists a process that uses heritable information and material acquired from the environment to create a copy of a self-replicator. This can be understood as an interpretation process of heritable information. As will be discussed later in more depth, the term "information" does only make sense with respect to such an interpreter process. This process relates the output of Equation 4.1.1 to a physical structure that is capable of self-reproduction again, whereby the latter requirement is a necessary condition for evolution to occur. We are, therefore, attempting to elicit the architectural requirements of a self-reproducing system.

Phenotype as physical implementation of the self-reproducing process Note that at this point, a distinction between the self-reproduction process itself P, and its physical implementation p is made: The latter can change in composition, as long as P remains functional. We call the physical system p that implements the process P the "phenotype" and denote the set of phenotypes that implement self-reproduction $\mathscr{P} \subset \hat{\mathscr{P}}$. For completeness, the set $\hat{\mathscr{P}}$ in turn denotes all possible phenotypes $\hat{p}$ that can be built by P, including dysfunctional ones that cannot self-reproduce. Thus, we include cases in which transmission errors during inheritance have catastrophic effects that preclude further self-reproduction, or the reproduction process itself is erroneous and results in defective phenotypes.

Dynamical systems formulation of self-reproduction With the set of environmental materials $\mathscr{E} \subset \hat{\mathscr{E}}$ from which a phenotype p can be produced (a subset of all locally obtainable environmental material configurations $\hat{\mathscr{E}}$) and the set of possible heritable information $\mathscr{H}$ we describe the interpreter process as

---

[2] But see chapter 5 for an extension and modification of this tenet.

$$\begin{aligned} P &= \left(D_P, \mathscr{E} \times \mathscr{P} \times \hat{\mathscr{P}} \times \mathscr{H} \times \mathscr{H}, T\right) \\ D_P :& \mathscr{E} \times \mathscr{P} \times \hat{\mathscr{P}} \times \mathscr{H} \times \mathscr{H} \times T \to \hat{\mathscr{E}} \times \mathscr{P} \times \hat{\mathscr{P}} \times \mathscr{H} \times \mathscr{H} \times T : \\ & (e, p, 0, h, 0, t) \mapsto (e', p, p', h, h', t''') \end{aligned} \tag{4.1.2}$$

Note that in Equation 4.1.2, the heritable information h' can be modified from h by virtue of imperfect inheritance, and consequently the newly produced phenotype p' will generally differ from the parent phenotype.

The environmental configuration $e \in \mathscr{E}$ can get degraded through the process of self-reproduction to a point where it is no longer viable. Then, $e' \in \hat{\mathscr{E}}\ \mathscr{E}$. For a natural example of e, think of sustenance, time and space that some animal uses to reproduce, which are, firstly, specific to given species and secondly contain many depletable elements. Food items for instance can typically only be used once by a given organism at a given time, and their supply successively depletes when the population of exploiters grows.

Self-reproduction as a combination of inheritance and interpretation processes Conceptually, we can further decompose D_P into the inheritance process Equation 4.1.1 and the interpreter process $\tilde{D}_P$ that produces the phenotype from the inherited information using environmental material. We first embed D_I into a more general map that includes the environment and the phenotype:

$$\hat{D}_I : \ (e, p, 0, h, 0, t) \mapsto (e, p, 0, h, h', t) \tag{4.1.3}$$

and extract the phenotype producing map from D_P:

$$\tilde{D}_P : \ (e, p, 0, h, h', t') \mapsto (e', p, p', h, h', t'') \, , \, t' < t'' \tag{4.1.4}$$

such that the two can be composed to yield the self-reproduction process D_P again:

$$\begin{gathered} D_P = \tilde{D}_P \circ \hat{D}_I : (e, p, 0, h, 0, t) \mapsto (e, p, 0, h, h', t') \mapsto (e', p, p', h, h', t'') \\ e \in \mathscr{E}, e' \in \hat{\mathscr{E}}, p \in \mathscr{P}, p' \in \hat{\mathscr{P}}, h, h' \in \mathscr{H}, t, t', t'' \in T, t < t' < t'' \end{gathered} \tag{4.1.5}$$

In this way, we identify self-reproduction as a process composed of transmission of heritable information and the "interpretation" of this information that uses and generally degrades environmental resources. We simplified this process in that only one copy is produced, while in nature, depending on the material supply and architecture of the organism, there can be many offspring at once of course. The ordering $t < t' < t''$ shows the temporal succession of the processes, which here is taken to be inheritance → interpretation. Note that this is not a necessary condition but a modeling choice, reversing the process only changes when mutations in h take effect: A new phenotype can alternatively be created using the existing set of heritable information h, after which h is copied and possibly mutated, and these mutations will take effect the next time reproduction happens. We also find that Equation 4.1.2 shows the characteristics of autocatalysis, in chemistry represented by $A + B \rightleftharpoons A + A$, although in our case we allow structural differences between educts and products, but not functional ones with respect to self-reproduction. Indeed, chemical autocatalysis is arguably the most basic self-reproduction process we find in nature and it thus is at the origin of biology [Hordijk and Steel, 2018].

4.1.3 Self-replication and ecological entanglement

As said before, not all environmental configurations e an organism can encounter or obtain from the environment are amenable to self-reproduction. A self-replicator is therefore faced with the problem to filter out viable environmental configurations $e \in \mathscr{E}$ from $\hat{\mathscr{E}}$ or transforming unviable configurations $\hat{e} \in \hat{\mathscr{E}} \setminus \mathscr{E}$ into viable ones. Examples include the acquisition and processing of food items, but also nest building, migration, the search for mates and many more. The evolutionary implications of this are further explored in chapter 6. Aside from that, organisms generally have internal conditions that must be fulfilled for reproduction to be possible, e.g., physiological states, nourishment, or location. Therefore, natural organisms transition within their lifetime to phenotypic states amenable to reproduction from states which are not, and, often, back again in a cyclic manner. This transition can be seen as yet another dynamical system, which we can envision as the life cycle from an initial state attained after the phenotype is produced from heritable information to a state of the phenotype and environment that supports reproduction. Consequently, this system models behavior and the ecological interactions organisms engage in - predator, parasite, herbivore, their entire behavior - essentially everything that the organism does. Consequently, the structure and function of this process is the main element on which evolution acts, because variation in this process has direct consequences on reproductive success. Consequently, most of the textbook examples of adaptations can be located in this subsystem. The process D_P therefore contains another subprocess L which implements these environmental interactions and transforms the environment and the phenotype into configurations that allow for reproduction. It can be formulated as

$$L = (\hat{e}, \hat{p}, 0, h, 0, t) \mapsto (e, p, 0, h, 0, t') \tag{4.1.6}$$

where we introduced $\hat{p} \neq p \in \hat{\mathscr{P}}$ to indicate that the phenotype p, in general not able to reproduce yet, is changed by L into a configuration $\tilde{p}$ which allows for reproduction, the same we already had included in Equation 4.1.5. The same is true for e, such that $L : \hat{\mathscr{E}} \times \hat{\mathscr{P}} \times \hat{\mathscr{P}} \times \mathscr{H} \times \mathscr{H} \times T \to \mathscr{E} \times \mathscr{P} \times \hat{\mathscr{P}} \times \mathscr{H} \times \mathscr{H} \times T$. It's worth emphasizing again that in this concept environmental states and physiological states are always modified together into a state that allows for the reproduction of the phenotype, as a direct consequence of the fact that self-reproduction can only occur in open systems. For example, feeding reduces and transforms the number of specific resources available, burrowing changes soil properties, and so on. Even more deeply, the production of the phenotype traps or alters the availability of environmental materials, which can have effects over geological times - e.g., in the carbon cycle of earth where long-term storages like limestone and mineral oil have biological origins. Environmental dynamics is therefore an inherent part of any evolutionary system. With Equation 4.1.6, we now modify Equation 4.1.5 to include ecological interaction:

$$\begin{aligned}
L &= (\hat{e}, \hat{p}, 0, h, 0, t) \mapsto (e, p, 0, h, 0, t') \\
D_P &= \tilde{D}_P \circ \hat{D}_I \circ L : \\
D_P &= (\hat{e}, \hat{p}, 0, h, 0, t) \mapsto (e, p, 0, h, 0, t') \mapsto (e, p, 0, h, h', t'') \mapsto (e', p, p', h, h', t''') \\
&\quad \hat{e}, e' \in \hat{\mathscr{E}}, e \in \mathscr{E}, p \in \mathscr{P}, \hat{p}, p' \in \hat{\mathscr{P}}, h, h' \in \mathscr{H}, t, t', t'', t''' \in T, t < t' < t'' < t''',
\end{aligned} \tag{4.1.7}$$

4.1.4 Reproduction as guided self-organization on multiple scales

In Equation 4.1.7, we modeled the process of self-reproduction as a system that creates new instances of itself by continually drawing on the environment it exists in for energy and resources.

Self-reproduction is conceptually inseparable from environmental changes These environmental elements of course are not inert but have their own complicated physics, which is often capable of forming complex– and complex adaptive systems by itself.

This, once more, constitutes a dynamical system $D_E : \hat{\mathscr{E}} \times T \to \hat{\mathscr{E}} \times T$. Note however that the environment relevant here includes not only abiotic elements but also biotic ones, as discussed in chapter 3. Given this view, we find that D_P and D_E are coupled systems that influence each other instead of comprising one-way dynamics. This picks up the assertion made in chapter 2 that complex adaptive– and complex constructive systems are generally open systems and self-reproducers fall into these categories. This coupling is generally very complex as it can be mediated by many biotic and abiotic factors.

Given that L and D_E unfold over time, we find that organisms create forcing on the environment via L, while D_E in turn modifies e in the arguments of L. We recognize this reciprocal forcing as one of the hallmarks of a self-organizing complex system, as we already mentioned in chapter 2. Recalling Figure 3.1, we can now identify the links that connect an agent node to the network as a representation of the elements transformed by the system L, i.e., L defines the ecological niche of the organism in question. The coupled system D_P, D_E therefore defines the self-organized unfolding of the ecosystem network. More concisely, environment and biology, and hence, evolution, are not dynamically separable.

Self-reproduction can be seen as guided or manipulated self-organization Next we focus on the subprocess $\tilde{D}_P$. As we did with L, we skipped over the details of the unfolding that leads from $(e, p, 0, h, 0, t)$ to (e', p, p', h, h', t'). This is another form of self-organization of matter from environmental elements and a phenotype to a transformed phenotype and a transformed environment and a possibly imperfect copy of p. If we employ the language of chemistry for a moment, the phenotype of the reproducing agent p thereby functions as a catalyst for the production of p', using e, h as reaction educts. From above, we understand that e has its own, generally complex, dynamics, capable of independent self-organization. It is this self-organization that p generally leverages to create a copy of itself. In other words, biology consists of systems that "guide" the pre-existing dynamics of abiotic systems into reliably (re)creating self-reproducing entities. Again drawing on chemistry, we can think of the folding of proteins, RNA[3] or the tendency of some lipid membranes to form spheres or droplets in water as self-organization that is used by self-reproducing organisms to form phenotypes or has been used in the past. Similarly, the population of different symbionts inherent to all multicellular life [Chiu and Gilbert, 2015] has self-organizing features which are manipulated by the host's immune system. Indeed, populations of cells in multicellular bodies do self-organize into tissues and organs via a dynamic interaction network (see subsection 4.2.5). The heritable information h sets boundary- and initial conditions and establishes forcing at each stage of the self-organized unfolding of the phenotype [Rocha, 2001]. In this way, it assures that the self-reproduction process creates reproductively viable phenotypes in dynamic environments, i.e., it shapes $\tilde{D}_P, L$ in such a way that a viable configuration of the phenotype-environment system is reached effectively. We, therefore, understand self-reproduction as a complex form

---

[3] Ribo-nucleic acid. Analogous to DNA but uses ribose as a sugar backbone and uracil instead of thymine.

of guided self-organization of matter [Rocha, 2001], where heritable information can be seen as encoding a scaffolding system that guides these processes. This "guidance" is consequently what is shaped by natural selection.

In summary, guided self-replication on the level of individual organisms reorganizes the local environment and forms an emergent complex system, the ecosystem, which feeds back to the dynamics of individual organisms. It does so by structuring how each individual can interact with the rest of the ecosystem, which creates variation in reproductive success and survival, i.e., natural selection. The self-replication process D_P decays formally into three parts which are the inheritance process D_I that creates a copy h' of the heritable information h which is provided to the offspring by the parent and serves as the primary guide to the self-organized creation of the phenotype p via $\tilde{D}_P$. The latter acts by drawing on the content of h' in order to reorganize environmental elements e into a new phenotype p'.

In this process, we can think of p as acting as a catalyst for self-organization, while the heritable information h' can be seen as providing forcing and boundary conditions, such that the whole dynamical system produces a new functional phenotype. To be able to unfold anew via $\hat{p}$, the environment and the phenotype must reach a viable state (e, p) first, which is represented by the process L, i.e., both are generally actively manipulated by organisms into such a state. This process constitutes all ecological activity an organism engages in.

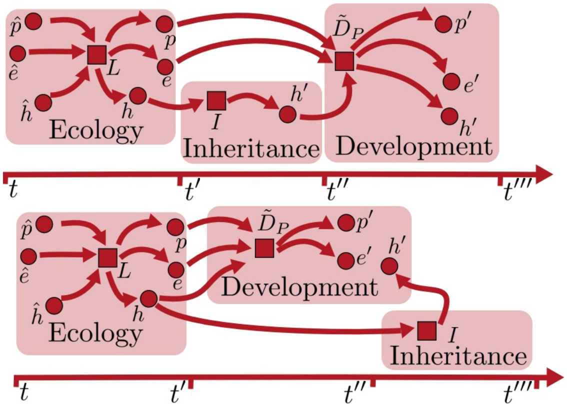


Figure 4.1: Self-reproduction as outlined in Equation 4.1.7 can be logically arranged in two ways, which differ in the point where variation acts. **Upper:** The scheme as presented in Equation 4.1.7, where development acts on the copied and possibly modified heritable information. **Lower:** A possible alternative is that the development process acts on the parent's heritable information directly, with heritable information being processed independently, such that variation takes effect later.

4.1.5 Corollaries

Because self-reproduction can be seen as guided self-organization, the process is dependent on initial conditions, i.e., on variations in environmental configuration e and in heritable variation h. Such sensitivity becomes less and less desirable the more complex p becomes because adaptations tend to be lost via the generation of variation. We can understand this as a lack of control, or insufficient guidance, of the self-organization process. Consequently, we expect natural selection to increase the control over the self-organization of e into $\hat{p}$ via h, i.e., natural self-replicators should be dominated by the forcing effects of heritable information to be able to maintain complex adaptations. Because the generator process for $\hat{p}$ generally has many degrees of freedom and is nonlinear, full control over the process cannot be generally achieved via initial conditions, because such systems tend to exhibit sensitive dependence on the latter. Therefore, heritable information must produce forcing of the system at many intermediate steps, i.e., the process must be broken down into many subprocesses with lower internal dimensionality, which are then feeding into each other. In this manner, the problem of development becomes a problem of control across multiple levels of organization. Moreover, we can expect this process to be able to correct at least some variations in this unfolding such that even if material or forcing is corrupted to some degree, a viable phenotype can still be produced. This expectation is derived from the fact that such robustness would allow for the maintenance of fitness even under suboptimal conditions and the creation and maintenance of more complex phenotypic traits.

There are arguably many ways in which this can be achieved. We can imagine functional redundancy to be one of them, i.e., of the myriad ways in which elements of the process can interact, many are equivalent. Another would be specificness, i.e., an element evolves structurally to only be able to interact in very limited ways with others. We can envisage both being present for different functions - redundancy provides robustness, while specificity allows for optimized adaptation to complicated conditions or the precise formation of complex structures. With this insight, we again arrive at the notion of biological systems as nested conglomerates of CAS and CCS. Indeed, a learning system that is able to recognize and reproduce certain patterns of forcing from incomplete or corrupted heritable information [R. A. Watson, G. P. Wagner, et al., 2014] would arguably constitute the upper end of the robustness spectrum.

Robustness via some form of buffering against harmful variation can happen in many ways, however. Examples include changing the local environment actively (called “niche construction” [Odlin-Smee et al., 2003] or “ecosystem engineering” [Jones et al., 1994], compare also subsection 4.3.2 and chapter 6), internalizing physical conditions into p, which arguably happened with endothermy or placental embryogenesis and which transfers elements of the reproduction process that where originally part of the surrounding ecosystem into elements of the phenotype which eventually become directly heritable [Laubichler and Renn, 2015; E. Danchin, Pocheville, et al., 2018], or by adapting the dynamics of L to changing environmental conditions during the existence of p - a process called “phenotypic plasticity” [Pfennig, 2021; Sommer, 2020]. Phenotypic plasticity takes on many forms, from the thicker pelt some animals get in winter to the formation of defensive structures as a response to predators [Agrawal et al., 1999] to learning from experience. It can affect both L (which is the part normally meant by “phenotypic plasiticity”) and $\tilde{D}_P$ (which is normally separated into “developmental plasticity”, [Gilbert et al., 2015], compare chapter 5).

As mentioned, variation in L between organisms can give rise to natural selection. Because the functional structure of L is strongly influenced by how h impacts the generation process $\tilde{D}_P$, natural selection will act to modify the distribution of heritable information in the population in favor of those L which are more efficient and more effective. Therefore, while the self-organization process that creates the phenotype will evolve to be more tightly controlled by heritable information, we expect the efficiency and effectiveness of L to increase by using more environmental cues to optimize its functioning on the fly (plasticity), by controlling environmental factors directly through modification (niche construction [Odlin-Smee et al., 2003; K. Laland, Matthews, et al., 2016]) or by progressive internalization of variable but essential environmental elements into the heritable information of the organism (Laubichler and Renn [2015] and E. Danchin, Pocheville, et al. [2018] and references therein, also compare chapter 5).

Adding evolutionary continuity, exaptations and cooption to these concepts, we can expect the dynamical system D_P to feature a nested, networked architecture that becomes deeper throughout evolution with its core elements being essentially universal to all self-replicating systems derived from a common ancestor, a concept that is explored more deeply later in subsection 4.2.6 and chapter 5. Consequently, we expect at least the basic system for constructing a phenotype to freeze out and evolve only extremely slowly, with further evolution being largely confined to the higher levels of the organizational hierarchy. Finally, where tighter control in order to limit harmful functional variation cannot be achieved, we should expect the process to evolve towards buffering unavoidable variation - e.g. through redundancy in heritable variation, or degenerate dynamics which maps many variants of some heritable forcing to the same effect in the phenotype.

4.1.6 Heritable information

Definition of heritable information Before discussing natural incarnations of the scheme presented in subsection 4.1.2 and subsection 4.1.4, we must clarify the nature of heritable information. Until now, we only postulated that it depends on the presence of an interpreter process that reacts to it non-randomly, but otherwise used the term "information" in its colloquial sense. There are at least two ways to approach this problem. The first is the physical one, pioneered by Shannon [1948], from whom we have the expected information content of a measurement of a random variable X

$$H = -\sum_i p_X(x) \log_b (p_X(x)) \tag{4.1.8}$$

with b being the logarithm base of choice, $p_X(x)$ being the probability that an event $x \in X$ occurs, and $-log_b (p_X(x))$ is the self-information of the event x. In that way, H can be interpreted as the average information obtained from a message, i.e., the lower the probability of a message to occur, the higher its self-information [Werner, 2008]. This definition renders information as a physical quantity closely related to thermodynamical entropy [Shannon, 1948]. As such, it does not include any notion of meaning or interpretation. The concept of information we need here, however, is one that explicitly takes into account its effect on a receiver, because we ultimately wish to arrive at a definition of heritable information h which is central to the self-reproduction process. Therefore, we understand information semantically - as some pattern that leads to a change in other patterns [Casagrande, 1999]. We can interpret the changed pattern as a receiver of information and the process that elicits this change as interpretation. Both patterns in the above concept are ultimately physically encoded. However, there can be many such encodings leading to the same changes in a receiver, and therefore it makes sense to conceptually distinguish information from its encoding, although nature always operates on the latter.

From this discussion, we can pick up on "biological information" as defined by Jablonka [2002], p.582: *"A source - an entity or a process - can be said to have information when a receiver system reacts to this sources in a special way. The reaction of the receiver to the source has to be such that the reaction can actually or potentially change the state of the receiver in a (usually) functional manner. Moreover, there must be a consistent relation between variations in the form of the source and the corresponding changes in the receiver."* They go on to clarify *"[...] form is the* ***organization*** *of the features and/or the actions of the source, and specifically those aspects of organization with which the* *receiver interacts[...]"* and *"[...] the processes in the receiver that result in a regular and functional response will be called 'interpretation'[...]"*. This is largely congruential with the definition of information used above, so in fact, although Jablonka [2002] only refers to biological information, there is no apparent restriction that would limit this definition to biology.

We specialize this definition to a semantic definition of heritable biological information in the following way: *Given a set of self-replicating entities, heritable information is all (biological) information that can be transmitted from one generation to another and for which variation in transmitted content can elicit systematic variation in the offspring's structure or function, which itself can further be inherited by future generations.* In other words, heritable information is biological information transmitted within a lineage of reproductively connected organisms that can influence the structure or trajectory of dynamical systems L and $\tilde{D_P}$. Heritable information is thus a special case of biological information associated with specific transmission and interpreter processes.

Corollaries The above definition does not include any notion of the origin of heritable information, nor does it specify its physical encoding system. It merely states that there is a process that transmits biological information sensu Jablonka during self-reproduction which, in turn, influences the organization of the offspring organism. It follows that heritable information is associated with specific lineages of organisms and is transmitted vertically from parent to offspring. This mode of transmission of biological information we call "Darwinian", in accordance with the "Darwinian Transition" from horizontally dominated to vertically dominated information exchange during the evolution of cells as proposed by C. R. Woese [2002].

However, the definition also allows non-heritable biological information to be transformed into heritable biological information. For example, horizontal gene transfer can change the genome of a bacterium which then is copied and transmitted to its offspring bacteria [Stearns and Hoekstra, 2005; Koonin, Makarova, et al., 2001], while at the same time influencing the bacterium's phenotype. Communication between adult organisms can modify their behavior and in turn is picked up by their offspring, a phenomenon common in social mammals. This transformation of non-heritable to heritable information can cross the chasm between encoding systems, as the Baldwin effect exemplifies [Baldwin, 1896; Badyaev, 2009]. Here, a behavioral trait that has originally been acquired by organisms during their lifetime as a reaction to some environmental state change becomes eventually genetically encoded and is subsequently inherited by future generations without the need for the original environmental signal that elicited the original behavior.

Therefore, the system described by Equation 4.1.7 has to be modified to accommodate the possibility of this assimilation of acquired information:

$$\begin{aligned}
L &= \left(\hat{e}, \hat{p}, 0, \hat{h}, 0, t\right) \mapsto (e, p, 0, h, 0, t') \\
D_P &= \tilde{D}_P \circ \hat{D}_I \circ L : \\
D_P &= \left(\hat{e}, \hat{p}, 0, \hat{h}, 0, t\right) \mapsto (e, p, 0, h, 0, t') \mapsto (e, p, 0, h, h', t'') \mapsto (e', p, p', h, h', t''') \\
&\quad \hat{e}, e' \in \hat{\mathscr{E}}, e \in \mathscr{E}, p \in \mathscr{P}, \hat{p}, p' \in \hat{\mathscr{P}}, h, h' \in \mathscr{H}, t, t', t'', t''' \in T, t < t' < t'' < t''',
\end{aligned} \tag{4.1.9}$$

Because the above definition does not specify any single encoding system, in principle it is possible to have multiple encoding systems for heritable information at the same time, although we lumped them all together into a single variable in subsection 4.1.1 and subsection 4.1.2 for simplicity. However, h can generally contain elements that are of vastly different physical nature but play a similar role in self-reproduction. The obvious encoding system we always think about in evolutionary theory is of course the genetic one, but cultural content learned from parents, peers or other members of a social group can similarly influence an organism's phenotypic development, as can environmental artifacts and conditions chosen or altered by earlier generations. More generally, environmental conditions can lead to developmental changes (recall the discussion of the dynamical system L above), which can be transmitted to future generations or can become encoded in the heritable information eventually. Heritable information consequently comes in different "classes", which, via their underlying physical structure, show varying degrees of stability against transmission errors and over which an organism has variable degrees of control. Along the same line, these classes can be monopolized by a single lineage to varying degrees.

Recalling that heritable information by definition leads to structural or functional changes in the receiver, and given that these effects can be adaptive and expensive to express and maintain, the more important consequence of monopolization appears to be the degree of control a lineage can exert over its heritable information. This can also be taken to include the stability of the encoding system – the more stable the encoding, the less uncontrollable transmission errors we can expect to happen. Consequently, heritable information can be qualitatively classified according to the degree of controllability by its carrier, split into the stability of the encoding structures first and the degree of monopolization its carrier lineage has second.

Different encoding systems for heritable information that has different effects can also lead to coevolution between them. Coevolution will arise whenever the presence of one trait influences natural selection acting on another trait. Concerning different encoding systems, gene-culture coevolution is perhaps the most striking example of such coevolution, with human evolution towards higher lactose tolerance in reaction to the advent of dairy farming being its most outstanding instance [K. N. Laland, J. Odling-Smee, and Myles, 2010]. Though most apparent in humans where culture is most developed, we find similar instances in other organisms [Whitehead, 2017; Whitehead et al., 2019]. Generalizing this idea, we would expect a network of coevolutionary interactions between traits governed by different encoding systems. Finally, assuming an autonomous origin of self-replication makes it probable that different inheritance systems arose at different times, and that, consequently, the most ancient ones are universal across the tree of life. More generally, the emergence, or disappearance, of different inheritance systems can, but is not required to, be associated with branching events in evolutionary history. Again, cooption of structures that may originally have had another function or no function altogether arguably played a major role in the formation, maintenance and shaping of inheritance systems.

4.1.7 Self-reproducing architectures

The question of the fundamental architecture of self-reproducing entities has been longstanding, tied as it is to questions about complexity and open-endedness in evolution. As such, we ask how a biological CCS must be constructed in order to allow for the continuous creation of new adaptations or the formation of novel complex structures. Throughout this thesis, we assumed self-reproduction to emerge and be maintained autonomously, i.e., without any guidance or dynamical influence outside of what is represented by Equation 4.1.7. The arguably most important part of any CCS is the trait memory structure because it is the fundamental element on which trait discovery and performance evaluation act. It must be noted that the inheritance process I was largely kept abstract in subsection 4.1.1. We did not make any assumptions on how transmission happens, nor did we prescribe any encoding system. On this abstract basis, the following section discusses possible architectures that could implement the system in Equation 4.1.7.

Template-based reproduction Rocha [2001] discusses two possible architectural principles, based on von Neumann [1966] and [Pattee, 2012b; Pattee, 2012a]. The first is template-based reproduction, which we can understand as "reverse engineering" or reproduction by self-inspection. The principle is that the produced structure, p in the outline above, does function as an encoding system of heritable information and as an interpreter process at the same time. In the scheme outlined above, there is no difference between I and $\tilde{D}_P$ in such a case, rather, heritable information is encoded in the structure of p directly. Therefore, h is not separate from p either. Natural examples include prions [Prusiner, 1998], which refold healthy proteins into their own shape, but also ribozymes [Scott, 2007], autocatalytic RNA molecules which play an important role in the RNA world hypothesis for the origin of life. More generally, this mode of reproduction includes all autocatalytic chemical reactions and thus constitutes the primordial mode of reproduction before the emergence of translation. An influential theoretical example is the hypercycle [Eigen and Schuster, 1977], which builds on the idea of stacked (auto-)catalysis as a precursor to cellular life.

For a deeper consideration, we recall that self-inspection-based reproduction, like all reproduction, must operate on some material substrate that is reorganized into the self-reproducing configuration. Adding autonomous unfolding, this means that self-reproduction is an emergent process, with the only currently known properties being that it allows for "complex enough" interactions among its elements to allow for this emergence. When and under what circumstances this is the case is by no means obvious, but this question goes far beyond the scope of this thesis. We thus postulate the existence of such a material substrate, picturing organic chemistry as its most natural example.

In this way, we understand self-reproducing systems as autocatalytic interaction networks. Consequently, self-reproduction is encoded in the structure of the system - the particular arrangement of interactions that enable autocatalysis - just as much as in the particular composition of elements it is made up of. Finally, this leads to the question of how control over the self-organization of material into a copy of such a system is achieved, and to what limits it is subject to. Rocha [2001], building on the idea of dynamical systems describing self-organization, frames this as a problem of recovering the initial conditions under which the autonomous self-organization process of material yields a copy of the particular self-reproducing structure. Taking into account that these initial conditions must be recovered from the fully articulated system, they then go on to posit that this limits the complexity such structures can attain because the self-organization of complex structures is generally not uniquely invertible (i.e., these systems are dissipative), such that the heritable information needed is lost in a template based scheme.

Aside from this, the conflation of heritable information and functionality results in conflicting selection, assuming that at some early stage in the evolution of these structures, there is variation in both: Adaptations for a given environment should be preserved under imperfect reproduction, so more stable, less dynamic configurations would be favored (compare the above discussion of evolution towards robustness). Because structure and encoding are identical in template-based reproduction, this would yield a self-organization process with a strong attractor and a large basin of attraction. In turn, this limits evolvability, such that adaptation to environmental changes is strongly restricted.

Furthermore, while template-based reproduction might be possible for comparably simple molecular entities like proteins for RNA sequences, the complexity of even the simplest cells we know of today would be impossible to evolve with such a system due to the multitude of intertwined and co-dependent subsystems these have. This holds even when we go beyond the notion of autocatalysis and envisage a system where some factory system builds new instances of another entity via inspecting existing ones, i.e., reverse engineering in the proper sense of the word. In such a system, the initial and boundary conditions needed for the factory to produce an exact copy must be recoverable from a fully articulated form, which is generally not possible for complex, dynamic arrangements as mentioned above. Paralleling the above discussion, the system ought to either remain simple enough for this to be possible or so general that essentially all possible trajectories yield the same attractor, which severely limits evolvability. Note that in such a system, complex structures still can form along the lines presented in subsection 2.3.2, but this does not alleviate the restrictions on evolvability discussed before, it merely pushes them further up the organizational ladder.

Code-based reproduction As an alternative to template-based reproduction Rocha [2001] discuss how a self-reproducing system can be physically split into heritable information encoded in some separate physical system(s) on the one hand and its interpretation that is created from environmental material by a separate process. The latter uses initial- and boundary conditions and forcing provided by heritable information. We can identify the encoded heritable information in Equation 4.1.7 with h, the created material interpretation with p and the self-organization process with $\tilde{D}_P$. Following Rocha [2001] further and restricting ourselves to the most simple cases, e.g., the ones probably prevalent at the origin of life, we find that $\tilde{D}_P$ depends on the properties of the material e that reorganizes into p: The structures p can attain depend on the dynamics of the elements of e. Heritable information restricts, or controls, these processes such that they result in a functional copy of the replicator, but only with respect to the physical dynamics of these “building blocks”. Therefore, there is no analogon to Turing-completeness [Turing, 1952] in biology, i.e., there is no singular material system that can be used to create any arbitrary phenotype. Rather, every material system on which guided self-organization unfolds will be subject to intrinsical physical limitations. In molecular biology, we can identify the building blocks with amino acids, and the self-organization process that yields the phenotype consequently starts with the folding of amino acid chains into proteins, which further organize into more complex structures within cells.

The fundamental difference that separates cellular self-reproduction systems from template-based reproduction is the encoding of heritable information into a structure separate from p. Employing the simple molecular case as a guideline, we find that h plays no direct mechanical role in $\tilde{D}_P$, but rather *represents* the elements that do, through the codon-amino-acid correspondence. Consequently, we can think of DNA[4] as a “symbolic” representation of the initial conditions for protein folding, namely the amino-acid chain. This representation follows certain rules, like the Adenine-Thymine and Cytocine-Guanin correspondence for base pairs or that three base pairs make a codon representing an amino acid. This, Rocha calls “syntactic”, as opposed to “semantics”, i.e., $\tilde{D}_P$, implementing the semantics, creates a physical system using this symbolic representation that itself only follows syntactic rules.

---

[4] Desoxyribo-nucleic acid. Macromolecule made up of a sugar-phosphate backbone with one of four bases (adenine, thymine, cytosine, guanine) attached to it, used in long chains to store genetic information in cells.

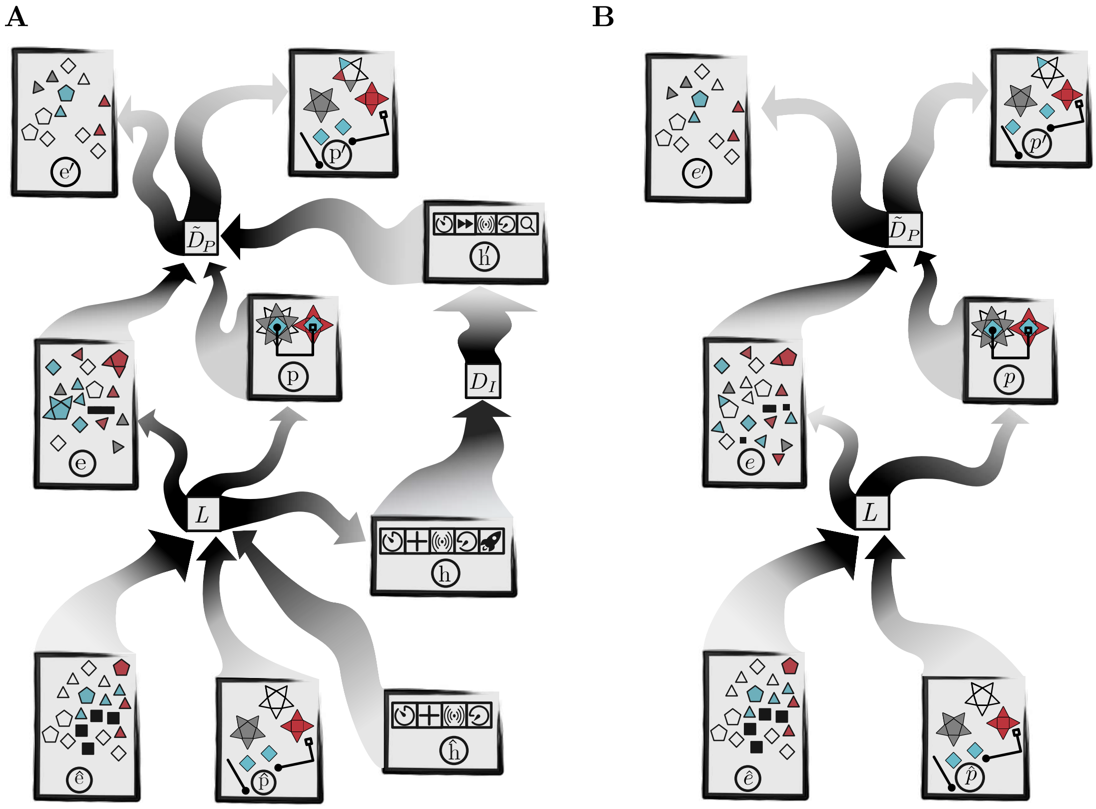


Figure 4.2: **A:** Self-reproduction as conceptualized in Equation 4.1.9. The dynamical systems $L, \tilde{D}_P, D_I$ are all implemented by the phenotype $\hat{p}$ in conjuction with the heritable information h. The latter provides forcing and boundary conditions on these systems. L thereby represents the interactions of the phenotype $\hat{p}$ with its environment $\hat{e}$ and transforms both into a state e, p that is amenable to the reproduction of p. Heritable information can have a restricting role in this process, but it can also be modified, $\hat{h} \rightarrow h$. Self-reproduction then proceeds via inheritance D_I, which at least partly processes symbolic representations, and construction of a new phenotype p' from environmental material, restricted and guided by inherited biological information. Inheritance is not perfect, and transmission errors can lead to phenotypic differences, creating a variation on which natural selection can act. Reproduction always influences the environment at least via material usage. **B:** Without a syntactic-semantics split of at least parts of the heritable information, $\tilde{D}_P$ and D_I are conflated, which does entail reduced evolvability due to a lack of universality and evolutionary conflict due to the aforementioned conflation of roles. Changes to the phenotype due to environmental interaction are the sole source of variation in this system and can be inherited by offspring. In nature, both extremes are generally mixed, because heritable information is partly symbolically encoded (genetic, partially cultural) and in other parts, it is conflated with functional structures (e.g., environmental or heritable symbionts).

Syntactics and semantics in code-based reproduction We recognize that D_I is only concerned with syntactics, not with semantics. More than that, given a certain encoding scheme, e.g. the genetic encoding of amino acid sequence or the syntactic rules of our languages, the complexity of the semantics is irrelevant for D_I. It only needs to be capable of processing symbols following the syntactic rules, e.g., the base pairs and codons, or the syntax rules of a language. If this is achieved, any "message" can be processed, irrespective of its content. This allows us to define a notion of universality for the syntactic system D_I, i.e., it is universal with respect to some encoding system if it can process any arbitrary message encoded in this scheme.

This restricts encoding schemes used for universal syntactic systems to those that have many more or less equivalent stable states, i.e., a system that has a strong local attractor would not be usable because encoded messages would eventually decay into this attractor, thereby losing the message's content.

Furthermore, natural syntactic systems do not make use of the full breadth of dynamics of their respective encoding system but use only a subset of it. For instance, we can produce a lot more sounds than are used in any of our languages. Especially with regard to DNA, it has not been fully resolved if the particular code and its material substrate are the product of natural selection or a frozen accident [Ridley, 2004; Vetsigian et al., 2006; Koonin and Novozhilov, 2017].

Focusing on semantics again, we find a similar notion of universality. An interpreter process is universal with respect to an encoding system if it can process any message composed in this system, i.e., it translates arbitrary messages into forcing and boundary conditions for phenotype creation. With an imperfect copying process in D_I, such a system can yield open-ended evolution within the confines of the physics underlying the self-organization process $\tilde{D}_P$ and the symbolic system that makes up h. In this way, a self-reproducing system that employs separate semantics and syntactics that attained universality with respect to some encoding scheme of heritable information is far more powerful than the template-based replicator discussed above, where semantics and syntactics are implemented in one and the same physical system. We thus gained insight into possible organizations of biological CCSs trait memory and their implications, showing that these are subdivided into more and less powerful subclasses with weaker and stronger evolvability.

The von Neumann universal constructor as a prototype for code-based self-replication Such a system has been implemented in nature at the basis of the genetic system and constitutes the basic function of a simple cell. Prior to the discovery of this system, it was developed theoretically by von Neumann [1966], who discussed the idea of self-reproducing automata. His concept splits the self-reproducing automaton into four parts, which we can identify with the elements presented in Equation 4.1.7 and Figure 4.1.

The first element is a universal copier, which we can identify with the process D_I, that is solely concerned with copying the heritable information h into the new replicator without caring for its semantics in any way. Next is a universal constructor that is capable of reading any instruction set and turning it into a copy of the entire automaton by using available material. Third is some functionality within its environment that allows for the acquisition of resources for reproduction. Lastly, an element is needed that governs the order in which the individual subprocesses are executed, which is called a developmental program in Figure 4.3 and Rocha [2001]. It must be noted that the von Neumann automaton uses its heritable information as a blueprint for the constructor to read and execute, which differs from the approach discussed above. Instead of a

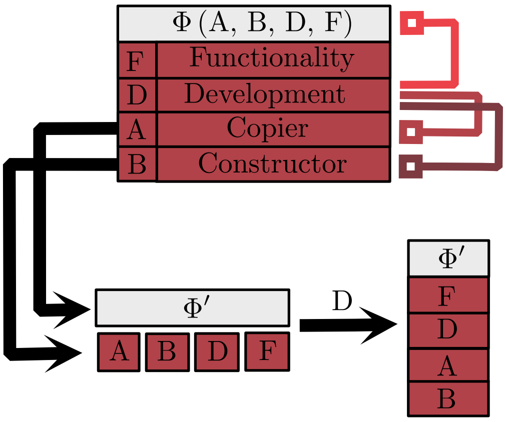


Figure 4.3: The axiomatic concept of a self-replicator von Neumann envisaged consists of four different parts, a universal constructor able to decode any description Φ (B), a universal copier able to copy any description (A), a developmental program (D) that coordinates the replication process and additional functionality (F), e.g. metabolism. If augmented with an encoding for each of the subprocesses $\Phi\,(A, B, D, F)$, the system is capable of evolving arbitrary complexity by accumulating mutations, $\Phi \to \Phi'$. Shades of red from bright to dark indicate a possible temporal sequence in which the developmental program organizes reproduction.

blueprint, h is understood to provide dynamical restrictions to a self-organizing system in the concept presented here. Hence, we identify the developmental program and the universal constructor in von Neumann's concept with the "interpreter" process $\tilde{D}_P$. This encompasses the construction/self-organization process and one part of the developmental program, with the other being supplied by the restricting influence of h. This conflation is necessary because within von Neumann's concept, heritable information has no dynamic role, it is only copied by the copier and read by the constructor. In contrast, the idea that it provides forcing for the unfolding process of the $\tilde{D}_P$ in addition to boundary conditions assigns it an active role in the dynamics of the process, i.e., the activation of specific parts of h at given times activate specific elements of $\tilde{D}_P$ and vice versa. Finally, we readily identify the functionality part F in Figure 4.3 with the process L in Equation 4.1.7. The cell cycle of natural cells therefore can be identified with $\tilde{D}_P \circ L$.

It is remarkable how well this concept fits the basic architecture of living cells, which have distinct mechanisms for transcription of DNA and translation into amino acid chains which fold into proteins largely by self-organization, and which across all known instances feature strong similarities in the genetic code. Moreover, the basic components of transcription and translation processes are also extremely similar across the tree of life, although there is astonishing diversity in the functional capabilities and ways organisms interact with their environment.

From a computational point of view, von Neumann's scheme can be interpreted as a Turing machine, where the encoding system specifies the instruction set it operates on. That being said, natural organisms are far more complex than von Neumann's scheme, and the fact that heritable information plays a more active role in reproduction than that of a blueprint or instruction set places limits on this analogy. Heritable information, as discussed in Equation 4.1.6, comes in many shapes and forms, some of which have a symbolic encoding system in which semantics and syntactic are separated while they are conflated in others.

Additionally, a semiotic split between syntactic and semantic processing brings also control over the construction- and developmental process by allowing for the fine-grained specification of timing and forcing. This is essentially impossible to achieve in a template-based system because the exact trajectory that must be followed to arrive at a particular attractor is not accessible anymore from the fully developed structure. Therefore, when trying to classify inheritance systems according to their controllability, the presence of this semantic closure [Pattee, 2012b; Pattee, 2007] is a useful empirical criterium. All in all, we conclude that the concept as presented in Equation 4.1.7 and Equation 4.1.9, which is largely parallel to von Neumann's system, is a useful and correct abstraction for discussing biological evolution, which by extension makes cells its fundamental building blocks. Of special importance is thereby the concept of evolution as guided self-organization introduced by Rocha, which we will use as a guiding principle throughout the rest of this thesis. This insight also allows us to fill in some gaps left in the concept of biological CCS, in that we now understand how their trait memory must be organized in order to allow for arbitrary adaptations and the emergence of complex structures from simpler ones.

4.2 Genetic inheritance

Genetic inheritance has been found to underly the entire domain of life, with closely related encodings, construction and copying processes having been found in every organism on the planet. This not only provides evidence for a common ancestry of all living things but also tells us that the biosemiotic separation between syntactics and semantics and the associated symbolic encoding arose early in life's history, and especially before the emergence of LUCA[5], although it might have been a lot more evolvable during these early phases than it is today. Genetic inheritance is also the only inheritance for which a fully formed, universal biosemiotic system exists - the mechanisms for transcription and translation within cells are such that they can process any genetic sequence within the boundaries of the underlying genetic code. It is beyond the scope of this thesis to discuss the origin or biochemical details of the genetic machinery. Rather, this discussion emphasizes the role the genetic system plays within the self-replicating system discussed before.

4.2.1 Genetic encoding of heritable information

All cellular organisms, and consequently all multicellular organisms as well, use DNA (Deoxyribonucleic acid) as the material basis for encoding heritable information.

Base-pairing as the lowest layer of semantics DNA is a macromolecule consisting of two parallel strands, each of which is made up of a sequence of "nucleotides" [Ridley, 2004]. Each nucleotide contains two parts - a sugar-phosphate backbone and a base. There are four bases in DNA - adenine (A), thymine(T), cytosine(C) and guanine(G). The DNA double strand is formed by pairing nucleotides in a specific fashion - A is paired to T and C is paired to G. Together, the two strands form a complementary double-helix structure [J. D. Watson and F. H. C. Crick, 1953]. The two strands run anti-parallel, which is important for DNA replication, see subsection 4.2.2. The direction is counted using the numbering convention of the carbon atoms in the ribose ring, with the phosphate group being attached to the $5'$ end, while the $3'$ still has its unmodified OH-group. Heritable information is encoded in the sequence in which the nucleotides occur in the DNA molecule. The complementary nature of base-pairing provides the first

---

[5] Last Universal Common Ancestor, i.e., the youngest organism to which all living things today are related.

layer of syntactics. It also means that the two strands of DNA are redundant – they contain the same information. As Rocha [2001] mentions, the chemical structure of the DNA with its repeating elements, bound together in the same way, makes sure that there is no preferred sequence that the molecule will tend towards, allowing for the encoding of arbitrary content[6]. Furthermore, the hydrogen bonds of the base pairs, while individually comparatively weak, are cumulatively strong enough to hold the double helix together while at the same time allowing for their frequent breaking and reforming via enzymatic activity.

Codons as the fundamental element of genetic semantics Above the level of individual nucleotides, the DNA molecule is divided into codons - triplets of bases which each *represent* an amino-acid [Klug, 2012; Ridley, 2004]. This is the main syntactic level of genetics and the one where symbolic representation is implemented. Because we can build $4^3 = 64$ codons from the four nucleotides, but there are only 21 amino acids from which all life builds proteins [Ridley, 2004], this code is degenerate, with multiple codons encoding a single amino acid [Koonin and Novozhilov, 2017; Ridley, 2004]. Codons have some internal substructure, with the first two nucleotides being more specific to the amino acid it encodes, while the third generally is more variable. Moreover, substitution in the first nucleotide generally leads to the substitution of a related amino acid, which typically still leads to a functional protein [Ridley, 2004]. Secondly, not every amino acid is encoded by the same number of codons; some are represented more often. This structure of the code entails a great deal of redundancy and structural optimization. Indeed, it has been shown that the code shows significant error-minimization properties that are interpreted as the result of evolution [Koonin and Novozhilov, 2017]. While this seems to lend credence to an evolution of the genetic code via natural selection [Vetsigian et al., 2006], the possibility of it being “just” a frozen accident remains [Ridley, 2004], although these two alternative hypotheses need not be mutually exclusive. As with many questions in evolutionary biology, the answer might not be one or the other, with an initial codebase possibly emerging from high levels of horizontal transfer of elements of the syntactic and semantic machinery, which then are optimized once a code had been, at least locally, become fixed [C. R. Woese, 2002; Goldenfeld and C. Woese, 2011].

Genes as the second fundamental syntax layer of the genome The next syntactic level consists of genes, sequences of codons that represent a single sequence of amino acids that folds into a specific protein, or parts of them [Ridley, 2004]. While codons are made up of three base pairs that code for a single amino acid, a protein can be made up of several hundred amino acids or more [Ridley, 2004]. Genes have no internal syntactic element that separates one codon from another. Because the reading frame is only determined by the codon length of three bases each, the system is particularly vulnerable to mutations that delete or insert a nucleotide, since the entire reading frame from this point on is shifted [Stearns and Hoekstra, 2005; Ridley, 2004]. The genes themselves are, however, separated by stop codons. The amino acid sequences provide additional redundancy because not every amino acid is equally important for the final shape and function of the protein they are a part of. Consequently, some errors in codons that change an amino acid can be tolerated without loss of function of the encoded protein. Genes therefore can come in different variants which are located at the same point in the genome and encode proteins with the same general function, but with more or less small variations. These variants

---

[6] While this is true in general, C-G bonds are somewhat stronger than A-T bonds because they form three hydrogen bridges instead of two and are consequently more stable.

are called *alleles*. Often, population genetic models of evolution are concerned with the evolution of the frequencies of different alleles.

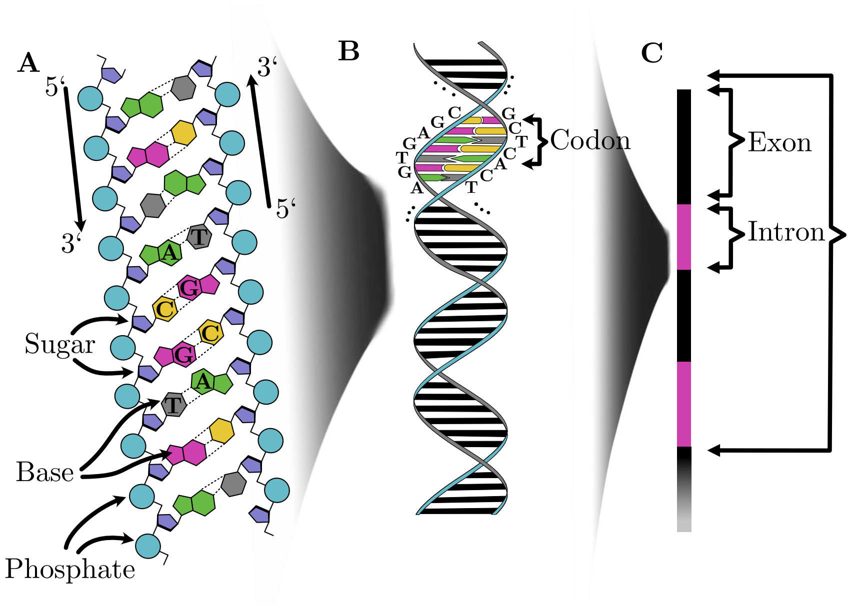


Figure 4.4: **A:** DNA is made up of a chain of sugar (Desoxyribose) -phosphate backbone to which one of four bases - Adenine, Thymine, Cytosine, and Guanine is bound. Together with their sugar-phosphate element, bases form nucleotides. In DNA, the bases form hydrogen bonds with their respective complementary base, Thymine for Adenine and Cytosine for Guanine. While Adenine and Thymine form two hydrogen bonds, cytosine, and guanine form three, which yields more stable bonding. The two strands of the DNA molecule are anti-parallel. **B:** While individually weak, the number of bonds formed bind the DNA strands stably together into a double-stranded helix structure. Strands form long macromolecules of often millions of base pairs. Within protein-coding regions, three bases form a codon that codes for an amino acid. This forms the primary syntactic structure of the genetic encoding system and implements the genetic code. **C:** Genes are subsequences that code for proteins or RNA. In eukaryotes, the coding regions (Exons), are interspersed with non-coding regions (Introns) which are cut out (spliced) during product synthesis. Aside from Introns, there are other kinds of non-coding DNA, e.g., transposable elements or dysfunctional virus genomes and more.

In prokaryotes, the genome is comparatively simply organized with genes being largely self-contained, while in eukaryotes, genes are not closed sequences of nucleotides that code for a gene in one unbroken chain but are divided into introns, elements that take part in syntactic dynamics but not in semantics, and exons, subsequences which take part in both the syntactic and semantic system [Klug, 2012]. This creates complications because it subdivides a gene into many units that can be put together in different ways. This "alternative splicing" occurs in eukaryotes [Klug, 2012] and endows the genetic system with additional degrees of flexibility, because one gene can be employed to code for a variety of related proteins, thus increasing the number of proteins that can be encoded for a given genome size at the prize of regulatory complexity. Therefore, eukaryotes possess an additional layer of syntax in their genomes that is absent in the genomes of prokaryotes, and which increases their encoding capabilities. Not all genes code for proteins, some indeed code for RNA which is a central element in the transcription and translation process [Ridley, 2004] (see below).

Non-coding DNA and chromosomal organization A large amount of the DNA of many organisms does not consist of genes at all, i.e., does not encode amino acid sequences, but can have different functions, not all of which are understood yet [Castillo-Davis, 2005]. For instance, some non-coding DNA plays a role in gene regulation, but other non-coding DNA contains ancient viral genomes or transposable elements which are either non-functional or are mostly thought to be DNA parasites. Importantly, non-coding DNA can be a source of novel phenotypic traits through exaptation [Fagundes et al., 2022]. In particular, it has been proposed that transposable elements play an important role in the evolution of gene regulatory networks [Feschotte, 2008]. In general, we consequently find that at least for more complex cases, the syntactic dynamics of the genetic system, operating on the molecular dynamics of DNA and RNA, implements a myriad of interactions that in their richness can be likened to the ecosystems formed by cellular life [Venner et al., 2009].

Finally, DNA is organized into multiple chromosomes in eukaryotes, while prokaryotes only have one circular chromosome and possibly a number of plasmids which contain smaller amounts of genetic material. Chromosomes are distinct large-scale DNA-protein complexes located in the cell's nucleus which contain the DNA of the cell in a folded-up form. This level of organization is missing in prokaryotes, where DNA is located within the cytoplasm and lacks the sophisticated folding and structuring that eukaryotes use.

While not strictly a syntactic element of the process, the chromosomal organization is nevertheless an important part of the transfer of heritable information, ultimately underlying Mendelian inheritance in diploid eukaryotes. Aside from the nucleic DNA in eukaryotic nuclei, their organelles have their own DNA, e.g., mitochondria. The genetic heritable information within a cell, therefore, is usually not fully concentrated within its chromosomes. The entirety of the DNA within a cell[7] is called its genome, while the set of proteins the organism can produce is called its proteome.

4.2.2 Syntactics – Genome replication

The syntactic processing system of DNA includes much more than the reading and copying of genetic information, e.g., proofreading or active error correction [Klug, 2012] which we only mention in passing here. It must be noted that the origin of DNA as information carrier and of its associated syntactic processes cannot be regarded separately. In a way, the origin of the syntactic part is a greater puzzle than the origin of the construction process, because it involves the emergence of DNA as information carrier, which, judging by the ubiquity of RNA in transcription and translation, was probably not the primordial state. Somewhere along the history of life, there must have been a switch from a likely RNA-based encoding mechanism to one based on DNA, which includes the emergence of the accompanying enzyme system that renders it functional. The situation is complicated by evidence that parts of the DNA replication system have been invented independently multiple times [Forterre et al., 2004]. We might place these origins deep in the pre-cellular stages of evolution which were likely dominated by high degrees of horizontal transfer [C. R. Woese, 2002]. It has also been theorized that the origin of the DNA-based encoding machinery lies with viruses [Forterre, 2006].

---

[7] Because all cells of multicellular organisms contain the same genetic information, this also applies to multicellular organisms

Furthermore, it must be noted that the process of DNA copying differs from the syntactical DNA transcription process during protein synthesis, owing to the former's parallel origin to the usage of DNA as information carrier, while transcription probably evolved in conjunction with translation (see subsection 4.2.3) which likely contains older, pre-DNA elements that have been coopted. This shows us that while it makes logical sense to distinguish between semantics and syntactics, their natural implementations within cells are intertwined and likely coevolved with each other. We will consider DNA replication here first because it is the natural analog to I in Equation 4.1.9 and implements inheritance in all cellular life. The process differs significantly between eukaryotic and prokaryotic cells, which reflects the more complex organization of heritable information in the former. More precisely, the eukaryotic and archaean systems for all DNA processing are much more closely related than each one is to the bacterial system. This, together with other differences in the semantic machinery hints at a minimum of two independent inceptions of the cellular architecture, with bacteria and archae (and by extension eukaryotes) constituting the two surviving lineages [de Farias et al., 2021].

A simplified overview of the prokaryotic DNA replication process For simplicity, we focus on the prokaryotic replication process. The process consists of three stages [Klug, 2012]: First, an initiator protein binds to a special site on the DNA strand - the replication origin. The enzyme "helicase" unwinds the DNA and breaks the hydrogen bonds between base pairs to separate the two strands, such that two replication forks form. In the next step, a second enzyme called primase attaches to each strand and assembles a short stretch of nucleotides that provides a molecular anchor. In the third stage, a DNA-Replicase enzyme attaches to this anchor and creates the new strand of DNA by attaching complementary nucleotides to the strand, thereby elongating the primer strand assembled by primase.

Because the two strands run anit-parallel and DNA replicase can only work in the $5'$ to $3'$ direction, only one strand, the so-called "leading strand", can be copied continuously. The other strand, the so-called "lagging strand", is assembled in steps that produce so-called "Okazaki fragments" [Klug, 2012] which later are joined together using different DNA polymerase versions and other enzymes to form the full copy. Many other components are involved in order to successfully replicate the DNA molecule. Binding proteins prevent the double helix from reforming after being broken up, and yet others are involved in holding the DNA polymerase in place during replication or in alleviating mechanical stresses that result from the breaking up of the double strands. The anti-parallel structure of the DNA molecule results in the assembly of the lagging strand being a lot more complex than the leading strand because it must be assembled from the Okazaki fragments. In many prokaryotes, there is just one DNA duplication origin and duplication proceeds along the circular genome in both directions from this origin, while in eukaryotes there are many origin sequences on each chromosome. Eukaryotes also make use of many more DNA polymerase enzymes, up to 14 as opposed to three, [Klug, 2012]. Eukaryotic DNA is additionally organized into linear sequences instead of circular molecules, which creates an additional complication for the lagging strand during replication because the Primase enzyme cannot attach an RNA primer to initiate replication toward the ending of the strand. Thus, the lagging strand's end remains unpaired. With repeated copying, the sequences at the end of the strand will grow shorter and shorter over time. While there are special repetitive noncoding sequences at the end of chromosomes, so-called telomeres, to prevent the loss of coding information [Klug, 2012]. Each of the resulting double-stranded copies contains an old strand and a new strand. DNA replication is, therefore, semiconservative [Klug, 2012], always retaining one template strand within the double-stranded final molecule. For material supply, the replication system relies upon the pool of free-floating base pairs and sugar-backbone molecules in the surrounding plasma.

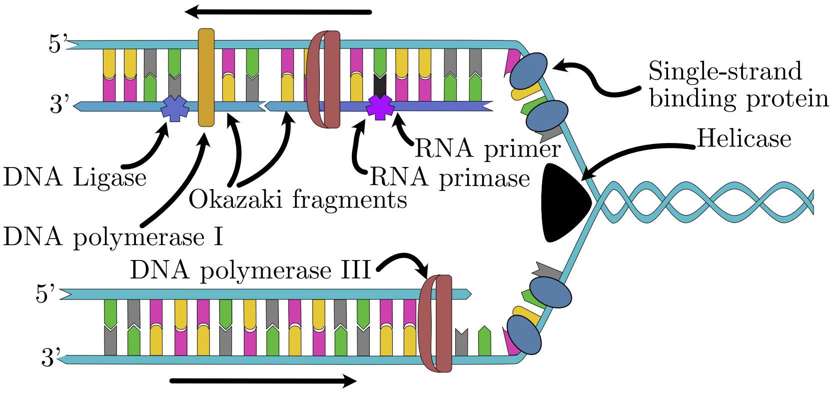


Figure 4.5: Schematic representation of prokaryote DNA replication. DNA replication is initiated at an origin point, special sequences of DNA where protein complexes are formed that can split the DNA strands into two replication forks. The responsible enzyme is called helicase, and single-strand binding proteins bind the individual strands to keep them from reforming the double helix. To initiate replication, a primase enzyme forms an RNA primer (note the black base representing uracil) complex on the single strand, to which DNA Polymerase III binds and synthesizes the complementary strand of DNA. Because this enzyme can only function in the $5'$ to $3'$ direction, one strand is synthesized continuously (lower, called “leading strand”) while the other (upper, “lagging strand”) must be synthesized in stretches, called “Okazaki fragments” [Okazaki et al., 1968]. Each stretch is newly initiated by a primer. After synthesis, DNA Polymerase I replaces the primer RNA with DNA, and DNA Ligase joins the Okazaki fragments together. In Eukaryotes, this process is much more complex with multiple origins of replication per chromosome, more complicated initiation procedures and more DNA polymerases involved [Snustad and Simmons, 2016]. After Ruiz [2022] with modifications.

DNA-Replicase has proofreading capabilities, i.e., it is able to detect errors in the nucleotide pairing and corrects them, which increases heritability. Such error correction is required in order to be able to encode the complex cellular machinery we observe today. Error correction thus was an important step on the way to more complex life. However, the required enzymes themselves must be genetically encoded which requires in itself thousands of bases because of their inherent complexity, well beyond the primordial error threshold. The resulting paradox is known as Eigen’s paradox [Eigen, 1971]. Different solutions have been put forward to this problem, e.g., by Eigen and Schuster [1977]. However, evidence has been obtained by studying real-world ribozymes, that mutation rates of RNA can be a lot lower than originally thought, which allows even primordial molecules to be able to grow to a length of several thousand base pairs, alleviating Eigen’s paradox [Kun et al., 2005]. These authors also point out that in an early RNA world, neutral or compensatory mutations can maintain a phenotype even when the exact sequence is not conserved, therefore maintaining fitness. We found in subsection 4.2.1 that redundancy in encoding can reduce the impact of mutations, an effect readily envisaged to play a role in primordial RNA and the evolution of the genetic code.

While error correction systems are highly efficient, they are not perfect, and different types of mutations can occur during the copy process. These are generally classified according to the level of organization they occur in. On the lowest level, point mutations can lead to the substitution, deletion or insertion of a base into the sequence. While substitution of a nucleotide is often of little consequence because of the redundant nature of the genetic code and the properties of protein folding, deletions or insertions, if they occur in a coding sequence, lead to a shift of the reading frame. Consequently, the entire amino acid sequence downstream of this mutation is changed or is rendered entirely nonsensical. Next are mutations that affect entire sequences. An example is slippage, where a random disconnect and reattachment of DNA replicase during copying and subsequent repair can result in the copied DNA missing a stretch of bases, or it being copied twice [Ridley, 2004]. We mention gene duplication as an especially interesting phenomenon, in which a complete gene is erroneously duplicated. Transposable elements, mentioned before in subsection 4.2.1 can copy themselves into other parts of the genome, which sometimes "picks up" neighboring stretches of DNA which then are duplicated too. Finally, in eukaryotes, many more mutations on the chromosome level can occur, which result in the translocation or duplication of genes or longer DNA stretches or their complete deletion, the inversion of a sequence, or the fusion or copying of entire chromosomes, or even whole genome duplications [Ridley, 2004]. Duplication events, or more generally all mutations that increase the length of the genome are thought to be an important generator for novel heritable information [Ohno, 2013; A. Wagner, 1994; Zhang, 2003] if they are such that fitness is maintained.

4.2.3 Semantics – A simplified overview of translation and protein folding

The construction process discussed here is concerned with translating the genetic nucleotide sequence into proteins. As mentioned, some DNA codes for RNA, which we will ignore here, however. The process is usually divided into two parts, *transcription*, and *translation*.

DNA-transcription as a purely syntactic process In the transcription process, the content of the DNA is transcribed to Messenger-RNA via an RNA-replicase enzyme. This is strictly a syntactic process, because the transcription system "knows nothing" about the correspondence of codons to amino acids, but processes only the sequence of bases. Transcription uses only one DNA strand and produces mRNA[8] that matches the unused DNA strand except for the thymine-uracil substitution. This transcription process is subject to a certain error rate and is therefore able to introduce phenotypic variation. However, this variation will generally not be heritable because it does not change the DNA itself.

As mentioned in subsection 4.2.1, in eukaryotic cells the transcription process involves splicing, i.e., the removal of introns from the primary genetic sequence, which can have multiple alternative outcomes. Additionally, other regulatory processes can intervene at this point and modify the pre-mRNA to prevent translation altogether or edit the sequence to create a different protein under some conditions. Syntactic preprocessing can therefore be complex in eukaryotic cells, which allows for more specific expression and a higher degree of differentiation without having to interfere with the base code in the DNA or

---

[8] Messenger RNA, used in the transcription process from DNA that is the first phase of gene translation to protein.

encode all the different possible alternative genes that can be created via alternative splicing into separate genes.

DNA-translation as the central semantic process The second stage of the construction process after transcription into mRNA is its translation into an amino-acid chain. This happens via specialized RNA-protein complexes called ribosomes, which catalyze the reaction of amino acids to form a polymeric chain. Amino acids are supplied by tRNA[9] molecules, which in turn are made up of two elements, one that is complementary to a codon and another one that can bind to the corresponding amino acid. Each tRNA can therefore only bind to a single amino acid. This stage is semantic, i.e., it is here where the genetic code is translated into its corresponding structure of material building blocks, i.e., amino acids. Moreover, the set of tRNAs defines the genetic code in that it determines which codons represent which amino acid.

In eukaryotes, the transcription process happens within the cell nucleus, while translation happens outside of the nucleus in the cell plasma. Prokaryotes of course do not have this distinction. RNA is the main active component that catalyzes the amino-acid links, and ribosomal RNA is highly conserved within archaea and bacteria, though has some recognizable variation between them. This provides another hint that the evolutionary transition to a cellular organization may have happened independently in both groups based on related ancestors. The eukaryotic translation system is much more similar to archaea than to bacteria, which hints at an ancestor of eukaryotes [Eme et al., 2017] in the archaea kingdom, which through repeated rounds of endosymbiosis with bacteria has given rise to eukaryotes [Zachar and Boza, 2020]. Indeed, tRNA is probably one of the most ancient elements of the genetic machinery, because it provides the translation element between genetic sequence and material implementation [F. H. Crick, 1958]. Thus, tRNA or a closely related molecule must have been involved in primordial translation precursors. It has been found that variants of tRNA can form replicator molecules that replicate binary nucleotide sequences with a fidelity of 85 % to 90% [Kühnlein et al., 2021]. Together with the older finding that RNA can act as a catalyst and not only as an encoding system [Kruger et al., 1982], this lends credence to the hypothesis that self-replication originated in some RNA-based system which subsequently must have diversified into a proto-biosphere we call RNA-world [Neveu et al., 2013].

Protein folding as low-level guided self-organization The third stage in the construction process of proteins from DNA is the reconfiguration of the amino acid sequence into a three-dimensional structure that is the functional protein. How a protein folds is to a large extent determined by its amino-acid sequence and the polypeptide chain's interaction with the aqueous environment of the cell, and thus protein folding represents indeed a process of self-organization. Rocha [2001] uses this as a guiding principle from which he derives the idea of evolution as guided self-organization. The protein folding process is, however, not always fully autonomous. Proteins can have many possible three-dimensional configurations, and often the functional structure is not easily accessible from the unfolded amino acid chain such that it either takes too long to fold into the desired structure or intermediate states are so unstable that the outcome becomes not determinate enough to yield reliable functionality. Cells employ special proteins called chaperones to stabilize unstable intermediates in order to allow further folding along the correct path by restricting the protein's possibilities or speed up otherwise folding that would be otherwise too slow

---

[9] transfer-RNA. Small RNA molecules that bind to an amino acid on one side and to a codon on the other. Important during the translation phase of gene expression. Implement the genetic code.

[Snustad and Simmons, 2016; Klug, 2012]. This provides evidence for the concept of heritable information providing the forcing for phenotypic development and reproduction on the lowest level of functional organization of gene products, and for the concept of reproduction as guided self-organization at least on some levels of biology.

Corollaries We now try to connect the properties of the DNA-based replication system with the theoretical insights discussed before. Any material substrate, like amino acids or RNA, has invariably physical properties that define possible physical interactions. If a self-replicating system uses this substrate in some way it gains access to this spectrum of physical interactions "for free", for example, the entire spectrum of chemical reactions proteins are capable of catalyzing, but also the possibly adhesive properties of some proteins, their ability to form large, stable structures and much more. The spectrum of physical interaction for a given class of physical building blocks is typically so large that it cannot be exhausted by a hypothetical biosphere within its possible lifetime, think for instance of the number of possible proteins that could be encoded in all possible genomes of size 5×10^9.

This sharpens the emerging picture of "life as guided self-organization" in that it shows how, from a basic self-replicating system that only uses a restricted subset of possible interactions of its substrate, a layered, self-organizing system of far greater capability can emerge by successively incorporating or coopting emergent structures or properties and the possible interactions with the surrounding environment these give access to. Within this system, not only does heritable information provide guidance and forcing for the self-organization of basic building blocks, but this guidance or force is transmitted indirectly from one layer of organization to the next by coopting independently existing physical interactions that the elements on a given organizational layer have anyways. More succinctly, an organism needs to encode in its heritable information structures that allow physical access to a given family of possible interactions, after which these can be coopted via natural selection for fitness-enhancing functions. The full specification of the entire range of interactions is not necessary. All this comprises just a more detailed account of the unfolding of complex constructive systems for the special case of a CCS based on self-replicating entities, which fills in some of the gaps left in the more general discussion with more mechanistic details. In that regard, it is worth mentioning that other self-replicating entities are always part of the environment with which organisms interact and to which evolutionary innovation provides access, which lends credence to the view that ecosystems, societies or multicellular bodies are essentially incarnations of the same underlying class of processes.

4.2.4 Limitations of information flow in the genetic system

The central dogma of molecular biology As described before, the DNA nucleotide sequence is transcribed into a sequence of RNA nucleotide sequences, which then is translated into an amino-acid chain that folds into a functional protein. Thus, biological information can be said to flow from DNA to RNA to protein. This scheme is called "the central dogma of molecular biology" [F. Crick, 1970]. Its main consequence is that information does not flow "backward", i.e., proteins do not, in turn, change the genetic sequence. In other words, there is no reverse translation from protein sequence into RNA sequence (or DNA sequence). This is one element in the genetic system that precludes the inheritance of acquired characteristics at least on the level of genetic sequence. Acquired characteristics like environmentally induced gene expression levels (see below), hinge on gene products, but while they do change the activity patterns of genes, they do not change the genetic sequence themselves. It has been suggested that this informational asymmetry hinges on the nature of protein folding as a self-organizing system that cannot be reliably reversed [Koonin, 2015]. That being said, its status as a “law” (or dogma) is subject to ongoing investigation [Koonin, 2012; Ille et al., 2022].

Weismann’s barrier in animals As mentioned before, Darwin’s concept of heredity involved pangenesis, i.e., the entire body contributed heritable information to the next generation [Holterhoff, 2014]. This theory was later refuted empirically [Ridley [2004] and Bonduriansky and T. Day [2014] and references therein]. In particular, Weissman’s concept of germline-soma separation [Weismann, 1892] in animals provided an explanation for why inheritance of acquired characteristics (i.e., changes to the somatic body) are not inherited by offspring in general, at least as far as could be known at the time. Rather, only mutations in the germ cells (eggs or sperm cells) are able to be inherited by the next generation genetically. It must be noted, however, that this does not apply to all animals (sponges, in particular, do not have differentiated tissues and also no clear germline [Fierro-Constaín et al., 2017].). Secondly, as described below, mechanisms have been found that show how the Weismann barrier can be somewhat porous and allow for the transfer of acquired characteristics from somatic to germ cells, in particular small RNA [Conine and Rando, 2021].

4.2.5 Genetic expression

Gene regulation refers to the manipulation of how genetic information is translated into phenotypic effects. This definition deliberately is so broad as to encompass all stages between genotype and phenotype, because regulation of gene expression can happen in most if not all of them [Klug, 2012], with many having been discovered only in the recent decades with the advent of better technology. Gene regulation is largely implemented by leveraging syntactic dynamics to control semantic outcomes, although there are mechanisms for regulating gene expression that happen after the translation into final products has occurred and hence can be regarded as strictly semantic [Klug, 2012; Sultan, 2015]. In discussing these mechanisms, we largely draw upon Klug [2012] and Sultan [2015] with a focus on the syntax-semantics paradigm laid out previously. Gene regulation is an important part of the implementation of the sub-processes L and $\tilde{D}_P$ of Equation 4.1.9, realizing therein a mechanism with which to react to environmental signals by adjusting the effect of one part of genetic information by leveraging other parts. In eukaryotes, gene regulation also is a key element in the development and differentiation of tissues in multicellular organisms [Gilbert et al., 2015], thus making it a key subject of evolutionary developmental biology. Indeed, we can understand the regulation of multicellularity as a highly evolved cooptation of environmental reactivity as it already is present in prokaryotic cells.

It must be noted that while we focus on genetic processes here, we should keep in mind that the genome functions in the context of a larger regulatory system that includes heritable elements on many spatiotemporal scales and in many forms. As we will discuss later in more depth, what is regarded as an environmental, and hence external, signal for genetic regulation is often part of a larger, heritable signaling system that includes genetic information, but also abiotic input from the environment, cultural content or input from symbionts or other organisms, and the effects of gene regulation can act, in turn, on these things. Therefore, gene regulation is only one element of L and $\tilde{D}_P$, albeit one that is singled out by its complexity and its fundamental nature to all life.

Gene regulation in prokaryotes Prokaryotes, being the primordial form of cellular life, are far simpler than their eukaryotic cousins, and so are discussed first. A variety of principles can be identified in them which we expect to find in eukaryotes in a refined form. First, we observe that cellular functions are rarely the product of a single protein or RNA. Rather, a function is usually achieved by the interaction of multiple gene products which create some structure or enable some functionality. A common example is the usage of lactose as a carbon source in *E. coli*. Normally, these bacteria prefer glucose whenever available. However, they can process other sugars to ultimately break them up into glucose if it is not available in raw form, albeit at a higher energetic cost. For lactose, this is mostly achieved by a combination of three proteins, which code for processing enzymes for lactose processing and for facilitating its transport into the cell [Klug, 2012]. All of these must be expressed simultaneously to allow for lactose metabolization. This functional interdependence makes common control possible, which in turn introduces a new syntactic element in prokaryotic (and some eukaryotic) genomes – the operon. An operon is a cluster of genes that together provide some function and are under common regulatory control.

We can think of it as a simple module of genetic heritable information. Operons generally consist of several functional genes preceded by an operator sequence which itself is not coding for a gene product. Rather, it provides binding regions for transcription factors. Transcription factors are proteins themselves [Klug, 2012] which can interact with the DNA to activate, deactivate, enhance or, suppress the transcription of DNA into mRNA. These transcription factors again are products of other genes and can interact with other such proteins or environmental signals (e.g., allolactose in the above example [Klug, 2012]). The regulatory region preceding the coding genes of the operon consists of a promoter sequence that allows for the attachment of RNA polymerase, while another part of it serves as an operator region that binds the actual transcription factors and allows or inhibits the transcription into mRNA.

Genes in prokaryotes are therefore networked into modules that are functionally linked to other genetic elements, thus generating a new level of syntactic structure within the genome. This organization, and its generally high specificity, explains in part the highly nonlinear effect of mutations. For example, mutations in the transcription factor gene or the operator region of the operon will render control ineffective and disable an entire functional module of the cell at once, or create binding sites for other transcription factors, thus rewiring the network. This rewiring allows for new combinations of gene products to be active within the cell, possibly facilitating new functionality without extensive rewriting of individual coding genes.

These networks are not governed by simple on/off switches responding to single inputs but can perform logical functions like `AND` or `OR`. For example, the *lac* operon of *E. coli* mentioned before is repressed normally by a specific inhibitor gene product, which undergoes a structural change that eliminates its functionality if it binds to lactose [Klug, 2012], and retains it if lactose is absent. However, if lactose `AND` glucose are available, it is more energetically favorable to prefer glucose because it “costs less” to metabolize, and so a logical switch between the two (or any number of available carbon sources) in favor of the energetically least expensive one would be favorable evolutionarily. Thus, the emergence of logical processing of environmental signals should be expected to be present in gene regulatory networks and to be universal across the tree of life. Biological regulatory networks, therefore, constitute processing systems for biological information *sensu* Jablonka (compare subsection 4.1.6).

Returning to the concurrent presence of lactose and glucose as food sources for *E. coli*, we consequently expect there to be a logical function that would have the structure of *glucose* `AND` *lactose rightarrow use glucose*, which, means that the normally activating effect lactose has on the *lac* operon must be secondarily suppressed if glucose is present at the same time. This secondary repression is achieved by inhibiting the action of catabolite-activating proteins (CAP). These normally raise the efficiency of RNA-polymerase binding[10]. Their effect depends on a secondary molecule with which it forms a complex binding to DNA whose production is inhibited by glucose presence. Thus, CAP action is inhibited in the presence of glucose, and the production of lactose-metabolizing enzymes is inhibited.

We note that the logic this system implements is not strict, but fuzzy when observed over time: The eventual transcription efficiency of the *lac* operon during some time interval is dependent on the concentration of lactose, glucose and other molecules, and the system does not yield a forever-on or forever-off switch. In other words, gene regulation can be both, analog or digital, with some regulatory systems reacting in one way and others in another [Lorberbaum and Barolo, 2013]. Moreover, transcription factors generally act combinatorially, such that the response of the regulation target is determined by the “cooperative” action of transcription factors [Stearns and Hoekstra, 2005].

It is instructive to observe that most of these regulatory effects are caused by changing the three-dimensional structure of DNA or signal molecules, which in turn facilitates or inhibits chemical binding. The structural flexibility of many biological molecules is also exploited in other ways to achieve gene regulation in prokaryotes. For instance, the process of attenuation exploits the temporal overlap of transcription and translation in prokaryotes to conditionally abort transcription based on translation. In *E. coli*, this happens in the operon containing genes for tryptophan synthesis, an amino acid. Here, several tryptophan codons are present in a separate DNA sequence preceding the coding genes (the leader sequence). Transcription of this sequence results in an mRNA with two mutually exclusive secondary structures. Which one is formed depends on the presence of tryptophan in the cell (and consequently the presence of tRNA loaded with tryptophan). If little tryptophan is present, the emerging secondary structure allows translation of the full mRNA, yielding tryptophan synthesizing enzymes, while in the presence of tryptophan, translation is stopped by the alternative mRNA structure, and no proteins are synthesized. More generally, metabolite sensing elements in leader sequences that change mRNA secondary structure are a common theme for gene regulation in prokaryotic cells, called Riboswitches. Finally, direct changes to the chemical state of the DNA are employed to regulate gene expression. The most widely discussed of these is arguably DNA methylation, in which methyl groups are attached to C- or A-bases and interfere with transcription [Seong et al., 2021]. Finally, the effect of gene products can be regulated after translation, too. However, because these mechanisms are more important for eukaryotes, we will postpone their discussion for now. This illustrates the complexity of gene regulation even in simple unicellular systems and shows that different control mechanisms can be implemented at different steps of the synthesis of DNA products, often concurrently. In other words, combinatorial control is achieved by not only modifying gene

---

[10] CAP plays a role in many other operons as well [Klug, 2012].

expression at a single step using different signals, but also by signaling at different steps of the translation process.

Gene regulation in eukaryotes In eukaryotes, the genetic organization is a lot more complex than in prokaryotes, and so are its regulation processes, although the principles stay largely the same, i.e., combinatorial control, (fuzzy) logic and temporally extended regulation all carry over from prokaryotic gene regulation. That being said, because of the symbiotic origin of eukaryotes, gene regulation must be one of the most evolved elements in their cellular organization because it ultimately provides the tools to ensure the cooperation (initially) and correct functioning (in more derived forms) of different organelles.

Unlike prokaryotic genomes which are organized into a singular genome, eukaryotic genomes are organized into multiple chromosomes which organize DNA into chromatin, where DNA wraps around histone proteins to form compact structures called nucleosomes. Therefore, there are more organizational levels of eukaryotic genetic information that can be manipulated in order to achieve fine-grained control over genetic expression. On the lowest level, DNA can be directly modified to change gene expression patterns.

It has been shown that gene methylation patterns are tissue-specific in multicellular organisms and that large unexpressed regions in eukaryotic genomes are often heavily methylated [Klug, 2012], although newer studies suggest that DNA methylation in eukaryotes may play a more complex and context-dependent role, being possibly important in both enhancement and repression of genes (Zhu et al. [2016] and references therein). DNA methylation can, for instance, facilitate the recruitment of chromatin remodeling complexes, leading to gene repression [Klug, 2012]. For example, histones can be modified to loosen their binding to DNA, making it more accessible to enzymes, transcription factors etc. Other modifications involve the repositioning or removal of histones to expose stretches of DNA.

On a lower level, gene expression is controlled by cis-regulatory elements which are located on the same chromosome as the genes they regulate. Three important such elements are promoters, which govern the initiation and base efficiency of transcription, enhancers and silencers, which regulate the achieved level of transcription [Klug, 2012]. A number of different, but specific, transcription factor proteins can bind to these sites and affect transcription efficiency, which again yields combinatorial control. Different signals that enhance or repress transcription act together to affect the final level of transcription [Stearns and Hoekstra, 2005; Klug, 2012], and this combinatorial control can be based on competition [Karreth et al., 2014].

Analogously to prokaryotes, transcription factors can be susceptible to additional intra- or extracellular signals such that they only become active after they undergo a conformational change mediated by another “signal molecule” [Klug, 2012]. In this way, transcription factor dynamics can implement time- and place-specific gene expression in multicellular development, or allow for the reaction of unicellular organisms to environmental states. Activators and repressors act in complex ways to modify transcription activity, for example by associating with general transcription factors necessary for transcription initiation, making binding of RNA-polymerase more or less efficient, by recruiting chromatin remodeling complexes to modify accessibility or by influencing the geometric shape of the DNA strand to manipulate binding efficiency [Klug, 2012]. Therefore, cis-regulatory regions can be thought of as signal interfaces through which genes are integrated into a self-regulatory network that reacts to external signals (see subsection 4.2.6). We understand that because the advent of eukaryotes precedes the emergence of complex multicellularity, these signaling systems would probably have been well developed by the time the first complex multicellular organisms arose, such that the cooption of these existing pathways provides a likely origin of today's developmental programs.

All of these regulatory actions depend on the DNA sequence to which transcription factors can attach. Therefore, gene regulation is achieved by manipulating the syntactic dynamics of the material encoding system that underlies genetics, as is the case in prokaryotes. It has been shown that, for instance, enhancer sequences can be reversed or moved to the vicinity of other genes, and they still retain their functions. Therefore, the eukaryotic regulatory elements discussed so far, while having semantic effects in that they determine what is translated when and where, are strictly syntactic in nature - their action is based on DNA nucleotide sequences, not on their protein translation. This stands in contrast to prokaryotes, where a temporal overlap between transcription and translation allows for a mixing of syntactic and semantic dynamics, see subsection 4.2.5.

Aside from directly manipulating transcription, other elements of the genetic machinery can be leveraged to implement gene regulation. For example, the average lifetime of an mRNA molecule determines how long it is available as a translation template and therefore can modulate the level of protein production. This half-life can be modulated by specific nucleotide sequences that facilitate attachment of degrading enzymes, thereby reducing the lifetime of the molecules [Klug, 2012]. Other elements of post-transcriptional regulation include the regulation of protein stability after transcription as a reaction to external signals, or RNA silencing. Indeed, RNA molecules have been shown to play an important role in gene regulation by silencing mRNAs in the nucleus or the cell plasma, or by directly inhibiting transcription via recruiting chromatin remodeling complexes after binding to DNA [Klug, 2012]. Much of eukaryotic genomes appear to be transcribed into small RNAs, such that the metaphor of "junk DNA" seems to be no longer accurate. These microRNA elements have been shown to regulate gene batteries in a tissue-specific manner [Klug, 2012]. Thus, much of DNA of eukaryotic organisms is devoted to regulatory functions instead of directly coding for proteins. Again, all these regulatory functions are purely based on the chemistry of DNA and RNA, and thus on the syntactic system of the cell.

Another important element of eukaryotic transcription is alternative splicing, as mentioned before. Alternative splicing can increase the size of the proteome of an organism, i.e., the set of proteins that can be built from the genome by an order of magnitude [Klug, 2012] or more. The downside is that while in prokaryotes only the gene or operon transcription itself has to be regulated, in eukaryotes, the ways in which proteins are spliced together must be specified additionally. Much like the regulation of transcription, splicing can be regulated by a variety of factors, from RNA secondary structure influenced by nucleotide sequence over specific proteins binding to the pre-mRNA to the specific composition of spliceosomes, RNA-protein complexes that catalyze the removal of introns. Because splicing adds another layer of syntactic structure on top of DNA regulatory networks that reacts in similar ways to external signals and is similarly structured, we can think of it as a second layer of regulation that controls the ambiguity of the primary level depending on incoming signals. In this way, the information encoded in the DNA does not just encode the primary structure of proteins but contains many elements that only exist to manipulate subsequent steps into a certain behavior, thus exerting indirect control over the final product as conceptually discussed before. Finally, DNA can be modified directly in some cells. For example, certain genes or DNA stretches can be duplicated to increase transcription activity in somatic cells, which contributes to cell identity and functional diversity. For example, in immunoglobulin production in mammals, programmed DNA recombination creates encodings for a variety of possible antibodies that by far exceed the number of genes in the genome [Klug, 2012].

In summary, eukaryotic gene expression is a very complex dynamical process and a highly active research field in which new phenomena are continuously being discovered. It implements a multi-layered network of interacting elements which for the most part act to regulate activity on the syntactic level of the genetic system through specialized semantic structures. Therefore, eukaryotic genomes, even more so than prokaryotic ones, can be seen as complex self-modifying systems in which multiple layers of organization interact to achieve global regulation in response to environmental cues or intracellular signals. Given that we identified prokaryotic gene regulation as a processing system for biological information, we find that its eukaryotic counterpart is a more sophisticated version of this system that integrates biological information on more levels of organization to achieve more complex behavior.

4.2.6 Genetic architecture

We discussed in the last subsection how different elements of genomes influence each other in order to create fine-tuned reactions of a cell's or a multicellular organism's phenotype to environmental conditions. We found that gene regulation constitutes a form of processing system for biological information that is based on sensing environmental conditions which then systematically precipitates gene activation, repression or activity adjustment. Multicellularity brings a new quality to these systems because regulation now must coordinate the interaction of large numbers of individual self-replicating entities which provide input signals to each other and form higher layers of organizational hierarchy, like tissues, organs and ultimately new forms of self-reproducing organisms. Although the same is true for eukaryotic cells with their organelles, the scale on which multicellular organisms achieve this is quite different, given that not only the body's own cells must be integrated, but a variety of symbionts or commencialists, too (see section 5.2 or Gilbert et al. [2015] and Chiu and Gilbert [2015] and references therein).

This subsection will concentrate on development in multicellular organisms with a special focus on animals, understanding that similar principles are at work in plants and fungi. The development at least of sexually reproducing multicellulars unfolds from an unicellular initial state which means that all cells of the developing organism have the same heritable information available to them [11].

Consequently, development into functioning phenotypes and intra-organismal cellular diversity is a consequence of the regulatory interactions of genomes with their environment. These regulatory mechanisms that have originated in unicellular life have been coopted over a long time to create the complex differentiated phenotypes we observe today. In subsection 4.2.5, we discussed low-level elements of gene regulation that integrate genes and their associated regulatory non-coding regions into functional elements. The *cis*-regulatory regions in these elements can be seen as a fundamental part of gene regulation because they provide interfaces for external regulatory input. The input signals for these regulatory elements are often other translation- or transcription products that act as

---

[11] There is asexual reproduction by fission in some multicellular complex organisms, which however normally recreates the body structure of the organisms irrespective of the place of fission. Therefore, the necessary complete heritable information still must be present in the cells of the regrowing organism, and regulation must be capable of recreating the complete phenotype from an incomplete starting point

sensors for environmental or intra-cellular states. In that way, individual genes can be thought of as being integrated into a genetic regulatory circuitry.

In many organisms, much of the genome has some regulatory function instead of coding for proteins. As discussed in subsection 4.2.5 and subsection 4.2.5, the genetic circuits underlying the regulation of gene expression perform logical operations conceptually not unlike the logical elements of programming languages, which is consistent with the interpretation of gene regulation as information processing. Research over the last two decades uncovered that gene regulatory networks in animals have a variety of structures that appear to be universal across kingdoms of life [Davidson, 2010; Erwin and Davidson, 2009; Davidson and M. S. Levine, 2008]. On a high organizational level, gene regulatory networks (GRNs) appear to be structured into structural motifs of differing stability. The most stable elements, called "kernels" [Erwin and Davidson, 2009] are those which perform the most fundamental function, like defining the basic structure of a body part, like its layout and symmetry. These appear to be highly conserved across evolutionary time, which has been hypothesized to be the reason for the stability of fundamental body plan organization in animals since at least the early Cambrian [Erwin and Davidson, 2009]. Below these, gene differentiation networks build upon the structures established by the kernels to further refine organ or tissue specialization, for example. Finally, differentiation gene batteries create the fine-grained differentiation of cells within tissues. The subnetworks responsible for each of these steps are of different complexity and depth, with the ones establishing the fundamental body structuring appearing to be deeper and more complicated than the ones that create the final cell differentiation in tissues or those which regulate cell cycles in response to metabolic needs or growth patterns [Davidson, 2010].

Individual network motifs can be connected by I/O-switches which control the activity of an entire subcircuit through their output [Davidson and M. S. Levine, 2008; Erwin and Davidson, 2009]. In this way, subnetworks can be redeployed in different contexts. The key to the evolutionary dynamics of these systems is the evolution of *cis*-regulatory modules [Peter and Davidson, 2017]. By changing DNA sequences within these modules, new binding sites for transcription factors can be created or lost, or their translocation to the vicinity of new genes, for instance via transposable elements, can change the wiring of regulatory circuits by reconnecting them to new network nodes. Therefore, the evolution of the linking of GRN elements can be identified as a major mode of evolution.

Similarly, mutations in the coding sequences of transcription factor genes can alter their binding affinity to regulatory regions, altering existing regulatory pathways or creating new ones. We observe that the redeployment of network modules in different contexts is similar to abstraction in mathematics or programming, which allows for problem-solving by recombination of elements that exist on a higher organizational level than the atomistic elements that make up the underlying syntax system. Because gene regulation is universal for all living beings, we find that an additional element has to be added to von Neumann's scheme if it is to be applied to biology, namely a syntactic processing system that reacts to environmental signals, i.e., a processing system that integrates multiple sources of biological information to determine and fine-tune the created phenotypic structures to the encountered environment. In Equation 4.1.9, this system makes up $\tilde{D}_P$ on the one hand, and on the other turns L into a feedback system between interactions with the environment and phenotypic reactions which occur as a consequence, and which in turn can influence future ecological interactions. These two parts are discussed as developmental– and phenotypic plasticity in the literature and will be revisited later in some more depth.

Returning to the notion of abstracted regulatory functions, their evolutionary advantages become readily apparent when considering them through the lens of software design: Abstraction allows for the redeployment of existing solutions to other data without having to rewrite each time and therefore allows for modularity which can be developed much faster and is more flexible than monolithic designs. During the evolution of animal development, the same principles seem to have emerged. Indeed, a set of network motifs has been uncovered which encode a function purely within their structure and not in the individual genes from which they are assembled, and which constitute building blocks from which known developmental GRNs are built up [Davidson, 2010]. For example, a common motif is the double-negative-gate, in which a gene coding for a repressor transcription factor is wired to a second repressor which in turn acts as a switch for some downstream regulatory circuit [Davidson and M. S. Levine, 2008; Davidson, 2010]. This motif can be used to activate the downstream circuit only when the input repressor is active, while under all other circumstances, the second repressor will be active and deactivate the controlled subnetwork. Others include positive feedback control or the implementation of logical functions like `AND`, which we encountered in the *lac*-operon of *E. coli* in another form. These elements appear to be wired together in an overlapping fashion [Davidson and M. S. Levine, 2008; Davidson, 2010] to create the complex and diverse gene regulatory networks we encounter in nature. More than this, it seems that each of these motifs is associated with a given function, much like statements in programming languages like `if-else` or `do... while...` are [Davidson, 2010]. Although the known set of these motifs cannot be regarded as complete [Davidson, 2010] and thus the existence of such a universal set of building blocks remains a hypothesis, the concept is intriguing. Assuming this hypothesis to hold, development would consist of a finite set of functions that need to be executed at different times and places in a developing body in order to create a functioning organism and which integrate ecological signals. These functions are implemented in a corresponding set of regulatory subcircuits which can be wired together to achieve a function. This hypothesis gains additional gravitas by the discovery that while the network motifs are found almost universally across the animal kingdom at least, they are often implemented using different genes internally [Davidson and M. S. Levine, 2008]. Once again, biological function is encoded in the arrangement of elements, not in the structure of the elements themselves. More fundamentally, the recurrent usage of a finite set of network motifs, implemented with different genes in different organisms, hints at the establishment of an additional layer of syntax analogous to the genetic code. As discussed above, the role of this layer of syntax is analogous to abstraction in human thinking or language. It allows for elements (genes or gene clusters) which at some very early point in their evolution have been either autonomous or served a specialized purpose, to be reorganized into larger-scale processing systems for biological information. As such, gene regulatory networks represent another instance of the emergence of a higher organizational level on top of another one by leveraging the emergent properties of this lower level. In this case, we can interpret the capability of *cis*-regulatory regions to manipulate gene expression dependent on input signals as this emergent capability.

In this sense, the architecture of developmental GRNs elucidated here adds to the mechanistic basis for an interpretation of the genome as a self-regulating system that reacts to environmental signals. Indeed, implicit in discussions of GRNs is always that they implement a dynamical system that can produce diverse outputs [DiFrisco and Jaeger, 2019] under different conditions without architectural change to the system itself, which is quite similar in principle to learning systems [R. A. Watson, G. P. Wagner, et al., 2014].

Through this insight, we recover information processing as an important factor in many biological systems on a variety of organizational scales.

Returning to inheritance, we find that while individual genes can be seen as “atoms” of inheritance at least for sexually reproducing organisms with their Mendelian dynamics [Mendel, 1866], the existence of GRNs shows that, generally speaking, their individual structure is only important insofar as they maintain their role in the regulation system they are a part of. Indeed, the hierarchical and modular nature of GRNs provide additional degrees of freedom for selection to act on, but at the same time, because they are built from modules with different degrees of conservation [Erwin and Davidson, 2009], focus variation into distinct elements of the genome. Others remain comparatively stable, thus constraining the phenotypic variation that can be exposed to selection in a given population. Knowledge about the architecture of genomes also helps us with interpreting the fossil record and the history of life, because architecture or GRNs correlates with body plan similarity [Erwin and Davidson, 2009]. Moreover, the organization of the genome into GRNs explains the apparent nonlinearity of mutations in DNA, because their effect depends on which part of a genetic element (*cis*-regulatory region, intron, exon etc) they affect, and what role the affected gene product plays in the regulation of other elements of the networks.

Gene regulatory networks beyond genomes and developmental symbiosis The importance of the regulatory function played by symbionts of a multicellular organism has come into focus in recent years [Gilbert et al., 2015]. In *C. elegans* the regulatory signal precipitating the formation of the body axis does not originate from a symbiotic bacterium *Gilbert2015*. In many organisms, symbionts provide vital phenotypic functions and are thus central to their evolutionary fitness. This tight interconnection between organismal functions and symbionts has prompted the interpretation of multicellular organisms as “holobionts”, i.e., as organisms that are not genetical individuals, but whose genetic heritable information is made up of their own genome and the genomes of all their symbiotic organisms [Chiu and Gilbert, 2015]. While many species found in the microbiome of humans and many other organisms do not occur outside of their bodies, and thus can be assumed to be vertically transmitted, which largely eliminates evolutionary conflict [Stearns and Hoekstra, 2005], it has been argued that while some are indeed symbionts, others are more likely commensialists or even slightly parasitic in nature [Douglas and Werren, 2016]. Thus, immune systems may have evolved to a significant extent as control mechanisms of these internal ecosystems. Irrespective of the ecological details, it is not surprising the microbiome of multicellular organisms is tightly integrated into the genetic regulatory system during development and beyond, while at the same time constituting a form of non-molecular epigenetic inheritance because many are vertically transmitted from parent to offspring. It appears likely that these symbiotic organisms have been free-living commensialists or even parasites, which have been coopted into symbionts by vertical inheritance and into regulatory nodes or phenotypic functions, coevolving with their hosts over evolutionary time. Thus, this provides an example of the evolutionary integration of ecological elements into self-replicators over time.

4.3 Heritable information beyond genes

As discussed before, the process of evolution does not refer to any particular material system of inheritance. Rather, it only requires the process of inheritance of *phenotypic* *functions* with variation to exist – the encoding system and if there is one or many of them is immaterial to its logic. Indeed, after the genome was the sole focus of heredity in evolutionary biology for the past century or so [E. Danchin, Pocheville, et al., 2018; Sultan, 2015], evidence has accumulated on many fronts that genetic information is not the only kind of heritable information [É. Danchin et al., 2011; Whiten, 2017; Gilbert et al., 2015]. Following the etymology of the word, we call all these inheritance systems *epi*genetic and discuss them together under this heading, although often the term is applied more narrowly to the inheritance of intracellular particles through the germline, e.g., microRNAs. The goal is not to give an exhaustive review of the scientific state of the art, nor a complete list of mechanisms over all kingdoms of life. Rather, this section is intended as an overview that helps us broaden our understanding of the evolutionary process.

4.3.1 Molecular epigenetics

Molecular epigenetics refers to the inheritance of gene activation patterns without changes to the DNA sequence. Because gene activation patterns are often the consequence of environmental signals, this constitutes a form of “inheritance of acquired characteristics”. This form of epigenetic inheritance involves the inheritance of molecular components of the machinery described in subsection 4.2.5. Because gene regulation is a complex, interconnected network of processes, molecular epigenetic inheritance is diverse, involving a variety of molecular components for many of which the precise role and transmission mode is still unclear [E. Danchin, Pocheville, et al., 2018; Sultan, 2015].

Molecular epigenetic inheritance has been found in many organisms to date. For example, the nematode *C. elegans* shows epigenetic transmission of physiological traits that persist for up to eighty generations, even when the originally inducing ecological factors are no longer present [Wang et al., 2017; Minkina and C. P. Hunter, 2018]. The flowering plant *Linearia vulgaris* shows two distinct flower phenotypes associated with changes in methylation pattern in a single gene which can be transmitted over generations [Klug [2012] and E. Danchin, Pocheville, et al. [2018] and references therein]. Indeed, in many cases, the material basis of molecular epigenetic inheritance has been tracked to small non-coding RNA, which’s role in gene regulation we discussed in subsection 4.2.5. As explained there, these can bind to DNA and bring about the modification of histones, chromatin remodeling or DNA methylation, which in turn silences stretches of DNA [E. Danchin, Pocheville, et al. [2018], Wang et al. [2017], and Skinner and Nilsson [2021] and references therein]. DNA regulation patterns, at least in animals, therefore are not directly inherited but are recreated because their generating factors are inherited. Among others, it has been shown that in some organisms, molecular epigenetic inheritance can breach the Weissmann barrier [Nilsson et al., 2020], i.e., somatic changes in biological information can enter the germline, which constitutes a mode of inheritance of acquired characteristics [Skinner and Nilsson, 2021].

Of potentially great importance is the fact that chromatin structure and DNA methylation significantly influence mutation rates of the affected DNA stretches. 5-methylcytosin can spontaneously decay into thymine which additionally is less likely to be repaired. In total, this can increase the rate of point mutation by a factor of 10^4 [E. Danchin, Pocheville, et al., 2018]. When patterns of methylation or chromatin modification are inherited through RNA transmission, the affected DNA will mutate faster and consequently create more phenotypic variation [E. Danchin, Pocheville, et al., 2018]. These authors generalize this concept and derive the hypothesis that molecular epigenetic inheritance and correlated changes in mutation rate provide a mechanistic basis for genetic assimilation without DNA sequence changes [E. Danchin, Pocheville, et al., 2018; Skinner and Nilsson, 2021]. Therefore, molecular epigenetic inheritance constitutes a mechanism for the inheritance of phenotypic plasticity (see chapter 5), thereby being a potentially important generator of phenotypic variation on which natural selection can act, and which either through classical[12] or epigenetically mediated genetic assimilation can ultimately drive genetic evolution.

Additionally, the mutation rate of epigenetic heritable information is up roughly five orders of magnitude higher than the genetic mutation rate [Skinner and Nilsson, 2021], so adaptation to short-term environmental fluctuations can be greatly facilitated by epigenetically mediated phenotypic variation. This concept extends the modern synthesis in that it acknowledges a directed influence of phenotypic expression on the evolvability of specific traits. While not directly violating the tenet of the occurrence of mutations being unrelated to their effect, it still extends the way in which phenotypic variation is understood to be generated with a feedback loop between phenotypic structure and genetic information. As *Danchin2018* points out, more theoretical and empirical work is required to elucidate its impact, however. Given the existing empirical evidence as reviewed in E. Danchin, Pocheville, et al. [2018] and Skinner and Nilsson [2021] however, it is safe to say that environmentally induced epigenetic inheritance is a significant element in evolutionary dynamics and can act as a dynamical link between ecological–, cultural– and genetic inheritance. Furthermore, it has the potential to link together ecological and evolutionary timescales because the epigenome is a lot more prone to mutations than the genome itself and thus provides a potential way in which ecological variation can lead to phenotypic variation without changes to DNA sequence. Consequently, epigenetics provides the natural motivation for the influence of L on h in Equation 4.1.9.

Continuing this line of thought, we can interpret molecular epigenetic inheritance as an emergent, secondary coding system for heritable information that emerges as a consequence of the organization of another level of heritable information, i.e., the primary and secondary structure of the DNA. We can hypothesize that it came into existence rather early in the history of life as a byproduct of the primordial evolution of gene regulation. Elements of it will later have been coopted because it resulted in phenotypic variation on which natural selection can act. The two encoding systems interact with each other by virtue of modified mutation rates of DNA which is altered by methylation or histone modification and via the influence of epigenetic traits on phenotypic traits, in turn influencing the action of natural selection. Because both are inherited via the germline and thus through the same transmission process, and because they influence the same phenotype, we cannot call it a secondary *inheritance system*, however, but a secondary encoding of heritable information that utilizes different degrees of freedom of the syntactic part of the genetic inheritance system. As such, it represents another example of organizational layers that emerge in the evolutionary process and are subsequently integrated into it.

---

[12] Classical genetic assimilation rests on the idea that phenotypic plasticity creates selection on cryptic genetic variation which can lead to the genetic fixation of the otherwise environmentally induced phenotype. Compare the so-called “Baldwin effect” for the related idea based on learning and behavior.

4.3.2 Ecological and cultural inheritance

Ecological inheritance Ecological inheritance refers to the transmission of elements of the ecosystem with which organisms interact to their offspring. Drawing on the picture of an ecological niche presented in Figure 3.1, we understand it as the inheritance of elements of the ecological niche. Its relevance, therefore, lies in the fact that natural selection, because it emerges from ecological interactions, is dependent on the ecological niches of organisms. Thus, ecological inheritance is the inheritance of patterns of ecological interactions, and therefore of patterns of selection [K. Laland, Matthews, et al., 2016]. The ecological niche of an organism does contain heritable information as defined in subsection 4.1.6 in various ways. As seen before, the environment is an important source of regulatory signals for development. On the other hand, ecological states can be actively changed by organisms to suit their needs, establishing a feedback loop that influences selection regimes and makes them partially heritable (more on this in chapter 6).

Ecological states can be bequeathed to offspring in multiple ways. As said, many organisms actively modify their environment in order to facilitate their survival and reproduction. Birds build nests, beavers create damns and ponds, a large variety of organisms both marine and terrestrial create burrows and tunnels and thus modify sediment porosity and ventilation, humidity and nutrient distribution and more. These actively constructed artifacts often persist for long times, frequently longer than the lifetime of the originator. In such cases, the changed environment is inherited by direct offspring or, more generally, the next generation of the same species, but usually influences other organisms as well. Indeed, for many costly environmental modifications, organisms try to monopolize their benefits by defending them from rival users of either the same or different species. The degree of monopolization has been theoretically shown to be crucial for the maintenance of costly constructions [Krakauer et al., 2009]. From this, we see that out of all the inheritance systems we encountered so far, ecological inheritance is probably the "softest", i.e., it lacks a dedicated encoding system analogous to genomes or language (see below), and it often requires an additional behavioral change to monopolize its heritable biological information. Indeed, the origin of such behavior lies in the effect of other heritable information, like genetics [Dawkins, 2016]. Genetically heritable environmental modifications have been recognized in the context of the gene-centered theory of evolution and have been called the "extended phenotype" [Dawkins, 2016].

Because these organismal modifications influence development and natural selection, we can see ecological inheritance as another example of the process where organismal traits, themselves products of the guided self-organization machinery that makes up organismal development, leads to the generation of new forms of heritable information by changing the structure of ecological variables.

Because, in the present case, these variables are by default accessible to the entire ecosystem, environmental modifications have typically effects on more than the originator species. Indeed organismal influence often changes the environment in many ways at once, in some cases so profoundly that the ecosystem functioning as a whole comes to rely on the action of one or a few species, which appropriately are named "ecosystem-engineers" [Jones et al., 1994]. Such species can be imagined as biotic hubs in the ecosystem network, which makes them so-called "keystone species" [Paine, 1969; Cottee-Jones and Whittaker, 2012]. Arguably, our own species is the most impactful ecosystem engineer of all, given the large and diverse transformations we exert on the entire system earth [Steffen et al., 2007]. Indeed, "ecological spillover" [Erwin, 2008] does modify flows and storage of energy and matter in ecosystems which can persist over geological times and modify the way organisms interact with their environment over timescales which are relevant for macroevolution. Therefore, ecological inheritance, because it is not strictly bound to an organismal inheritance system and therefore affects many organisms at once, operates most impactfully on the level of communities of populations which exploit a set of overlapping ecological niches, i.e., it transcends the limits of the lineage.

Not surprisingly, such ecological effects played an important role in the evolutionary history of life [Erwin, 2008] by adding ecological interactions to the repertoire of the biosphere. Turning this around and taking into account the network view of ecosystems, we also understand how individual extinctions of species can have vastly varying effects on the rest of the ecosystem, from hardly any to precipitating a wave of secondary extinctions, which also shows how mass extinctions ultimately unfold. We note that all this applies not only to active and adaptive ecological changes but also to the passive effects of organismal action. Organisms process their nutrition into different forms, thereby transforming matter and energy into often more accessible variants. It has been shown for instance that the recent extinction of the Pleistocene megafauna had a substantial effect on the nutrient transport in the global ecosystem [Doughty et al., 2015]. Similarly, water plants trap sediments in rivers, thereby creating meandering riverbeds [Corenblit, Steiger, et al., 2007] fixating sand dunes, such that organismal action can be said to form landscapes on the largest scale. Other passive ecological modification effects include the very bodies of organisms, e.g., treetrunks or –crowns with the myriad of arboreal species that are dependent on them.

Finally, the transmission of ecological heritable information can happen, much like molecular epigenetic information, either by biased recreation of an environmental state, which relies on other kinds of heritable information to bring about the behavior that leads to its construction, or by the sharing of modified environments between different generations. Recreation of environmental states, if encoded genetically, is represented by the aforementioned concept of the “extended phenotype” [Dawkins, 2016], but other encoding systems or forms of inheritance, especially cultural (see below) are possible. Sharing the same environmental state as the parent generation can come about by parental imprinting, where organisms tend to share habitat preferences with their parents [É. Danchin et al., 2011] at different points in their lifecycle. Another possibility is that offspring stay in the environment they were brought up in or disperse to nearby habitats not very different from the parental one, such that correlations between parental and offspring habitats are strong. As mentioned, the interaction of heritable information and development can have important consequences in that the inherited, and partly constructed, developmental environment provides signals that precipitate regulatory responses in the developing organism and in this way shape phenotypic variation, which in turn influences how ecological modification will proceed in the next round. This process is, theoretically, capable of shaping phenotypic variation over generations without the action of natural selection or genetic drift, indeed, without changes to genetic information at all, thus providing an avenue for the creation of adaptations that is not dependent on genetic changes.

Cultural inheritance Cultural inheritance is the final form of inheritance we shall discuss because it extends the farthest away from what is normally considered evolutionary biology. Cultural inheritance relies on a distinct transmission system - social learning. Social learning means that behavior is acquired from other organisms by observing their actions and the ensuing consequences [Whiten, 2019]. Note that this definition of social learning does not require the involved organisms to be related or from the same species, the learner must just be able to recognize the “other’s” behavior and its consequences. We can therefore understand social learning as building on top of the capability of individual learning. Individual learning can be thought of as a form of phenotypic plasticity (compare chapter 5), which involves the modification of an internal model of aspects of the ecosystem an organism lives in using its experiences. This allows predictions about the behavior of these aspects in the future. We recognize that these capabilities are quite sophisticated and highly complex, but emerge ultimately as part of a developmental process that is based on environmentally informed genetic regulation. We consequently recognize the ability to learn and the underlying existence of internal models for ecosystem elements as a new organizational layer in phenotypes that processes biological information. However, the physical basis of this processing system lies in the structure of the nervous system instead of the genetic content. Indeed, learning systems constitute a CCS that emerges within the evolutionary CCS normally recognized as biological.

When social learning occurs between generations, the traits transmitted in this way can constitute traditions [Whiten, 2019], which in turn gives rise to cultural inheritance. We follow the definition of culture given in Whiten [2019] and reference therein, where “culture” is defined as “the totality of traditions characterizing a community”. This culture can be cumulative via successive addition and refinement of traditions [Whiten, 2019]. Social learning as a transmission mode can be based on communication or observation, both of which can be associated with “transmission errors” and hence yield imperfect inheritance. It must be noted that while social learning through observation has been demonstrated in a wide variety of animals ranging from cockroaches and other insects over fish to ungulates, whales and primates [É. Danchin et al., 2011; Whiten, 2019], communication-based learning appears to be rare, possibly even limited to *Homo Sapiens* in its full form.

Indeed, while not as easily formalized, the question about cumulative culture leads to a problem analogous to the “error threshold” in genetics. Cultural traits must be such that they can be memorized and transmitted reliably via social learning. If they are too complex to be either memorized or observed in sufficient detail (a problem paralleling the template- versus code-based reproduction system discussed in Figure 4.1.7), the trait will be lost eventually because it cannot be transmitted without errors that will render the trait non-usable eventually. Drawing on our own experience in learning and memorization, we might suggest that cultural traits that are used often are transmitted with higher fidelity than those that are used rarely as a consequence of the nervous system’s plasticity. Language appears to parallel the genetic code system to some extent by providing the aforementioned syntax-semantics distinction that makes it possible to communicate cultural traits in detail and repeatedly. This alleviates the problem of observation-, i.e., template-based reconstruction of cultural traits via observational social learning and consequently opens the door to a much richer and far more complex cultural phenomenology which might well have been the central innovation of *Homo Sapiens* when compared to other species of *Homo* or other animals in general. Therefore, the invention of symbolic communication can be seen as a major evolutionary innovation that opened a new domain of biological information for the partaking species and thus more opportunities for adaptation. However, in contrast to the genetic system, our linguistic system has no universal code, with languages differing substantially in their grammatical structure or the number of employed sounds. However, while social learning (with transmission errors) in whatever way is a necessary condition for cultural evolution, it is not sufficient [É. Danchin et al., 2011]. Rather, it must be demonstrated that variation in cultural traits leads to variation in the reproductive success of their carriers and that these traits are transmitted through social learning. While intra-generational social learning has been found in many species there are comparatively few instances where this has been demonstrated. For songbirds, it has been suggested, though not conclusively proven, with sexual selection remaining a valid alternative [H. Williams, I. I. Levin, et al. [2013], according to Whiten [2019]]. Finally, cultural traits can influence natural selection on traits encoded in other systems of heritable information, mostly genetic. This gene-culture coevolution appears highly plausible but has mostly been demonstrated only for cultural traits shaping genetic selection, the classical example in humans being adult lactose tolerance [Ségurel and Bon, 2017]. It must be noted that other organisms from which can be learned are part of the ecosystem of the focal organism, and thus social learning over multiple generations can be seen as a form of ecological inheritance. However, it concurrently constitutes a form of horizontal transfer of biological information, which can be seen as a source of cultural innovation much in the same way as horizontal gene transfer is a source of genetic innovation for bacteria. The transmission mode is the same in both cases, i.e., social learning.

Cultural inheritance can form the basis for a new form of complex constructive system, in which imperfect social learning or individual cognitive processes constitute a trait discovery process, trial-and-error in a given ecosystem constitutes a performance evaluation process and the collective memory of the population forms the trait-reservoir. Depending on how well-developed the syntax-semantic-split is in such a system, how large the population is and how complex the underlying learning systems of the organisms in question are, this system will be more or less powerful, as discussed in the context of genetics before. This system can be intertwined with natural selection and classic biological evolution by sharing parts of the performance evaluation system for example, e.g., if an organism dies due to cultural maladaptation, which will lead to the aforementioned coevolution.

4.4 Summary

We started with an abstract concept of self-reproduction as a network of dynamical systems (Figure 4.2), modeled on observations in the natural world. Aimed to be general, it separates the initial creation of a functional phenotype $\tilde{D}_P$ from its maturation into a reproductively viable from, L. This is based on the observation that organisms must establish the necessary preconditions for reproduction first, e.g., acquire material from the environment, but often, physiological changes are part of this lifecycle, too. This leads to the insight that this process always transforms the environment, i.e., the surrounding ecosystem, as well. This process includes all organismal actions, not only foraging. Therefore, the phenotype and the local environment undergo a transformation process from a state that is not amenable to reproduction to one that is. We understood the creation of phenotypes from some material as a form of self-organization of matter, during which heritable information plays a central role in that it restricts and steers the self-organizing processes into forms that implement specific self-reproduction again. While we mostly envisaged this self-organization as operating on pre-existing material building blocks, the same principle applies to elements that exist only as part of biological processes.

We defined heritable information as a special form of biological information, adopted from Jablonka [2002], that is transmitted within lineages. Notably, this excludes horizontal exchange by choice, although, in nature, there is significant overlap between horizontally and vertically transmitted biological information, e.g., in horizontal gene transfer or cultural evolution. This does not limit the usefulness of the concept, however, because these processes just as well can act on heritable information. They just do not constitute inheritance. In keeping with the structure of Darwin's theory of natural selection, the exact physical nature of this heritable information is immaterial.

Heavily building on concepts discussed by Rocha [2001], we extend these by assigning heritable information an active role in the self-reproduction process, which includes the exertion of dynamical forcing, but also the possibility of it being modified during lifetime. Consequently, heritable information is not the central blueprint, but a system of, mostly indirectly, guiding functions for the dynamics of other parts of the organismal unfolding. Asking for possible implementations of this system, we discussed template-based and code-based self-reproduction, following the concept of von Neumann [1966] as discussed in Rocha [2001] and Pattee [2012b], and arriving at the insight that template-based reproduction is limited to comparatively simple systems because it must provide encoding of heritable information and physical function in the same structure, whereas the possibilities for a code-based replicator are only limited by the physical interactions its material building blocks are capable of.

This fills with life the assertion of heritable information as a guiding and regulatory influence on self-organizing matter because the autonomous self-organization of, say, amino-acid chains into folded proteins can be said to be the first layer of the organizational network of natural phenotypes. It follows that heritable information can be assessed according to different properties - stability or heritability across generations and the syntactic-semantics split, i.e., the existence of a symbolic encoding system and the ability to monopolize its content by a lineage the most important ones. We then discussed different forms of heritable information, starting with the genetic system for which the syntax-semantics split is most well-developed and which generally is the most stable form of heritable information. Its activity and dynamics are almost entirely controlled on the syntactic level, except in some prokaryotic organisms in which transcription and translation overlap in time. The complex chemical processes involved in the regulation of gene activity first lead to molecular epigenetic inheritance, i.e., the inheritance of gene-activity patterns over generations. This shows how a different inheritance system, lacking independent symbolic encoding in this case, can emerge on top of the dynamics of another one by employing degrees of freedom that arise from the syntactic dynamics of its underlying system.

A similar pattern has been discussed with respect to the architecture of genomes into gene-regulatory networks, where the hypothesis of Davidson [2010] has been discussed that there is a limited number of abstract network motifs, implemented by different genes in different lineages, which are used to build up complex, deep GRNs. This would create a second system of syntactic dynamics on top of the genetic one, which has been likened to abstractions employed in the design of programming languages. The central insight from all this, backed up by the finding that in many organisms, regulatory elements numerically far outweigh protein-coding genes, is that the genome cannot be regarded as a blueprint containing independent instructions to build specific phenotypic properties. Rather, it is a self-regulating system that evolved to dynamically react to environmental inputs in order to enable the self-reproduction of complex structures.

The last two forms of heritable information discussed here where cultural inheritance, which while not a new concept, only relatively recently received significant attention [Whiten, 2019; É. Danchin et al., 2011], and ecological inheritance, which has found a central place in a contemporary debate about whether evolutionary theory needs a fundamental revision [K. Laland, Matthews, et al., 2016; Müller, 2007; Scott-Phillips et al., 2014]. Ecological inheritance refers to the concept that the structure of the ecosystem in which an organism finds itself is a carrier of heritable information. This can happen by direct construction of artifacts, e.g., nests, burrows, and dams, that exist longer than the organisms that constructed them originally and thus can be seen as being bequeathed to the next generation. Indeed, costly constructions like beaver damns are preferentially re-used by direct offspring and are defended against conspecifics and other organisms. This behavior can be linked to genetic heritable information, in which case we once more find an emergent inheritance system that is based on the output of the underlying genetic one, although local variation in ecological construction can also be based on cultural traits (see below or subsection 4.3.2) or abiotic environmental conditions. Ecological inheritance depends on limited or directed dispersal, i.e., offspring either stay in the habitat of their birth or only disperse to similar neighboring habitats, such that they “inherit” the ecological conditions their parent chose. In many cases, parental imprinting plays a role [E. Danchin, Pocheville, et al., 2018; É. Danchin et al., 2011], such that offspring as adults develop a preference for the habitat of their parents. Ecological heritable information is, in general, difficult to monopolize for a single lineage. Indeed, the arising inter-lineage or population-wide ecological spillover effects may be the more important influence it has on the evolutionary process, because organisms, through their passive and active actions, change diverse elements of their environment for themselves, but also all others that interact with it, thereby creating ecological niches, changing matter and energy fluxes and storage and influencing the long-term trajectory of system earth as a whole [Erwin, 2008]. Ecological inheritance, consequently, is of more general scope than the other inheritance systems discussed here.

Cultural inheritance can be seen as a subprocess of ecological inheritance because the ecosystem an organism is a part of contains others through which cultural interaction and social learning can happen. The term “cultural evolution”, in its narrow sense, means that variation in cultural traits leads to their differential transmission to other carriers which results in the spread or reduction in the frequency of a cultural trait in a population.

Therefore, this theory of cultural evolution results from a direct application of Darwin’s theory of natural selection to cultural content [Whiten [2017] and Whiten [2019] and references therein]. The fundamental difference that distinguishes cultural inheritance from virtually all other inheritance systems is that the transmission of cultural traits proceeds via social learning and not by phenotypic reproduction as is the case with the genome and epigenome. Therefore, cultural evolution in that sense can be seen as proceeding largely independently from non-cultural evolution in that cultural traits. It can produce an independent kind of CCS analogously to biological evolution, which nevertheless can be intertwined closely with biological evolution. Along with epigenetic inheritance, it is one of the ways in which the unfolding of L over time can influence heritable information h, which allows for the inheritance of acquired characteristics on a cultural as well as epigenetic level. Cultural inheritance stands out in that it represents an inheritance system that emerges from another complex phenotypic trait, nervous systems, i.e., as an emergent property of the regulatory unfolding of underlying genotypic information, which at the same time creates its transmission- and encoding system. We can draw an analogy in social learning based on observation and copying to template-based reproduction as discussed in Figure 4.3, while social learning based on communication can be likened to the syntax-semantics split Rocha [2001] professes and which underlies von Neumann's universal constructor. Human language is arguably the most sophisticated example, giving human cumulative culture a qualitatively different character from all other animals.

5. Evolutionary developmental biology

5.1 Basics of Development

Developmental processes evolve in multicellular organisms All multicellular organisms that produce sexually, and some asexual ones, pass through a unicellular stage during their lifecycles [Grosberg and Strathmann, 1998; Du et al., 2015]. A multicellular lifecycle therefore rarely consists of a single, mature phase that can reproduce immediately. Rather, organisms develop in a complex, temporally extended process from their single-celled form into a reproductively viable state which, for some organisms, can make up the majority of their lifetime. We call the process that generates an adult body from the unicellular initial state "development". Taking up the notation introduced in Equation 4.1.5, this corresponds to the functions L and $\tilde{D}_P$. We note that this divides development into two phases, one that creates a functional multicellular body from heritable information and environmental resources that is capable of interacting with its environment (corresponding to $\tilde{D}_P$) and a second phase (corresponding to L) that transforms this structure into a stage that is reproductively viable. This latter phase includes invariably influencing the environment in which an organism exists.

While this clear distinction can be helpful as a theoretical guide, it must be noted that, in nature, it is rarely as clear-cut as this model posits. Growth and differentiation can continue long after an organism reaches a reproductively viable state, and in many complex animals, the ability to interact with the environment is gradually attained over a prolonged period of time. Thus, the process $\tilde{D}_P$ does not just create a physical structure, it creates a system that is capable of implementing a specific L, i.e., a specific growth- and development cycle. Heritable information thus must contain the necessary signal processing systems to guide this temporally extended process. Development is consequently subject to evolution, and many different strategies for implementing L and $\tilde{D}_P$ have evolved in different organismal lineages. While we immediately think about the quality of the phenotypic traits, it must be emphasized that adaptation just as much refers to the timing of their emergence and the order in which they do so in relation to environmental conditions or internal states. Therefore, all evolutionary change in multicellulars involve development in some form [Stearns and Hoekstra, 2005].

Development has ecological consequences Because organisms develop over time, their interactions with their environment change over time, too, as part of the growth and unfolding from zygote to juvenile to adult. This entails changes in the ecological niches they occupy. As a consequence of different phenotypic traits, the environment organisms seek out at different stages of their development can differ substantially from each other. Thus, the function L does entail complex shifts and reorganization of organismal features for many multicellular organisms, with consequences for the selection pressures active at any time.

In insects that undergo metamorphosis, this is probably the most apparent. Dragonflies are aquatic predators as larvae, but aerial predators as adults. Similarly, the larvae of most amphibians are fully aquatic, while the adults are largely or fully terrestrial. An even more striking example comes from the fossil record of dinosaur ecosystems, which often appear to be less species-rich than modern ecosystems, a fact attributed to them growing through different ecological niches associated with different body sizes [Schroeder et al., 2021] during development. This, in turn, is a consequence of size restrictions of size at birth because eggs have an absolute size limit.

Such *onthogenetic niche shifts* are common in nature [Nakazawa, 2014] not only for animals, but also for plants [Dayrell et al., 2018], and are doubtlessly a result of natural selection acting to improve the survival chances of organisms during non-reproductive stages leading up to the adult phase, or the general preconditions to become a successful reproducer. As mentioned, natural selection acts on each stage of development in possibly different ways, with the performance evaluation process associated with the combined lifecycle being given by the attained lifetime reproductive success. Because of this coupling, life-history trade-offs emerge, where the optimization for survival at one stage can impede overall lifetime reproductive success by inhibiting the later development of another trait. In turn, the developmental unfolding of organisms influences ecosystem assembly, and therefore their stability against external shocks or invasive species, but also the coevolution between different populations over time, because the timing and quality of interactions can shift from one developmental phase to another. Moreover, such niche shifts at different stages of the developmental process can serve to minimize competition between adults and juveniles of the same species, or, more generally, between different stages of a givens lifecycle.

Developmental processes are thus not just an organizational necessity in multicellular organisms. Rather, they allow them to exploit different resources at different stages of their lifecycle, thereby permitting the expansion into different habitats and ecological niches, or evolve different ecological interactions at different times. Thus, developmental processes can contribute to ecological complexity. Ecology consequently gains an additional dimension, in that a single biological (or morphological) species can represent a multitude of ecological species over the course of the developmental unfolding of its individuals.

Development and “developmental selves” The developmental process does not only create spatiotemporal and functional organization in a population of cells but also establishes a form of “self–non-self” recognition. Because this self–nonself- recognition is a feature of all living things, including individual cells in a developing organism, this entails the integration of originally independent “selves”, i.e., cells, into a larger whole. The cell population as a whole is eventually able to distinguish cells and other particles that belong to this unit from those that do not, and this new “self” does transcend the analogous feature of individual cells. These are still able to make such distinctions concurrently, albeit in a less complex and differentiated form [Burroughs and Aravind, 2020] and one that, in the multicellular organism, is “subject” to the greater whole. The emergence of this distinction is a hallmark of evolutionary transitions in individuality [Szathmáry and Maynard Smith, 1995; West et al., 2015], although this concept is normally applied to the emergence of self-reproduction of the whole and not to the emergence of developmental selves.

This intrinsic self–non-self distinction provides us with one way to distinguish environment and individual, which we can call a “developmental individual”, and which is defined as the set of interacting entities for which a process emerges that accepts some of these elements and rejects others. We note, however, that this “self” is an emergent entity, or property, of the multicellular system. Individual cells can indeed lose this notion of a larger “self” and become singular again, which is why cancer has been described as a loss of multicellularity [Trigos et al., 2018]. In that way, evolution of cellular life shows the emergence of higher-order elements in a population of complex adaptive systems as described in chapter 2.

Development is paramount for evolutionary theory On the other hand, because development creates all the structures by which a multicellular organism interacts with its environment, development is of paramount importance for the evolutionary process, because it creates the phenotypic variation upon which natural selection acts. More than that, variation in the developmental process itself is subject to and consequence of natural selection. Therefore, development is the central element that links evolutionary dynamics, self-organization and complexity and genetic dynamics, as illustrated in Figure 4.2. This gives a more nuanced meaning to the notion of “self-reference” in discussions of evolution: Natural selection can shape the process that generates the variation upon which it, in turn, acts. More than that, evolutionary developmental theory emphasizes the dynamic nature of individuals, which must be regarded more as comprising a set of continuously unfolding and interacting processes that reproduce themselves rather than as the more or less static “interactors” or “vehicles” [Dawkins, 1976] they are often interpreted as. A corollary is that “what evolves in evolution” are in fact different developmental processes, or lifecycles, instead of only genes or organisms, although their evolution is a necessary consequence.

Although development is concerned with multicellular organisms, the assertion that lifecycles are the evolutionary foci goes beyond that, and so some attention must be given to unicellular organisms in this regard. While they do not undergo development in the same way as multicellular organisms do, they undergo repeating cell cycles that differ in complexity and mechanics between prokaryotes and eukaryotes. These cell cycles can be seen as a unicellular analogon to multicellular development because both follow broadly the same principles. They constitute an ordered pattern of conception, growth and resource acquisition, which leads to a reproductively viable state. Furthermore, unicellular organisms establish a kind of “self-recognition” in much the same way as multicellular organisms do, i.e., have their own immune systems [Westra et al., 2012]. Because cells are the fundamental entities of biological evolution, unicellular cell cycles do not involve the emergence of a higher-order, compound “self” that multicellular lifecycles entail, and are correspondingly simpler. Nevertheless, we can see the emergence of developmental selves as a manifestation of biological information processing, which consequently is present on all scales of biological organization.

The principle that evolution refers to self-reproducing adaptive processes instead of static vehicles can, therefore, be maintained for the entire domain of life[1] In other words, evolutionary change in multicellular organisms always involves development in some way

---

[1] A particular challenge in that regard, however, is posed by viruses, which cannot be said to establish some form of developmental self or undergo development in any appreciable form. However, because they are comprised only of syntactic elements (DNA or RNA) and do not maintain the syntax-semantic split inherent in cellular life, they cannot be said to be “alive” in the same way as cells can, and are thus excluded from the discussion here.

[Stearns and Hoekstra, 2005]. Consequently, developmental processes can be nested and can coopt different stages within each other, i.e., the unfolding of developmental processes on a lower level can provide capabilities that are employed in the higher-order developmental process, while the higher-order process, in order to remain stable and robust against selfish elements, must restrict these lower-level processes by setting their boundary conditions or parameters, i.e., by providing forcing [Fields and M. Levin, 2020b]. This concept has been encountered multiple times already, in particular in chapter 2 and chapter 4, and finds a natural implementation here.

5.2 Plasticity and developmental symbiosis

5.2.1 Development integrates environmental signals

Development does not just create a life cycle according to a plan, as complex as this might be. Rather, developmental unfolding is able to integrate signals from the environment and create an adaptive response in the phenotype of the organism [Gilbert et al. [2015], Sultan [2015] and references therein]. Therefore, the environment a population is embedded in not only plays a role in natural selection and (co-)evolution but is also a source of developmental information. The phenotypic variation a population of organisms will show is thus dependent on the environmental properties it finds itself in.

Integration of environmental signals gives rise to plasticity, i.e., multiple different phenotypes can be generated without change in genetic sequence and in response to environmental changes. Consequently, we identify developing organisms as complex adaptive systems. With that comes the whole spectrum of phenomena known from nonlinear dynamics, most importantly well-defined attractors in the collective behavior of the cells of the developing organism, with their accompanying basins of attraction. Thus, we can interpret the body plans of organisms as attractors in a complex nonlinear system that attains different forms depending on external input. This also connects back to the concept of genomes guiding self-organization laid out in chapter 4.

Developmental variability of a phenotype is represented using norms of reaction, which measure the change of a trait of some organisms over some environmental gradient, assuming the genetic background to be fixed [Stearns and Hoekstra, 2005]. Often, these are abiotic factors like temperature, but biotic factors often induce plastic reactions, too. True to their nature as nonlinear dynamical systems, developing organisms do not just exhibit a continuous reaction of their traits under changing environmental conditions. Rather, akin to bifurcations in simple nonlinear systems, these changes are often qualitative and can transform the developed phenotype substantially.

Indeed, many natural examples of developmental plasticity are characterized by qualitative and conditional changes. For example, the exposure of *Daphnia* water fleas to predators can lead to the creation of protective structures [Agrawal et al., 1999]. Nematodes of the genus *Pristionchus* are able to switch from a bacterial diet to macropredation, i.e., preying on other nematodes based on environmental conditions [Wilecki et al., 2015]. In general, plasticity is thought to be more relevant for sessile organisms which have to cope with the conditions in their habitats as they are, while mobile organisms can move to find more favorable conditions.

Developmental systems thus can be seen as an information processing system (compare subsection 4.1.6) that incorporates environmental signals as well as internal heritable information into its dynamical unfolding to create a phenotype capable of successful reproduction and survival.

5.2.2 Gene-regulation as mechanistic underpinning of environmental signal integration

Genetic architecture implements development In chapter 4, we discussed gene regulation, gene-regulatory networks and epigenetic inheritance in the context of heritable information, and explored various mechanisms for how this regulation is achieved. Most noteworthily, cells continuously adapt to environmental conditions by adjusting the levels of gene expression as a reaction to environmental signals, e.g., to the presence or absence of lactose as has been discussed in subsection 4.2.5. Cells can consequently be seen as complex adaptive systems as well, which makes multicellular organisms, but also eukaryotic cells, nested CAS.

Because the development of multicellulars is a process in which a growing colony of cells cooperates and diversifies, and gene regulation is the fundamental mechanism by which cells adjust to their environment, gene regulation is the fundamental developmental mechanism, in that it's the differential expression and manipulation of heritable genetic information by means of other heritable information and non-inherited environmental signals and their temporal coordination that implement a developmental program.

Genetic organization confers and constrains evolvability and structures the evolutionary process As mentioned in subsection 4.2.6, the genome is organized into coupled gene regulatory networks which are sensitive to input signals and produce some outputs that act on other GRNs again or have another function in the cell. In unicellular organisms, these signaling systems create a temporal coordination of functions that results in the cell cycle. In multicellular organisms, spatial organization attains a greater role, because many mulitcellular organisms, in particular animals, are highly spatially heterogeneous and segmented. As discussed in subsection 4.2.6, GRNs form circuits and motifs connected together. Different GRNs and genes thus can be reused in different contexts whenever a specific function is required, with the switches that link them being a deciding factors for the spatiotemporal developmental process (compare subsection 4.2.6). The total network of gene regulation thus implements the developmental process. As a consequence, some of the regulatory network motifs are deeply conserved across the animal- or metazoan kingdom, and are reused in different contexts. Once established, such deep motifs are hard to change because of evolutionary continuity (see subsection 3.3.5). Therefore, the discovery of the function of GRNs in development provides a mechanistic explanation of why the differences in body plans are greater for different animal phyla than for individual species or clades within them, and why all members of a given phylum, e.g., Arthropoda, share a number of distinctive features (segmented extremities or compound eyes) that other animals, e.g., tetrapods, lack. The same logic can be applied to higher or lower levels of classification.

In other words, genetic architecture can restrict evolutionary change because low-level, highly conserved genetic motifs are so important for the functioning of the lifecycle that any mutation to them would disrupt it and reduce fitness to the point where they will quickly be lost [Erwin and Davidson, 2009]. As a consequence, evolutionary changes are biased towards downstream elements of the organism's GRNs, where changes are less disruptive to the overall functioning of the organism. This architecture thus provides a mechanistic example for when and how evolutionary continuity (subsection 3.3.5) creates trade-offs and higher-than-expected complexity, and it can leave a signature in the evolutionary trajectory of a lineage [Brakefield, 2006].

On the other hand, this hierarchical organization allows for evolution to proceed in a far more flexible manner than would be possible in hierarchy-less genotype-phenotype-map, because mutations in genetic switches can relink or disable entire network motifs (see subsection 4.2.6) that confer novel functions at different times or places or under different ecological conditions. For example, over the course of animal evolution, the HOX-gene cluster has been duplicated multiple times, with different versions having different functions in animal development [Stearns and Hoekstra, 2005]. This gives exaptations, i.e., the reuse of functions in a different context [Gould and Vrba, 1982], a greater role in evolution and, at the same time, provides a mechanistic explanation for it. Concurrently, their ability to integrate environmental signals allows for developmental plasticity by conditionally enabling or disabling or up– or downregulating genetic subcircuits or individual genes.

Because of evolutionary continuity, the evolution of the GRN-based developmental process must have involved a cooption of the existing gene regulatory signaling pathways in eukaryotes when multicellulars evolved from unicellular groups. This came about because variation in gene regulation in cells of some primordial multicellular conglomerate that affected the reproductive success of the group (and thus of the individual cells) came under a changed natural selection regime compared to a unicellular state, leading to evolutionary divergence. Thus, multicellular organisms retained the reactivity to environmental conditions during development by coopting this reactivity from their constituents. In this way, they inherited also the CAS property from their unicellular ancestors.

5.2.3 Development and ecology revisited - developmental symbiosis

Until now, we conceptualized organisms as a collection of interacting processes that form a "developmental self" and share a set of heritable information and which all acted on a collection of cells that arises from a single-celled starting point. However, empirical analysis shows that multicellulars are settled by a multitude of other unicellular organisms [Gilbert et al., 2015]. These symbionts often provide important functions to their hosts. For example, the gut of termites is settled with a set of unicellular organisms that allow them to digest cellulose and which produce methane as a waste product [Breznak, 1982]. A similar dependency can be found in ungulates, and, in fact, the large majority of animals in one way or another. This symbiosis extends also to other tissues, e.g., skin [Ross et al., 2019].

More than just useful helpers, it turns out that these unicellular symbionts are, in fact, important sources of developmental signals which are often crucial for the successful unfolding of ontogenesis of organisms [Gilbert et al., 2015; Carrier and Bosch, 2022]. In particular, the gut of sterile mouse embryos fails to develop correctly, but can later be induced to do so by reintroduction of the normally present symbiotic bacteria. Similarly, in the nematode *Brugia malayi*, the body axis fails to develop correctly in the absence of *Wolbachia* bacteria [Landmann et al., 2014], and the light organs of some bobtail squid species depend on *Vibrio* bacteria living in seawater in a similar way [McFall-Ngai and Ruby, 1991]. These examples are specific in that the developmental function depends explicitly on the bacterium species involved, i.e., the association must have come about by coevolution.

Symbiotic organisms, therefore, are a widespread and vital part of multicellular developmental processes, and the structure of this “internal ecosystem” has a significant influence on their reproductive success. For the symbionts, this dependency is often equally deep, with many symbiotic microbes being exclusive to certain phylogenetic groups. This association often mirrors their evolutionary history of divergence [R. M. Fisher et al., 2017]. Indeed, changes in organismal microbiomes have been implicated to play a role in species divergence [Gilbert et al. [2015] and references therein].

It must be noted, however, that not all organisms that settle a multicellular body and are part of its internal biome are necessarily beneficial symbionts. Many can be more or less neutral additions that neither significantly harm nor benefit their host, while others can indeed be harmful and must be suppressed by the immune system, and yet others are fully beneficial [Douglas and Werren, 2016]. Moreover, if an organism is one or the other can significantly depend on the location it settles in the developing body. Thus, rather than thinking of natural organisms and their microbiome as mutually beneficial integrated wholes, it is arguably more accurate to see them as a kind of self-managing community, part of which undergoes a collective replication process, while other parts are assembled from its environment over time [Douglas and Werren, 2016].

Nevertheless, because of their frequent importance for the timing and quality of developmental steps, the symbiotic ecosystem of organisms can be seen as part of the developmental signaling network, and its formation over time as being part of the developmental process. Indeed, the evolution of the forcing the multicellular whole exerts onto its microbiome can be seen as a central theme in the evolution of multicellular development, in much the same way as the forcing via heritable information on protein self-organization and cellular assembly. The self-nonself-recognition introduced above must include the management of this internal ecosystem, i.e., a plausible argument can be made that immune systems of multicellulars have been formed at least in part by the requirement of maintaining this vital element of developmental signaling [Gerardo et al., 2020].

5.2.4 Plasticity and evolution

Plasticity, as mentioned, can affect the phenotypic variation exposed to selection, thereby influencing ecological interactions. As a consequence, this can change the evolutionary trajectory of populations.

Cryptic genetic variation Cryptic genetic variation is genetic variation that normally is not exposed to natural selection, i.e., is neutral but can have fitness effects when environmental conditions change [Gilbert et al., 2015; Paaby and Rockman, 2014]. Elements of a gene regulatory network that are disabled or suppressed under normal conditions can accumulate variation over time, which then can become exposed to selection once an outside influence disables or reduces this suppression [Ng and Kinjo, 2022]. Thus, organisms harbor, generally, much more genetic variation than is exposed to selection at any given time [Paaby and Rockman, 2014; Gilbert et al., 2015], allowing for greater evolvability under varying environmental conditions and for accelerated evolution [Zheng et al., 2019]. This is of particular importance in the case of rapid changes to a given environment, where cryptic genetic variation can provide phenotypic variation as a response to environmental change, thus allowing for greater evolvability “right when it is needed”. Cryptic genetic variation thus contributes to solving the problem of gradualism for large environmental change, because genetic variation that contains potentially adaptive traits under the new conditions does not need to be first evolved via small steps while the environment already has moved away from the configuration the population is adapted to. Rather, populations harbor a large pool of genetic (-regulatory) variation that is not expressed under established conditions but can be expressed when these change. In other words, the trait memory of the population contains many neutral or unused elements. It's no large jump to connect this to learning systems, which can memorize past events and reproduce them upon encountering similar conditions again. In particular, models for gene-regulatory networks have been shown to be equivalent to RNN[2]s (R. A. Watson, G. P. Wagner, et al. [2014] and references therein), and developmental systems have been discussed as learning systems by these authors.

Genetic accommodation Plasticity can lead to organisms expressing new traits under changed environmental conditions. If these traits are adaptive, and there is variation in their expression between individuals, natural selection should favor the variants that express the plastic trait most reliably and advantageously [Ng and Kinjo [2022], Gilbert et al. [2015] and references therein]. This results in an originally plastic trait becoming assimilated into the heritable information of the organism such that it is expressed independently from the original environmental signal. Genetic accommodation thus provides a mechanism in which plasticity can be the first step of evolutionary change, with genetic variation reacting to it via natural selection, in contrast to the classical notion of genetic mutation being the first step towards novelty. Indeed, genetic mutation is not required to initiate the process at all, although it does require some preexisting genetic variation. The key to this mechanism is gene regulation, which is envisaged to change under new environmental conditions because different environmental signals lead to cryptic variation in regulation being activated. These regulatory pathways can then evolve to become less and less dependent on the environmental signal in order to be expressed if their product is adaptive, and thus variation in them is selected against. This concept of plasticity-led evolution also includes the notion of plasticity being universal to developmental systems [Ng and Kinjo, 2022], and the assumption that it generally is able to produce adaptive phenotypic variation.

Plasticity-led evolution Plasticity-led evolution can be seen as a generalization of the Baldwin-effect [Baldwin, 1896], bringing together the above concepts of cryptic genetic variation, genetic accommodation and phenotypic plasticity. In the Baldwin-effect organisms change their interactions with their environment through learning, which then is assimilated into the genome via genetic accommodation. The generalization thereby happens in the source of the original change, from learning of new behavior to any form of adaptive phenotypic plasticity. While cryptic genetic variation allows for populations to persist under environmental change due to larger adaptability, plasticity allows for the rapid formation of new traits through which organisms can survive and thrive initially. The changed ecological interactions will uncover cryptic genetic variation, which will produce wider phenotypic variability and thus modify natural selection on these newly expressed traits. Genetic accommodation then can happen when the originally plastic trait is also, at least partially, expressed through uncovered cryptic pathways, such that genetic variation in the reliability and quality of the adaptive plastic trait exists on which natural selection can act. It must be noted again, as Ng and Kinjo [2022] point out, that for this to be a general process in evolution, adaptive plastic responses need to be a general feature

---

[2] Recurrent neural network, a kind of artificial neural network that allows connections between nodes to form cycles.

of natural genetic regulatory systems, which is an as of yet not sufficiently validated assumption.

Epigenetic inheritance as a mechanism for the inheritance of plastic responses An intermediate step in the genetic accommodation process is sometimes provided by epigenetic inheritance of regulatory signals like microRNA (compare subsection 4.3.1), which allow for regulatory effects originally brought about by environmental change to persist even after the environmental signal no longer exists, thereby exposing the plastic phenotype to selection for longer times. Epigenetic inheritance therefore can, in principle, bridge the gap between the timescales of environmental signal persistence and natural selection taking effect, and thus facilitates genetic accomodation.

Genomes as dynamic instruction sets instead of blueprints Classical evolutionary theory interpreted phenotypic variation as a consequence of genetic variation, modulated by non-evolutionary processes like learned behavior. The phenotype was thus regarded as a consequence of the genotype, which gave rise to the idea of the genotype-phenotype map. That being said, it was long since known that organisms show variation in their phenotype under different environmental conditions (Stearns and Hoekstra [2005] and references therein), but this data either failed to be integrated into larger evolutionary theory or was interpreted as arising as a consequence of genetic variation alone.

Modern research into developmental processes suggests that a more integrative approach must be taken, and genomes should be regarded more as "*repertoir of environmentally contingent possibilities* rather than a single determined outcome" [Sultan [2015], p. 20], as discussed multiple times in this thesis. With development being identified as a complex adaptive system, and understanding that organisms achieved von Neumann's syntax-semantics-split, we can go one step further and interpret genomes as self-modifying instruction sets that react to environmental signals. Indeed, Fields and M. Levin [2020a] went as far as suggesting a symmetry between evolution and development, based on the idea that all life on earth can be seen as a succession of cell divisions. It must be noted that such a view would have to be reconciled with the aforementioned emergence of "developmental selves" during the unfolding of a lifecycle – a task beyond the scope of this thesis.

Organisms as evolving holobionts Often, symbionts are acquired through environmental interaction, such that the environment becomes a source of phenotypic variation in yet another way. Secondly, this environmental acquirement may be an additional source of phenotypic plasticity, when variation in the acquired community exists and is environmentally dependent. On the other hand, many of the most vital symbiotic organisms are in fact transmitted vertically from parent to offspring [R. M. Fisher et al., 2017], although, in principle, these two modes of transmission need not be as clearly separated as they may appear. For example, parent organisms may construct or seek out specific environments to rear their young, and this may entail shaping them to allow the infection of offspring with symbiont organisms.

If symbiotic organisms are transmitted from parent to offspring, this aligns their fitness interests, i.e., their reproductive success is interdependent and too selfish symbionts will likely die out by harming their host. In such a case, variation in the symbiont composition of an organism will translate into phenotypic variation upon which natural selection can act. Therefore, the evolving developmental process is a conglomerate of entangled developmental unfoldings of the multicellular organism and its symbionts. These processes form a network that coevolves, thus extending the classical notion of what a developmental or evolutionary individual is. Indeed, it has been shown that in some organisms the microbiome is responsible for creating reproductive isolation, thus having a macroevolutionary impact [Brucker and Bordenstein, 2013]. On the other hand, the concept of the holobiont has also been criticized for placing too much emphasis on symbiotic interactions and ignoring the variation in degrees of mutualism involved in natural symbiotic relationships.

5.3 Summary

Development is the process that generates the reproductively viable multicellular body from a unicellular initial state in most animals, plants and fungi. This process thus creates the phenotypic variation in a population upon which natural selection acts and is thus relevant for evolution. Indeed, all evolutionary change in these organisms can ultimately be seen as change in development. More than that, the developmental processes active in organisms can have significant consequences for the assembly and dynamics of ecosystems, and thus for ecology.

Developmental unfolding is heritable, and thus development and lifecycles are subject to natural selection, leading to life-history evolution and the emergence of different strategies for growth and reproduction. More than that, the developmental process does produce a "developmental self" which is able to distinguish elements of itself from other elements in the environment. This developmental self is an emergent property based on the interaction of other "selves", namely individual cells.

Development does, however, not just create an interaction platform that reproduces in the end. Rather, it integrates environmental information and is able to adapt the developmental trajectory itself as well as the character of individual traits to these signals. Therefore, it forms a complex adaptive system. Because all stages of the process are relevant for the eventual lifetime reproductive success of an organism, we can see the evolution of organisms as the evolution of dynamic adaptive processes – lifecycles – instead of only the evolution of genetic sequences or more or less static organisms [Smith, 2023]. Consequently, because of all these far-reaching properties of development, organisms can be seen as self-reproducing processing systems for biological information and the developmental systems can thus be seen as the focus of evolution. This even extends to unicellular organisms, which, while lacking true development, still operate in a cyclic manner similar to the one outlined in Equation 4.1.9, which also constitutes a process which dynamically incorporates external signals.

With this, environmental signals become important agents of the evolutionary process, thus elevating the environment from a background filter to an active element in the evolution of the biosphere. Mechanistically, this is accomplished by the effect of environmental signals on gene regulation. The hierarchical, modular nature of GRNs, where elements are up- or down-regulated or enabled or disabled in reaction to external signals, thereby not only allows for the information processing that underlies phenotypic– and developmental plasticity but also leaves a pattern in the evolutionary history of multicellular life through the frequent cooption and repurposing of existing GRN circuits in different contexts. Thus, the evolution of developmental processes can help explain patterns of variation we see in the biosphere and the fossil record.

Plasticity can shield genetic variation from selection and allow organisms to persist in otherwise unfavorable environments. The evolutionary implications are multiple, from allowing for increased colonization of new habitats to increased adaptability, but also to potential maladaptation [Langerhans and DeWitt, 2002]. Plasticity-led evolution is an interesting hypothesis in this regard, which allows for a population to overcome the problem of gradual change to genetic heritable information by externalizing evolvability into their plastic developmental systems. In particular, epigenetic inheritance and genetic accommodation are mechanisms by which plasticity can facilitate evolutionary innovation without relying on an initial change in genetic sequence.

Finally, the developmental unfolding of organisms is not only dependent on external environmental signals and genetic makeup. but crucially on the presence of symbiotic unicellular organisms that interface with signaling pathways to influence the timing and expression of developmental steps. Without them, the developmental process cannot function correctly, i.e., it will produce defective individuals with reduced fitness. Often, symbionts are vertically transmitted and thus heritable, leading to direct coevolution between the organism and its symbionts, although acquisition from the environment is also frequent. Thus, they constitute not only a source of heritable variation that natural selection can act upon or act as the source of plasticity but form a crucial element of the developmental process that regulates other steps in the process but is also regulated in return. Finally, there have been documented instances of reproductive isolation and thus speciation being dependent on symbiont composition. All this led to the notion of organisms as a collection of cells plus their symbionts, or “holobionts” [Chiu and Gilbert, 2015].

6. Niche construction

6.1 Ecosystems and niche spaces are biotically constructed

6.1.1 Niche change and the role of organisms in constructing ecological niches

Organisms modify their habitat as consequence of their metabolism The lifetime reproductive success of organisms depends on their interactions with their local environment throughout their lifecycle. Thus, active manipulation of their environment at specific stages of their developmental unfolding appears to be a natural strategy that we should find in nature. Indeed, because organisms must be open systems connected to their environment via matter and energy flows, it is unavoidable that their actions will modify the environment around them, at least by means of resource acquisition or –spoiling and waste excretion.

Grazing animals foster the growth of certain plant species through their dung and trampling or provide nourishment for coprophages. The most potent such effect was probably the oxygenation of the Earth's surface by photosynthesizing organisms. Indeed, through their metabolism, biological organisms process matter into different forms often not produced abiotically, thereby contributing to the dimensionality of the niche space in a habitat. The most consequential such metabolic influence was arguably the oxygenation of Earth's surface, which fundamentally formed the history of life.

Physiological structures contribute to niche space dimensionality This contribution or organismal activity goes beyond their metabolism, however. Many organisms form body structures that serve others as a habitat. In the tropics, many animals live entirely arboreal [Kays and Allison, 2001], i.e., their habitat exists solely as the consequence of an individual tree's lifecycle. The mere existence of these plants, therefore, creates a whole subspace of the local niche space that is absent wherever no trees exist. Similarly, our skin is the home of a plethora of organisms ranging from mites to bacteria, as is our gut, mouth, or mucous membranes. We discussed already that this is more than just ecological association because symbiotic organisms are an important source of developmental signals.

Coral reefs are another striking example, where populations of individual corals change shallow marine habitats into very diverse habitats [Plaisance et al., 2011]. Physical structures of organisms can also have strong influences on the emergence of landforms like dunes or meandering riverbeds, whose dynamics is often determined by the trapping of sediment by plants [Corenblit, Steiger, et al., 2007; Corenblit, Baas, et al., 2011].

Habitat changes through behavior Beyond contributing to a habitat's niche space by metabolism or body structure, many organisms actively modify their habitat over the course of their lives through their behavior. For example, earthworms change the soil through their burrowing and other modifications in order to facilitate water intake, turning it into an environment in which they can live despite their originally aquatic physiology [Turner, 2002]. Beavers, among the most prevalent examples, create ponds and transform terrestrial habitats into swampy or aquatic ones, changing the structure of the local niche space dramatically in the process [Pelletier et al., 2009]. Other organisms build nests or burrows that outlast their physical bodies and which, therefore, change the local environment. The cultural evolution of *Homo Sapiens* is a particular case in this regard, where collectively learned behavior, e.g., the construction of cities and the use of agriculture, has planetary-scale influences on ecosystem structure. It is important to note that behavioral changes to a habitat are usually adaptive, i.e., they have been shaped by natural selection to be a part of an organism's lifecycle, such that removing them will negatively impact its fitness.

Ecosystem engineers Some organisms, including beavers, humans, corals, or earthworms, exert, through physiology, behavior, or metabolism, a fundamental influence on the structure of the niche space within their habitat, such that their presence or absence precludes or facilitates the persistence of many other species and forms the ecological network they are a part of. Such organisms are called "ecosystem engineers" [Jones et al., 1994; Jones et al., 1997]. Their importance shows that ecological interactions go beyond resource consumption or trophic interactions and can entail the fundamental transformation of local habitats within one or a few lifetimes of the acting organisms. Therefore, ecosystem structure can usually be seen as biotically facilitated as much as adapted to abiotic conditions, and so the ecological niches organisms occupy during their lifecycle are largely the result of biotic influences. Because ecosystems are interaction networks, coevolution is the norm rather than the exception.

6.2 Niche construction in evolution and development

6.2.1 Consequences of organismal change for the actors and the wider ecosystem

Consequences of environtmenal influences for the focal organism As a consequence of natural selection, the environment-changing traits of organisms can be thought of as part of the phenotype of an organism. Thus, the phenotype of an organism does not end at the physical limits of its physical body, which we normally think of as constructed through the developmental process. This boundary is somewhat arbitrary, as, e.g., behavior is doubtlessly influenced by genetic makeup. Within the framework of the gene-centered view of evolution, this leads to the concept of the "extended phenotype" [Dawkins, 2016], which develops the argument that the effect of a gene, i.e., a phenotypic trait, is not bound to the physical body of its carrier, but can manifest itself in changes to other environmental properties, e.g., physical artifacts or manipulation of the fitness, i.e., ecological interactions, of other organisms.

Extended phenotypes are common, and given that genes are the most stable form of heritable information, we can expect environmental modifications brought about by organisms to have a genetic component, influences from other sources of heritable information notwithstanding. This is especially important where culture is not a major component of an organism's developmental trajectory and where modification behavior is complex and costly, such that variation in it will likely reduce fitness. Given the discussion in chapter 4 and chapter 5, it is easy to extend the concept of the extended phenotype to accommodate all effects organisms have on their environment, no matter their underlying heritable information.

An organism-centric view has been developed within the framework of niche-construction theory [K. Laland, Matthews, et al., 2016; F. J. Odling-Smee et al., 2003], which is often contrasted with the gene-centric theory of the extended phenotype. Niche construction is defined as the change of the evolutionary niche of a population by their own or others' actions [K. Laland, Matthews, et al., 2016]. The evolutionary niche is thereby defined as the sum of all selection pressures [K. Laland, Matthews, et al., 2016; F. J. Odling-Smee et al., 2003], which does not in all cases coincide with the ecological niche [Trappes, 2021]. Indeed, not all possible ecological interactions of an organism may be under selection at the same time or to the same degree. Therefore, niche construction theory explicitly emphasizes that organisms are at least partially the architects of their own selective environment, such that natural selection is not an effect of a static environment but the product of a feedback process between a population and the rest of the ecosystem. Among other things, it has been shown in computational models that niche construction can lead to evolutionary inertia, where a trait is shielded from selection through the construction of an environment that buffers selection pressures on it [K. Laland, Matthews, et al., 2016], which is consistent with the earthworm example given before.

Niche construction theory is more inclusive than the concept of the extended phenotype in that it does not build on a preconceived concept of inheritance. Consequently, cultural niche construction has become a field of interest [K. N. Laland and O'Brien, 2011].

Moreover, niche construction theory explicitly includes ecological inheritance (see subsection 4.3.2) and thus emphasizes the connections and feedback processes between environment and evolution. At the same time, it encompasses the concept of the extended phenotype as a special case that is bound to direct genetic effects. It must be noted that, in general, behavior and its environmental effects are integrated into the lifecycle of an organism and, therefore, must be regarded as the result of the developmental system of the organisms in question of which individual genes are a part. It's argued here, therefore, that the gene-centric view of evolution is only part of the picture, as has been discussed in chapter 4 and chapter 5. Niche construction not only encompasses physical changes but also the movement to other habitats with different properties when this is based on an evolved "choice", called "relocatory niche construction" [R. L. Day et al., 2003] to distinguish it from niche construction brought about by active modification, called "perturbatory niche construction".

Niche construction theory as a scientific framework posits that niche construction must be seen as an evolutionary process in its own right, on par with natural selection [K. Laland, Matthews, et al., 2016], a position that has elicited significant debate [Scott-Phillips et al., 2014]. Because niche construction alters the selective regime of an organism at a specific stage of its lifecycle, we can expect variation in active niche construction to yield variation in natural selection, such that niche construction traits become subject to natural selection. This is trivially fulfilled for extended phenotypes, which are based on genetic heritable variation, but it is much harder to prove for epigenetic causes, in particular for the effect of variation in ecological inheritance. This has led Dawkins [2004] to argue that niche construction proper, where this correlation is observed, is fully explained by the extended phenotype, while other examples, especially the ones that occur as side effects of metabolism or body structure, he labeled "niche change". Similarly, the explanatory power and conceptual consistency of niche construction theory have been called into question [Scott-Phillips et al., 2014]. This, however, does not diminish the validity of the phenomenon as such, which doubtlessly has important consequences for biological evolution, be they mediated by genetic or epigenetic inheritance.

Consequences of niche construction for the embedding ecosystem Niche construction not only affects the natural selection regime that the actors are exposed to. Rather, it also has an effect on the natural selection regime of other organisms that interact with the modified properties of the ecosystem, e.g., bacteria or insects that live in animal burrows or ponds created by beavers. Therefore, in much the same way as physical structures of an organism's phenotype or metabolic effects influence niche space, active construction can lead to coevolution between organisms and thus can be seen as a structuring force in an ecosystem that is not limited to the cases where the actor species is a full ecosystem engineer. Indeed, we can deduce from this discussion that the relative importance of constructive activities for ecosystem structure is not a fixed scale, such that organisms that act as ecosystem engineers in one habitat may be less important in another. Niche construction thus enables environment-mediated coevolutionary feedback between organisms that can be indirect and does not rely on direct physical interaction. This process can span timescales larger than the lifetime of the organisms in question, such that ecological interactions can be mediated across large timescales.

On the macroscale, niche construction can have important effects on niche space complexity that, in turn, facilitate speciation and evolutionary innovation and are thought to be an important modifier for the carrying capacity of an ecosystem [Erwin, 2008]. One example is the Cambrian substrate revolution [Bottjer, 2010], which made the sediment at the ocean floor, previously a more or less two-dimensional habitat dominated by bacterial mats and probably early animal grazers, into a three-dimensional, oxygenated, complex sub-ecosystem which increased the available ecological niches and thus allowed for many species to evolve new adaptations, creating a more diverse ecosystem in the process. Thus, the emergence of novel niche construction adaptations can lead to major evolutionary shifts that can persist for long times. Possibly, some of these behaviors lead to coevolution between the actor and other organisms, leading to fixation of particular forms of niche construction in the biosphere, such that some ecosystem elements become fixated in a superficially similar fashion to network motifs in GRN[1]s.

6.2.2 Niche construction, development and plasticity

Bringing development and niche construction together yields a more complete picture of how niche construction influences individuals. Indeed, niche construction complements plasticity in that it shows how organisms are able to manipulate an important source of regulatory signals for themselves and others.

Consequently, adaptive niche construction is sometimes considered separate from developmental niche construction [Stotz, 2017], with the latter describing niche construction to facilitate stages in the developmental process of an organism or its offspring. Indeed, providing an environment that is conducive to the developmental process of offspring is an important part of the lifecycle of many organisms. This shows how variation in niche construction can yield variation in plastic responses, thus facilitating plasticity-led evolution (compare subsection 5.2.4). Recalling the view of the genome as a processing system for biological information and the developmental process as the subject of evolution,

---

[1] Gene regulatory network

this makes clear that these systems do not just integrate environmental signals that are independent of them but evolved active structures to produce and control these environmental signal sources. Therefore, the evolution of developmental programs can be seen as the evolution of the processing *and* manipulation of biological information. Therefore, niche-constructing behavior in itself can, like every other organismal activity, be seen as part of the developmental process of an organism [K. N. Laland, J. Odling-Smee, and Gilbert, 2008]. It can have similar effects to phenotypic plasticity, e.g., promote persistence in unfavorable environments and facilitate colonization of new habitats or shield genetic variation from selection.

In fact, niche construction and plasticity combined open a path to phenotypic evolution without genetic change, in that a change in niche construction can bring about the environmental change mentioned in subsection 5.2.4 that elicits a modified plastic response and thus a change in ecological interactions, i.e., phenotypic variation in a population. Genetic evolution then can follow as a consequence of the changed selection regime. In such a way, niche construction and plasticity may have significant consequences for evolutionary trajectories. This is especially relevant for learning, which is an important and powerful plasticity generator. Thus, we can see culture as a collectively learned niche construction.

With regard to developmental symbiosis, niche construction by the host on the ecological niche of the symbionts can be seen as a mechanism that can cement the symbiotic relationship and ensure specificity in the symbiont's signaling behavior, its location and timing [Gilbert et al. [2015] and references therein]. Niche construction, therefore, may also play a role in evolutionary transitions in individuality [Torday, 2016]. Niche construction on this level, thus, is no different, conceptually, from niche construction on the ecological level.

6.3 Summary

Natural organisms not only adapt to an existing ecosystem and receive developmental signals from it, but they also actively and passively change it. These changes can come about as part of their metabolism or physiology but also as actively enacted modifications of environmental structures or other organisms, i.e., through organismal behavior. These influences modify how organisms interact with biotic and abiotic entities in their local ecosystem, such that organismal modification can structure ecosystem networks. Moreover, some species change the character of a habitat so fundamentally that they can be seen as ecosystem engineers, i.e., they create conditions that drive the local ecosystem into a fundamentally different state from the one it had if this species was absent. Consequently, organismal activity changes the selective regime of natural populations.

More generally, organismal activity changes the flow of matter and energy and their availability within an ecosystem, thus creating opportunities for adaptation and contributing to the local biodiversity. Besides, through physical changes like structure formation, burrowing, or bulldozing, they change the local complexity of the niche space, thus contributing fundamentally to the evolutionary trajectories of different populations in the ecosystem. On the macroscale, construction effects can thus lead to speciation and precipitate adaptive radiations, such that biological diversity over geological timescales is, to a large part, a consequence of evolutionary innovations in ecosystem modification.

Because niche construction behaviors are not bound to the physical boundaries of the acting organism, the concept of the “extended phenotype” has been invented, in which the ecosystem-altering effects of organisms are interpreted as the result of genetic expression. In contrast, the concept of “niche construction” does not reference any source of heritable information explicitly and describes organismic activities as a modification of the sum of selection pressures (the evolutionary niche) to which local populations are subject. Thus, it explicitly includes epigenetic inheritance and professes an organism–centered instead of gene–centered view of evolution. The theory that has been built around this term emphasizes ecological inheritance as a mediator for changed selective regimes and as a driver of biological evolution.

Merging these concepts with development allows for a reconciliation of the concepts of extended phenotypes and niche construction in that extended phenotypes or any niche-constructing activity can be seen as part of the unfolding of a developmental process, which we previously identified as the focus of evolution. Plasticity can be brought about by niche construction through the modification of environmental signal sources, which thus implies that organisms are able to modify their own developmental signal environment, but it also can be present in niche construction traits themselves. Therefore, a feedback loop is identified that allows for phenotypic evolution without genetic change by successive changes of niche constructing behavior, which leads to changed plastic responses, which can lead to changed niche construction, and so on. This can facilitate plasticity-led evolution but also the production of phenotypic variation without genetic change. Moreover, if the environmental modifications that bring about a specific phenotype are heritable by future generations, a non-genetic mode of evolution has been established, which can be seen as part of epigenetic inheritance.

Indeed, developmental niche construction as a behavior that facilitates some developmental steps of the actor or its offspring is often separated from adaptive niche construction, as discussed above. Indeed, niche construction appears to play a particularly important role in developmental symbiosis, where host organisms engineer the environment of their symbionts in order to create specificity in signal location, timing, and quality. More generally, merging development and niche construction shows that developmental systems, interpreted as processors for biological information, not only rely on pre-existing biological signals that are then integrated but that evolved elements in this system specifically modify or create environmental signals that can be relevant for developmental unfolding and reproductive success. These concepts can be extended to evolutionary transitions in individuality, i.e., the formation of a new evolutionary–, and thus developmental–, self from a group of initially individual ones. However, this has been treated in-depth elsewhere [Sevinchan, 2021].

In summary, like evolutionary developmental biology acknowledges the nature of organisms as continuously unfolding, reproducing complex adaptive systems, so does the concept of niche construction acknowledge the nature of the environment, or the local ecosystem, as an active player in the evolutionary process instead of a passive filter for phenotypic variation. Similarly, merging development with niche construction shows how developmental processes are not just passive processors but active creators of biological information.

Part II

Application

7. The Utopia project

7.1 Motivation and Design

With in-silico experiments and large-scale computer-based data processing being on the rise for decades, software development has turned into an important skill for most natural scientists. By now, this goes well beyond the craftsmanship of programming. The term "in-silico *experiment*" in itself entails that we expect such software to allow us to make inferences, or at least informed conjectures, about natural processes. As such, they must be held to the same standards of reproducibility and documentation as traditional experiments. With this, testing and quality control have attained a central role in scientific software development. At the same time, the increase in computing power available has enabled us to tackle more complicated problems, especially in complexity science and biology. Consequently, the software to describe these problems has become larger and more complex, often no longer being created and maintained by a single developer but by groups of contributors distributed all across the globe. With this increased complexity comes the need for more sophisticated software engineering techniques that enable the effective maintenance of such complicated software packages. At the same time, software engineering and computer science are not the primary focus for most natural scientists, and so many software packages have arisen that enable users to address complex research questions without an avid knowledge of software development and design [Wilensky, 1999; Datseris et al., 2022; Kazil et al., 2020].

More recently, these packages have come to harness the power of collaborative, open-source development, which directly brings together developers and user feedback and thus can result in agile adaption of the software to new demands and effective bug fixing. Furthermore, usage of an existing software platform alleviates us from "reinventing the wheel" each time and allows the user to progress much more quickly from development to analysis, which, after all, is the step they normally are interested in, significantly boosting research productivity.

All that being said, many packages available are built upon a specific toolset, which has little overlap with the adopted, usually deeply entrenched tools and workflows of a group of users. For the research group in which this thesis has been carried out, this was reinforced by a very broad range of interests, reaching from geomorphology over the evolution of cooperation to opinion formation in human societies. With this plethora of research topics comes a corresponding range of modeling techniques, from ordinary differential equations over cellular automata to agent-based models with elements organized in various kinds of networks, with the latter three having taken precedence over the first due to their comparative ease of formulation, all of which would have to be accommodated by an existing modeling framework. Thus, no such system was adopted initially. The need for a unifying software framework became more pressing, however, when the majority of the group began working on methodically related modeling projects. Each member devoted a considerable percentage of their limited project time to developing infrastructure code, which resulted in code of widely varying quality and reusability and created unnecessary redundancies. Thus, the Utopia Project [Riedel et al., 2020; Sevinchan, Herdeanu, Mack, et al., 2020] was born to [Gaskin, 2021]:

- Create a versatile and expandable framework for simulating coupled complex systems that range from geomorphology to evolving populations of diverse entities.
- Build a set of computationally efficient, reusable algorithm building blocks from which to assemble models easily
- Make complex modeling and model coupling accessible to non-experts, thus boosting scientific productivity

Additionally, we wanted to achieve a tight integration of data generation and -analysis and a high degree of automation for repetitive workflows. This led to the integration and continued development of the Dantro framework [Sevinchan, Herdeanu, and Traub, 2020], which is used in Utopia for automating data analysis and visualization. Furthermore, we wished to leverage our existing expertise in modern C++[1] and Python3. The project has chiefly been developed, designed, and maintained by four Ph.D. students (Benjamin Herdeanu, Lukas Riedel, Yunus Sevinchan, and this thesis' author), for all of whom it was a major part of their work. The maintainer team has since been reinforced by Julian Weninger and Thomas Gaskin, who came in touch with Utopia as Master's students and have continued to use it since as part of their respective Ph.D. thesis outside of Heidelberg University.

At the time of this writing, the project is actively developed via GitLab and has its own web presence [Gaskin, 2021]. All in all, over twenty people provided contributions to the framework, most of them master's students in the group, all of whom successfully used Utopia to carry out their respective thesis. Moreover, Utopia has successfully been used by students as their primary simulation tool in the "Chaotic, complex and evolving environmental systems" lecture and the "Modeling and Simulating Interacting Complex and Evolving Systems" master's seminar given at Heidelberg University between 2018 and 2021 by Prof. Roth.

Taking up ideas from Helbing [2012], we envisage modeling a complex system to be a multi-stage process.

- Conceptualization: Analysis of the phenomenon to be modeled, extracting the core dynamics as far as possible, and choosing how to turn this into a computer model.
- Implementation: Production of computer code according to the produced concept. Frequently, oversights in the design become clear only once implementation has started, so this is tightly intertwined with the conceptualization phase.
- Plausibility testing: It is not clear why a complex computer program intended to simulate a (much more) complex system should produce reliable results when these depend on differences in single characters in the code, rounding errors due to limited accuracy, and so on. Therefore, code testing (unit tests, functional tests, integration tests, etc.) takes a central place in the developing cycle.

---

[1] Modern C++ is usually taken to mean C++11 (released 2011) and later.

- Model validation: While the code per se might be correct, the mechanics envisioned during conceptualization might lack some crucial element or might have some other fundamental flaw. In order to discover this, the model output should be compared to well-known observations from nature. Note that this is not always possible to its fullest extent, as some systems (many evolutionary systems among them) are hard to observe in detail.

- Sensitivity analysis: Every model is complemented by a set of parameters that determine the interactions between elements and set their relative timescales. In analogy to bifurcation diagrams, scanning the parameter space of a model yields important insights into its character and limitations. Sensitivity analysis in itself can yield important results because the comparison between the state space of the model and the real-world system often allows conclusions about the latter or yields clues on how to improve the former.

- Production: Here, the research question of choice is addressed. This might change comparatively frequently, so a jump between the production–, conceptualization, and testing phases is a common mode of operation when working with computer models.

Of course, these phases are not usually separated as clearly as outlined above and are, in reality, much more intertwined, as results from sensitivity analysis frequently elicit a change in conceptualization. Also, data analysis takes up significant amounts of computing power and working time for the latter half of the cycle. In order to meet the goals outlined in section 7.1 and best support the development cycle, Utopia features a modular design that allows the user much freedom in choosing predefined algorithms or employing their own implementations. In particular, the model presented in chapter 8 does make use of only a minimum of Utopia's facilities, using custom-built solutions for the most part. The Utopia framework can be dissected into four modules (see Figure 7.1), which can coarsely be divided into data generation and data analysis. Beyond this modular code design, quality control had a high priority in our group, and we employ automated testing to monitor code quality continually. Additionally, we adopted a policy by which code could only be merged into the main framework when accompanied by appropriate tests and after it had been reviewed by at least one maintainer.

7.2 The Core module

The foundation of the framework is formed by the "core" module. This module provides predefined templates for a user to build their own models on. Its design revolves around the basic idea of a complex system, i.e., a system made up of many components that change their internal states over time through local interactions with other components. In most cases, an interaction is thought of as belonging to one of only a few different types, and when talking about their software implementation, these classes are usually called rules. This structure might remind the reader of cellular automata, and in fact, it is this direct similarity between the structure and cellular automata that makes the latter so formidable a tool in modeling many complex systems. The concept employed by Utopia, however, goes beyond the classical cellular automata in that it leaves open the meaning of state, locality (i.e., space), and the nature of the rules, therefore providing a

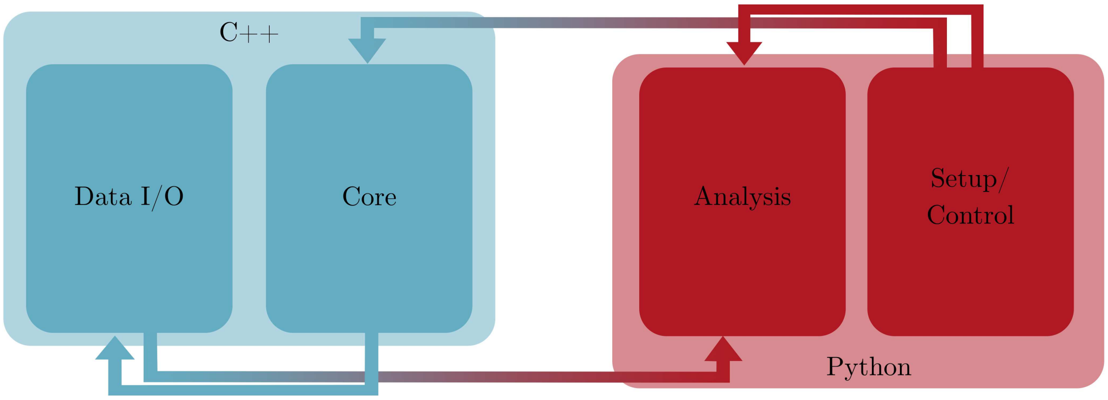


Figure 7.1: The Utopia framework is composed of four modules, which, in conjunction, provide a fully automated data-generation analysis cycle. The "setup/control" module is responsible for processing configuration files defining the simulation and analysis to be made. It then sets up the simulation, parallelizes runs for different parameters (if so specified), and sets up the analysis to be executed afterward automatically. The "Core" module comprises a library of modern C++ software modules that allow easy and performant definition of models. The "Data I/O" module is built on top of the C-based HDF5 library and provides largely automated interfaces to the C++ standard library, such that the peculiarities of HDF5, are abstracted away as far as possible. Where data is stored is largely specified in the configuration file controlling the simulation and can further be automated by employing a provided manager class. The "Analysis" module finally uses the produced data and supplied configuration files to execute predefined data-analysis and visualization procedures, completing the cycle. The Analysis and Setup modules are written in Python because of the latter's flexibility and feature richness, whereas the Data I/O and Core modules are written in modern C++, which provides fine-grained control and performance.

much more flexible concept. It inherits from cellular automata its fundamentally discrete nature, however, especially with respect to time, which is assumed to proceed as an iterated sequence of rule applications. Moreover, complex systems are by their nature organized into multiple levels of organization because, by definition, they are made up of many interacting entities (the first level of organization) that organize autonomously into a coherent whole (resulting in emergent properties of the complete system, the second level of organization). Modular architecture is frequently exploited in software engineering to make systems more flexible, reusable, and easier to design, maintain, and optimize. Therefore, a comprehensive modeling framework for complex systems should allow for the exploitation of such properties. The Utopia core module implements these basic ideas in the form of a base class (just called "model") that provides customizable time-discretization, monitoring, and I/O capabilities.

The user, therefore, needs to focus on three things: The definition of the entities making up the system, for instance, what variables can be used to describe a fish in a swarm (e.g., X position, V speed, $[N]$ its nearest neighbors), the rules by which these variables change, possibly depending on other entities (in case of a swarm of fish, these might be "go into the average direction of your N nearest neighbors") and the combination of the rules into a step which can be iterated to create the model trajectory ("for each fish: measure the direction of N nearest neighbors, compute their mean direction, move with speed V into that direction”). For all the infrastructure code like data output, model configuration, or progress monitoring, the user only needs to fill the parts that explicitly pertain to their model, for instance, what variable to write to file for each entity. Beyond this very basic and abstract concept, helper classes are provided that cover the most common incarnations of complex systems models: Cellular automata (Euclidean grids of immobile cells carrying a dynamic state in fixed local neighborhoods) and agent-based models (free entities that can appear and vanish and which dynamically form interacting neighborhoods). While both paradigms are of the same expressive power, each is more easily applied to some problems than to others. For instance, a CA formulation seems natural for a simulation of geomorphology like an eroding river valley, a sand dune, or the formation of mountain ranges, owing to their naturally prescribed, fixed spacial structure. Agent-based models, on the other hand, are usually employed to simulate pedestrian flows through cities, opinion formation in complexly networked social systems or ecosystems made up of multiple species with complicated behavior. In short, agent-based models often appear more suited to problems where the individual behavior of entities becomes more complex and interdependent, and their interaction partners are not fixed but depend on their inner state. It’s important to note that most of the aforementioned example systems often have some form of ODE representation as well, for instance, the Lotka-Volterra equations for predator-prey systems [Volterra, 1927; Lotka, 1910] or the SIR-model for disease propagation [Kermack et al., 1927]. Utopia does, however, not provide facilities of its own to solve ODEs because there is no shortage of high-quality ODE software in virtually any programming language used in science or engineering. Instead, the user has to rely on external tools to provide this functionality if so desired. Given programming language compatibility, however, Utopia does not hinder such integration. The aforementioned helper classes for cellular automata and agent-based models, respectively, are called `Cell-` and `AgentManager`. These manager classes are fed with a collection of entities and provide facilities to produce and remove them or apply rule functions defined in the model class to them. Their interface is designed with usability and simplicity in mind, placing the focus again on the design of the model itself and away from implementation details. To leverage the parallel processing performance of modern CPUs, the rule application can be optionally parallelized. Beyond that, an implementation for continuous Euclidean spaces has been provided, onto which different types of discretization can be layered (hexagonal or rectangular at the moment) with support for von Neumann- or Moore neighborhoods. Moving around agents on such grids is also automated via the manager class. As mentioned, many complex systems’ fundamental feature is their networked architecture, and so special attention has been given to supporting complex networks in agent-based models via the `Boost::Graph` [Siek et al., 2002], by making the manager classes and output facilities as far as possible agnostic to the data structure in which the system components are stored. Utopia thus supports grid-based, grid-less, and networked systems within one set of algorithms.

Returning to the modular architecture of complex systems mentioned above, it is worthwhile to take a look at how algorithmic model building proceeds. Here, this is thought of as proceeding through three steps, which are iterated to result in the final model; see also Figure 7.2. The focus, of course, is on complex systems. Naturally, a degree of knowledge has to be gained first to form hypotheses about the structure and dynamics of the system at hand, which results in a mental model that involves a separation of the natural phenomenon into different parts and subprocesses. Such a dissection enables us to identify entities and processes that are (thought to be) of relevance for the phenomenon to be modeled, discarding the rest (usually only to re-add them later once a deeper understanding has been gained). A priori, two roads open in front of us at this point. One is to build a monolithic model system that includes all the processes and entities into a single step of abstraction, a road often taken for relatively simple models and one that quickly closes when modeling complexity. The second way is not to aim at the big abstraction leap called for above but rather model the different subprocesses independently at first, coupling them together in a second step. While this quickly results in more complex models and more development work, it is arguably more powerful for cases where an abstract, full-scale representation is not attainable or which strongly depends on the details of their elements, as is the case for evolution, for instance. Utopia supports both approaches by allowing for models to be nested into each other, transforming them into "sub-" and "super-models", respectively. This results in each submodel, as well as the super-model, having their individual time-stepping algorithms, which need to be coupled

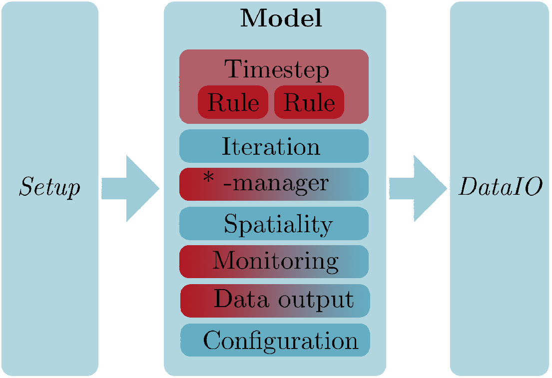


Figure 7.2: **Left**: We envisage modeling a complex system as a three-stage process. **A**: A real-world system is generally a continuum of processes linking entities on multiple scales. Studying these natural systems allows us to build mental models, or intuition, which allows the identification of entities and processes relevant to the phenomenon at hand. **B**: This can lead to a "monolithic" model in which processes and entities are directly linked into one dynamical system. While often desirable, this approach normally calls for a single large step of abstraction, which is not always possible. The alternative would be a modular approach that models individual components independently, linking them together by introducing explicit couplings in a second stage. While usually leading to more complex models, it often allows progress where a monolithic approach would not. **C**: The Utopia framework allows both approaches to be followed by allowing model instances to contain other model instances that are linked via the rules of the super-model. **Right**: A user-defined model is pre-structured via a "model" base class, which provides a variety of infrastructure code (solid blue), which partly has to be completed by the user to make it usable (gradient from red to blue), e.g., the manager classes for cells or agents need to be told how and when to construct and destruct entities in which configuration or the data-output code needs to be told what to write out. Blue arrows indicate the flow of data from the Setup module, which preprocesses configuration files, to the model class, which hands the generated data to the *Data I/O* module to write out. Both are operated from the model class via simple, predefined interfaces. The user's focus can thus stay with the definition of the rules for the system's dynamics (solid red), which must be bound into a time-stepping function that can be iterated to produce a trajectory.

in some way. How this nested dynamics proceeds is under full user control. In the same way as the user defines for a single model how rule application proceeds within a single time step, in a super-model, they define when and how the rule application of each submodel proceeds and how and when these interface with each other to produce the coupled dynamics. Additionally, new submodel instances can be dynamically generated during runtime within a super-model.

7.3 The Setup/Control and Analysis modules

7.3.1 Setup and simulation control

Utopia provides sophisticated simulation control functionality. While the template library provided by the core module is very helpful in making model development more efficient, keeping track of all the parameters in the code and initiating every run by hand can still be tedious. The Setup/Control module of Utopia aims at overcoming these complications, making usage of computer models easier. The design goal is to allow users access to the power of sophisticated computer models even without any coding ability. The two main ingredients are a configuration system based on the YAML language and the provision of an easy-to-use command-line interface for running and managing simulations. Central to the module is the concept of a “universe”, i.e., a model instance (analogous to a set of natural laws acting on a collection of matter or energy in time and space) together with a specific set of parameters (fundamental constants) that determine the algorithmic rules a simulation follows. A collection of such universes is consequently called a “multiverse”. Multiverses arise when a model is run with multiple sets of different parameters. The model itself, however, cannot change within a multiverse. With sensitivity analysis being a major part of model development, multiverses arise routinely.

In Utopia, each model must have its configuration file written in YAML, which defines the model’s parameters together with their default values. Such parameters might, for instance, include the default grid size, energy influx per timestep, number of species, or random death rate of individuals. A universe is defined by writing a new YAML file that contains a subset of the model parameters in question. This file is then added as an argument when calling Utopia in the command-line interface. The Utopia Setup/Control module will then take the supplied configuration file, check it against the default model configuration, and replace all default parameters found in the supplied file with their respective values while keeping the rest. A level above the model configuration sits the default configuration of the entire Utopia framework, to which the same process is applied. This “framework configuration” supplies things like the number of parallel workers employed, the path to Utopia’s output directory, how often the simulation reports its progress and many more. All of these parameters can be overridden by the “universe file” supplied. This hierarchical structure reduces the amount of parameterization the user has to deal with, as things not pertaining to the model are kept out of sight until explicitly called for. A “universe” becomes a “multiverse” if the “sweep” tag is specified in the configuration. This tag is attached to a single parameter and tells Utopia to create a simulation, i.e., a “universe”, for each value of the parameter that follows. Multiverse runs are parallelized by default to make use of modern multi-core CPUs. This feature makes conducting sensitivity analysis of newly built models significantly easier. Finally, the Utopia command-line interface not only initiates model runs but also facilitates the creation of new models (by creating an appropriate folder structure and all necessary files), allows for quick overrides of parameters to help with debugging or model drafting, and automates the application of analysis routines to existing or newly created data. Therefore, it is arguably the biggest contributor to usability in Utopia.

7.3.2 Automated Analysis

The Analysis module arose out of observed commonalities in the data produced by models for complex systems. Most of the time, data produced by the models we employ are of a high-dimensional, complicated, interdependent structure. Analyzing such data, of course, has become a mainstay of modern data analysis software, so it does not present an obstacle in general. However, becoming proficient in any given analysis tool to the desired degree is no small investment in time and effort. The Analysis module aims at simplifying both analysis and visualization by taking over many of the repetitive tasks inherent to working with simulations. This includes the automatic loading of the universe- and multiverse data into appropriate data structures and the provision of default plotting functions for many common visualization tasks. High-quality visualization has been a priority during development. While appearing as a rather mundane task, it is also notorious for requiring a lot of time and for being hard to generalize. Luckily, when simulating complex systems, visualization tasks are more limited, mostly requiring facet grids or layering for visualizing parameter effects, grid- or graph plots for visualizing structures, or time-series plots. Utopia provides ready-made abstractions for these, which can be customized with little to no coding. When this does not suffice, custom plotting functions can be integrated into the module as well. Data visualization can be configured via the same YAML-based system that is also used to specify model parameters (see subsection 7.3.1), such that in many cases, little to no actual Python code needs to be written. On the analysis side, data processing is handled by forming a directed, acyclic graph of processing functions that are specified in a user-supplied configuration file. This combines with the visualization logic to form a system in which a complete data analysis and visualization pipeline can be specified within YAML configuration files. In this way, no code needs to be modified when the pipeline changes. The entire system is based on Python's xArray [Hoyer and Hamman, 2017], Dask [Rocklin, 2015], and Matplotlib [J. D. Hunter, 2007] packages, which are widely adopted in the community. All that being said, the usage of the data-analysis module is not forced upon the user, and individual solutions can easily be applied as well. This has been done in this thesis, for which the Julia language has been employed instead of Python to analyze and visualize simulation data.

7.4 The Data I/O module

Data handling is a mundane task compared to the creative challenge posed by model development. However, as with many mundane tasks, failures or inefficiencies makes themselves felt quickly and keenly, and so it deserves its place in this exposition. As mentioned before, simulations have now attained a status where they routinely complement and, at times, replace physical experiments. It is thus not surprising that the data they produce is of a similarly complicated nature. It is desirable to be able to group the actual simulation data together with relevant metadata (date, time, experimental parameters) such that different experiments can easily be identified and reproduced. Moreover, large amounts of data are routinely produced by modern simulation codes and need to be efficiently accessible to analysis tools while at the same time allowing for space-efficient storage and archiving. Furthermore, as analysis tools have radiated greatly over the last twenty years or so, a high degree of acceptance of the file format chosen and the reliability thereof is important. While the classical CSV files have the advantage of being human-readable and support for them is nearly universal, they lack the efficiency of access bundling of meta-data and need more disk space when compared to binary formats. Similar objections might be raised when considering XML, YAML, or JSON formats. Custom binary formats are a possible alternative for simple data but are difficult to efficiently implement and maintain when data becomes more complex and interrelated. We, therefore, chose the widespread HDF5 file format [Koranne, 2011], which is a well-established binary data format that allows for the bundling of simulation- and metadata, efficient access, and full user control. Furthermore, its structural similarity to a UNIX filesystem (see Figure 7.3) makes its usage intuitive, and its widespread use across the natural sciences has led to the ubiquitous availability of HDF5-packages in nearly every programming language in use in science. Finally, HDF5 natively supports parallelization.

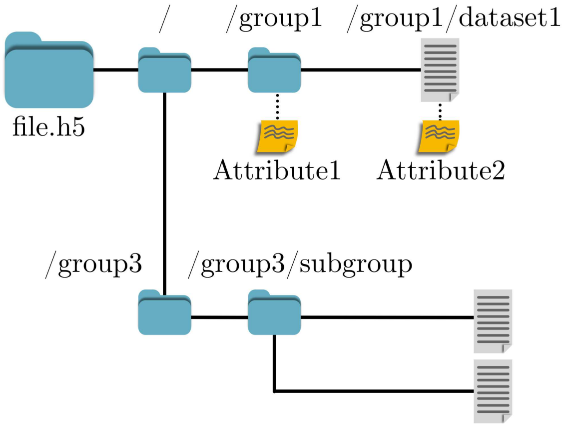


Figure 7.3: A HDF5 file works essentially like a UNIX filesystem, which contains multiple folders (named groups in HDF5) accessible through a path in a tree-like structure. These groups can be arbitrarily nested and contain an arbitrary number of files that hold the actual data (hdf5 datasets). Each group and each dataset can have metadata attached to them, which, for instance, give details about the experiment that produced the data. Datasets a priori can hold arbitrarily structured data and also can be compressed.

Utopia's Data I/O module consists of two parts: a low-level wrapper of the HDF5 library into modern C++, standard library compatible structures, and a high-level manager class that is designed to concisely execute complex data writing tasks. The design and implementation of this module have been my primary contributions to the Utopia framework and represent a large part of the work carried out over the course of this thesis. Other contributions have been made to the core module but to a lesser extent. Auxilliary functionality has also been contributed by Yunus Sevinchan [Sevinchan, 2018], who worked on automating parameter detection for efficient data output and provided helpful critique [Sevinchan, 2019].

7.4.1 Low-Level: A C++-HDF5 Wrapper library

The foundation of Utopia's data-I/O capabilities is formed by a set of custom C++ wrappers around a subset of the HDF5 C-library. While HDF5 is a very powerful library,

Listing 1 Example of using the plain HDF5-C-library to write an array to file.

```
#include <hdf5.h>
#include <vector>

int main() {

  // ———————————— Preparation ——————————————————————————
  //data to write to dataset. Pretend this to come from somewhere else.
  std::vector<int> v{1, 2, 3, 4, 5, 6, 7, 8, 9, 10};

  //data to write to attribute
  std::vector<int> a{12};

  // Actual size of the dataset to be allocated immediatelly
  std::vector<hsize_t> dims{10};

  // Attribute dimensions.
  std::vector<hsize_t> attrdims{1};

  // Maximum size of the dataset. Unlimited in this case.
  std::vector<hsize_t> maxdims{H5S_UNLIMITED};

  //chunksize for the data to be chopped up into.
  std::vector<hsize_t> chunks{4};

  // —————————————— Actual writing ————————————————————
  //Create a new, empty HDF5 file with default properties.
  hid_t file = H5Fcreate(”outfile.h5”, H5F_ACC_TRUNC, H5P_DEFAULT, H5P_DEFAULT);

  // Create a default group to put the dataset into
  hid_t group = H5Gcreate(file, ”/G1”, H5P_DEFAULT, H5P_DEFAULT, H5P_DEFAULT);

  // Create dataspace. We need only one because the data looks the same in file
  // and in memory
  hid_t space = H5Screate_simple(dims.size(), dims.data(), NULL);

  // Create the dataset named ”data” in group ”/G1”.
  hid_t dset = H5Dcreate(group, ”data”, H5T_NATIVE_INT, space, H5P_DEFAULT,
                         H5P_DEFAULT, H5P_DEFAULT);

  // Write the data to the dataset. Pretend this to have to be preprocessed
  // without changing the original.
  std::vector<int> v2;
  for (auto &&i : v) {
    v2.push_back(i * 2);
  }
  herr_t status =
      H5Dwrite(dset, H5T_NATIVE_INT, H5S_ALL, H5S_ALL, H5P_DEFAULT, v2.data());

  // Create an attribute to write a single integer into.
  hid_t attrspace = H5Screate_simple(1, attrdims.data(), NULL);

  hid_t attr = H5Acreate(group, ”attribute”, H5T_NATIVE_INT, attrspace,
                         H5P_DEFAULT, H5P_DEFAULT);

  status = H5Awrite(attr, H5T_NATIVE_INT, a.data());

  // close everything
  H5Dclose(dset);
  H5Gclose(group);
  H5Sclose(space);
  H5Sclose(attrspace);
  H5Aclose(attr);
  H5Fclose(file);

  return 0;
}
```

it is arguably not the most user-friendly, and using HDF5 directly from C is notoriously verbose and unidiomatic in C++. Writing out a simple array of numbers to HDF5 involves setting up not only the datafile (which naturally must always be done) but also the employed datatypes on disk and in memory, specifying the shape and topology of the array in memory and on disk and, possibly, the chunk size into which the data is chopped up when writing and possibly the compression rate, among other things. It was thus imperative for us to simplify and automate this process as far as possible. Before diving deeper, however, it must be noted that the Utopia Data-I/O library is by far not the only C++ wrapper around HDF5. Among others, there is a C++-03 version by the HDF-Group [The HDF Group, 1997-2022], H5CPP [Wintersberger et al., 2019] and HighFive [Devresse, 2016]. All of these, however, focus on comprehensive support of HDF5's features and thus provide a very general interface which, while being easier than the C-interface, nevertheless demands all the manual steps the latter does. For Utopia, with its focus on modeling complex systems, we went another way and focused on ease of use of a subset of HDF5 instead of full library support. This subset includes the output of arbitrary but fundamental datatypes (i.e., no structs) in one- and two-dimensional configurations, including data compression, attribute storing, and other standard HDF5 procedures possible on datasets. The dimensionality limit arises from the desired support for nested C++ containers, which, when nested deeper than two, limit the degree of possible automation we can attain. Therefore, the library provides an interface to the

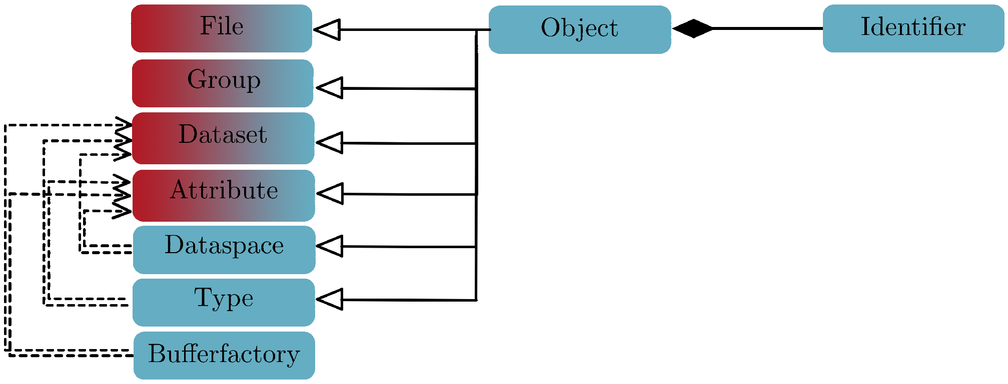


Figure 7.4: Simplified class diagram of Utopia's HDF5 library. Blue boxes are backend classes that the user needs not touch, while gradients from red to blue indicate that the class contains part of the user interface to the library. `Identifier` and `Object` handle resource acquisition from the underlying HDF5 C-library, while `Dataspace` provides the extent and topology of the data to be written, in memory as well as on disk. `Type` takes the data to be written and identifies a suitable HDF5 datatype to be used, while `Bufferfactory` does turn the user data (which might be nested and non-contiguous) into a contiguous representation that HDF5 can deal with. Finally, the user interface consists of a `File` class that provides the structure into which data is to be stored, the `Group`, which is analogous to a folder in a UNIX filesystem, and the `Dataset`, which handles the actual output/input operation. Finally, the `Attribute` provides access to the HDF5 Attribute concept, which allows the attachment of small but arbitrary data to `Datasets` and `Groups`. The connections follow UML convention: Filled diamond represents composition, empty triangle represents inheritance, and dashed arrow represents dependency.

C-implementation of the HDF5 file, group, dataset, and attribute structures, plus classes for handling resource acquisition, which in the C-interface of HDF5 naturally has to be done by hand (see Figure 7.4). Additionally, an algorithm to automatically determine the size of chunks of data into which the dataset should be split when using compression or when the final data size is initially unknown has been provided (contributed by Sevinchan [2018]), templated on an analogous implementation available for the Python HDF5 package [The HDF Group, 1997-2022]. This is especially important because the chunk size is one of the key factors determining the performance of input and output operations with HDF5.

Beyond that, the library design is guided by the character of complex systems modeling. Complex systems, in general, cannot be represented conveniently by simple arrays of numbers (simpler cellular automata notwithstanding), but their elements are modeled in terms of many kinds of data of different types and structure. It is rarely desirable to save the entire data structure representing an element. We would like to be able to run over the entities, extract arbitrary elements from each one, and possibly transform these in some way before finally writing to disk. Moreover, usage of the module should be similar to using the well-known standard library algorithms, which usually accept a begin and end-point into some range that contains data, along with a callback function determining what to do with that data along the way.

The design does significantly simplify usage of HDF5 but, at the same time, does not preclude expert users from implementing their custom solutions. See Listing 1 and Listing 2 for a comparison of plain C-HDF5 versus Utopia.

Listing 2 The same example as Listing 1, but with the HDF5 wrappers employed by Utopia.

```
#include <vector>

#include ”/path/to/Utopia/utopia/include/utopia/data_io/hdfattribute.hh”
#include ”/path/to/Utopia/utopia/include/utopia/data_io/hdfdataset.hh”
#include ”/path/to/Utopia/utopia/include/utopia/data_io/hdffile.hh”
#include ”/path/to/Utopia/utopia/include/utopia/data_io/hdfgroup.hh”

using namespace Utopia::DataIO;

int main() {

  // ---------------------- Preparation ------------------------------------------
  // data to write to dataset
  std::vector<int> v{1, 2, 3, 4, 5, 6, 7, 8, 9, 10};

  // data to write to attribute
  std::vector<int> a{12};

  //--------------------------- Actual writing -----------------------------------
  // Create empty datafile ’outfile.h5’
  HDFFile file(”outfile.h5”, ”w”);

  /*
   * Create group ’/G1’ in the file, open a dataset ’dset’ with unlimited
   * capacity inside it and write the data ’v’ to it, preprocessing
   * along the way via callback. Chunksize determined automatically
   */
  auto group = file.open_group(”/G1”);
  group->open_dataset(”dset”, {H5S_UNLIMITED})
      ->write(v.begin(), v.end(), [](auto &&i) { return i * 2; });

  // Create attribute to ’group’ which contains the ’a’ vector.
  auto attr = group->add_attribute(”attribute”, a);

  return 0;
}
```

7.4.2 High-Level: The Data Manager

Even with a relatively simple interface to the HDF5 library, data handling remains a complex task. Modern computer models of complex systems easily produce more data than is practical to handle on disk if the entire model is stored every time. Rather, typically, we would like to store different data with different resolutions, depending on their characteristic timescale of interest, especially if some kind of preprocessing, like filtering, is done within the simulation itself. Handling this in Utopia's model class, however, is cumbersome and laborious to maintain and adjust to different situations. For instance, some models implemented in Utopia produce data with more than twenty different variables, all written out at different intervals or under different conditions to different paths in the model's HDF5 file. It would, therefore, be much better to store all the necessary information (intervals to write at, data to write out) in a configuration file, provide the code for how to extract desired data from the system, and then let an internal class handle everything else. Not only does this prevent the user from writing all the code themselves, repeating errors in the process, but it also allows them to describe and handle very complex data-output tasks concisely and efficiently. In Utopia, the `DataManager`-class fulfills this role. The system is designed around the idea that all the users should need to specify when to write what to which place. This is conceptualized as a `WriteTask`, i.e., a structure that associates a data source with a data sink augmented with information on when to move data between them. In HDF5 terms, this means that five separate tasks have to be specified:

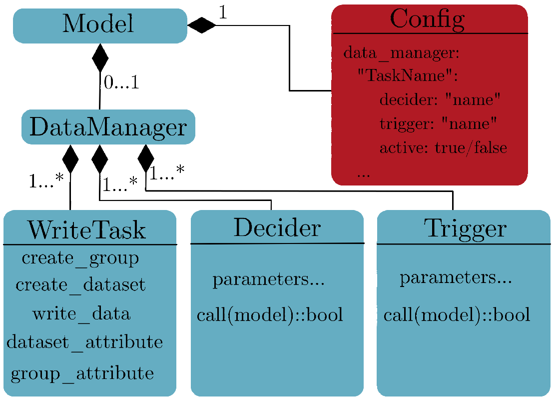


Figure 7.5: Simplified class diagram of the data manager module, following UML conventions. Blue indicates C++ code; red indicates YAML code preprocessed by the Python setup module of Utopia. Each model can have a `DataManager` instance, which contains lists of `WriteTask`, `Decider`, and `Trigger` objects. The `WriteTask` class contains all the facilities to write data to file, while the `Decider` and `Trigger` classes must at least be able to check the condition for data output (former) dataset creation (latter). Tasks and decider- and trigger objects are associated via the configuration file upon construction of the data manager.

- In what HDF5-group to store the dataset to be created
- In what dataset to store the data and whether it should be compressed or not
- How to move data from their source to the sink, i.e., the dataset. This can include preprocessing
- What attributes should the group should have, e.g., some string specifying what data is stored in there and from which model it comes

- What attributes the dataset should have, e.g., to what timestep it belongs

These five elements are grouped together in a structure and are represented by callback functions. The user needs to give a list of these structures to their model upon construction. What remains to be added is the information on when to execute them, i.e., when to build a new dataset or group and when to write data. For this, functions need to be implemented that return true whenever a given condition is met, e.g., an interval between timesteps has passed. Utopia provides a range of predefined functions, and the user can add more if desired. In principle, there can be arbitrarily many such functions checking arbitrary conditions on the model the data manager works on. In practice, however, the predefined functions all pertain to the timestep of the model in question because this is by far the most frequent variable conditioning data output. These functions come in two flavors, “triggers”, which trigger dataset construction, and “deciders”, which decide when to write out data, i.e., when to call functions two and three, respectively, in the list above. Secondly, these functions need to be parameterized (for example, a general interval decider function is provided by Utopia, but it has to be given when to start and stop working and in what intervals to return true) and associated with a `WriteTask`. Parameterization and association are done in the configuration file that defines a Utopia-Universe/Multiverse. The `DataManager` class is fully customizable and can, in principle, be used with other libraries as well, i.e., not just HDF5. With all the facilities in place, however, this need never has arisen in our experience. For an overview of how `DataManager`, `Model`, and configuration work together, see Figure 7.5. In keeping with the Utopia philosophy of not restricting the user in their creativity, usage of the data manager is optional.

8. *Amee*: A model for eco-evo-devo

In this chapter, the computer model, called *Amee*[1] developed as part of this thesis will be discussed. The model represents, to the best of my knowledge, the first agent-based model that unifies developmental plasticity, phenotypic plasticity, and niche construction together with an independent building block system based on Net-Rewriting-Systems [Llorens, Oliver, et al., 2004; Llorens and Oliver, 2004] and Petri-Nets [Petri, 1962] in a spatial setting. Each building block and several individual combinations up to and including a combination of plasticity, niche construction, and evolution do exist in the literature, however, a comprehensive system on the scale of *Amee* has not yet been developed. We proceed by first giving a short overview of the literature on modeling these processes, before presenting the conceptual architecture of the model. Finally, we dive into the implementation of this architecture with Net Rewriting Systems.

8.1 Background and previous results

The following attempts to list some of the work that has been done in modeling plasticity, eco-evolutionary dynamics, and development as well as their interaction amongst each other and with other effects, mainly genetic adaptation. Given the size of the field, this list is naturally incomplete, so its goal is more to show the broad lines that have been considered than detailed results. Computer models of evolutionary systems have become increasingly common over the past decades, paralleling the growing prevalence of computing technologies in other parts of society at large.

Many of these models are based on ordinary differential equations. K. N. Laland, F. J. Odling-Smee, et al. [1996] present a system based on two-locus theory in which the fitness consequences of niche construction were investigated. By using two-locus theory without niche construction as the baseline for comparison, they found a variety of ways in which niche construction influences overall fitness outcomes. These include evolutionary inertia, i.e., a tendency to keep evolving towards a phenotype although independent environmental selection would favor a different one, over unexpected presence or absence of polymorphism, to the fixation of deleterious alleles.

In a follow-up article [K. N. Laland, F. J. Odling-Smee, et al., 1999], they investigate the influence of independent renewal and depletion of the constructed resource and find that these independent resource dynamics influence the occurrence of polymorphism compared to K. N. Laland, F. J. Odling-Smee, et al. [1996], as well as the fixation probability for normally deleterious alleles. Extending this line of work, Silver and Paolo [2006] add spatiality in two dimensions. Their findings include a significant increase in the

---

[1] **A** **m**odel for **e**co-**e**vo-devo. Name for the agent-based computer model presented in this thesis.

frequency of niche construction through spatial effects and an increase in the range of initial conditions for which niche construction alleles become fixed compared to the non-spatial model. Moreover, in a situation where two species engage in opposing niche construction on the same resource, the chance for stable polymorphism is increased by clustering effects. In concordance with K. N. Laland, F. J. Odling-Smee, et al. [1996], they find that niche construction can lead to the fixation of otherwise deleterious traits via genetic linkage disequilibrium. This suggests that niche construction may be able to increase the genetic variation in a population and with it its ability to respond to changing conditions. Finally, the presence of niche construction dampens the effects of external perturbations, an effect they call "environmental homeostasis". They suggest several further research directions, among them the extension to a multi-species model, different implementations of spatiality, and an analysis of the sensitivity of their results to the strength of the association between genetic and ecological inheritance. Silver and Paolo [2006] is of particular interest because of its emphasis on the potentially stabilizing effects of niche construction on ecosystem structure in a spatial system. Another important contribution has been made by Krakauer et al. [2009]. The authors investigate the tradeoff between investment in fitness-enhancing niche construction versus investment in reproduction. From this situation, the "tragedy of the commons" ensues for a costly constructed niche that can only incompletely be monopolized. The authors emphasize that their model does not rely on frequency-dependent effects like the one employed by K. N. Laland, F. J. Odling-Smee, et al. [1996]. Rather, it treats the niche affected by the organisms as a dynamical system in its own right instead of subsuming it into a variable in the equations governing population density. They furthermore forego the modeling of genetic details present in the preceding work, and focus on the ecosystem-level effects, with genetics and inheritance mechanisms being abstracted away. Therefore, Krakauer et al. [2009] comprises almost a complementary approach to the line opened by K. N. Laland, F. J. Odling-Smee, et al. [1996]. Their model consists of an ODE system that describes the rate of change for the density of two populations exploiting a single niche that changes through construction and niche decay, i.e., the loss of construction effects over time. They introduce three different effects of niche construction: construction that increases carrying capacity, construction that increases fecundity, and construction decreasing mortality. They find that, in general, niche construction is only viable if the constructed niche can be at least partially monopolized by the constructing organisms. If so, intermediate levels of niche construction dependent on life history and competitive regimes evolve, although the investment into niche construction increases with increased monopolization. Especially the dependency on life history is of interest here, because in nature, niche construction, or ecosystem engineering, is often dependent on a specific phase in the lifecycle of organisms. They note that their general assumption of a trade-off between investment into reproduction and investment into niche construction might be violated in reality, and construction may increase the available resources for reproduction in many species.

Indeed, niche construction and the problem of cooperation seem intuitively related, given that niche construction is a costly process the results of which are often accessible to other organisms, thus leading to a public goods problem. Based on previous work that showed how cooperation could arise by genetic hitchhiking on advantageous mutations ("Hankshaw effect") [Hammarlund et al., 2016] in an agent-based model based on a meta-population structure, Connelly et al. [2015] explored how negative niche construction can lead to persistence and even invasion of cooperative alleles by continuously providing adaptive opportunities that keep a cycle of adaptive hitchhiking going. Interestingly, this discusses *negative* niche construction, i.e., a kind of change to the environment that reduces the fitness of the constructing population. Indeed, the invasion of cooperators, even when cooperation yielded high adaptive benefits, hinged on the continuous change of the environment through niche construction and thus on organisms being adaptively limited.

Taking an ecological point of view, Gilad et al. [2004] investigate another ODE model of two ecosystem engineers in drylands (plants and crust forming cyanobacteria) with a focus on habitat creation performed by these engineer species. They find that species richness depends a lot on the patterns formed by the engineer species, i.e., the plants. They propose a mechanism for species loss based on a change in formed patterns as a response to an environmental change.

T. Taylor [2004] build upon a system presented in Rocha [2001] to construct an agent-based model for niche construction in evolving populations, with a focus on the evolution of complexity. The environment is represented as an explicit dynamical part of the system. Phenotypes of organisms and environmental properties are implemented in analogous ways. The central finding of their analysis is that assuming the organism found a way to circumvent the error-threshold [Eigen and Schuster, 1977], niche construction is sufficient to introduce an autonomous increase in complexity over time. It must be noted here that "complexity" was interpreted as genome length, thus ignoring much of the conceptual complexity inherent in the term. Additionally, because the model features explicit implementations of environment and phenotypes and thus makes more assumptions about their structure than we encountered in the ODE-based systems before, these results must be interpreted within the confines of these assumptions. While we are dealing with algorithmic modeling in this thesis, a number of inspiring analytical results must be mentioned. Lehmann [2008] has contributed an analytical approach in which they analyze the selection of niche construction in populations with different mobility and spatial segregation. They introduce the concept of a "posthumous extended phenotype", i.e., a trait expressed beyond the lifetime of the organism carrying the effecting allele, and assume spatiotemporal homogeneity and weak selection. They find that selection on costly niche construction traits can feed back to the evolution of these traits through indirect inclusive fitness effects that extend over generations. Because they base their model on inclusive fitness theory [Hamilton, 1964], population relatedness is an important factor. Too high mobility in space does reduce selection for niche construction because it reduces relatedness in local populations. Among other topics, they suggest that ecological inheritance may be an important driver for altruistic niche construction, e.g., in social insects. Furthermore, they investigated the evolution of resource utilization restraint by introducing a resource utilization rate as a new trait of organisms, and interpret this as a niche construction trait. The authors found that restraint in current resource utilization can evolve if this results in higher availability of the resource in the future, given that the available amount of the resource does not only depend on the current state of the system but includes also past states, i.e., that ecological inheritance exists. Because resource-consumer systems are the basis of ecology, this is a fundamental finding that places the focus away from direct intra-generational interactions and highlights the temporally extended nature of evolving systems brought about by inheritance. This started an ongoing line of work that mainly investigates eco-evolutionary dynamics [Mullon and Lehmann, 2018].

Staps et al. [2019] modeled the emergence of multicellularity, including a representation of gene-regulatory networks. The emergence of multicellularity is of special importance because it is arguably the point at which development, as we understand it today, emerged. A key element is that multicellularity requires spatiotemporal organization of cell differentiation, and it is thus relevant for the origin of development. They find that diverse lifecycles emerge depending on ecological conditions and ancestral constraints.

In another line of work, Scheiner [2014] started by investigating how developmental instability evolves as a form of bet-hedging in a spatial system, where it was found that developmental instability evolves as a reaction to spatiotemporal heterogeneity. Later, they investigated the role of habitat choice and heterogeneity on the evolution of phenotypic plasticity as an alternative to genetic differentiation and jack-of-all-trades strategies. A similar investigation was made with respect to niche construction [Scheiner, Barfield, et al., 2021a], where it was found that niche construction was often favored over habitat choice, but that in more detail, the eventual outcome depended a lot on life-history details and ecology. Later, they added plasticity as an alternative to niche construction for coping with environmental spatiotemporal heterogeneities [Scheiner, Barfield, et al., 2021b], and found similarly that life-history effects had a great influence on whether plasticity or construction traded-off against each other or which one was favored, but also that construction was favored when direct offspring benefited from it, in agreement with what Krakauer et al. [2009] found before. Finally, in [Scheiner, Barfield, et al., 2022] they investigated how adaptive niche construction evolves in comparison to non-adaptive niche change, and show again that construction evolves the most when it is monopolized by a lineage, while at the same time, it is hampered when movement pattern and environmental spatial heterogeneity are mismatched.

Ng and Kinjo [2022] reviewed past modeling efforts for plasticity-led evolution, mostly based on the model for genetic regulation introduced in A. Wagner [1996]. They review several works that extended this model and suggested a number of improvements and future research directions, among them the effects of hierarchical network structures in GRNs on plasticity, or the breakdown of robustness in relation to environmental cues. Crucially, the concept of plasticity-led evolution relies on the assumption that plastic responses are a universal feature of developmental processes, which, while plausible, has not been empirically demonstrated sufficiently.

In light of subsection 2.4.1, we focus on evolution and learning for a moment, and start with R. A. Watson, G. P. Wagner, et al. [2014]. In their work, they build on the Wagner model to show how such systems can learn to express multiple different phenotypes even when presented with incomplete or erroneous genetic information, analogously to Hebbian learning and demonstrated some generalization capabilities of these systems. Indeed, they point out that models for GRNs as used in the literature are mathematically equivalent to recurrent artificial neural networks. Their results emphasize the analog between Hebbian learning in neuronal systems (“fired together, wired together”) with directional selection on a system of reciprocally influencing gene expression. Ng and Kinjo [2022] discusses further research in these directions.

8.2 On modeling evolution and other complex adaptive systems

8.2.1 Required elements for comprehensive modeling of open-ended evolution

When trying to model evolving organisms capable of open-ended evolution, we need to surmount four challenges:

- Find a substrate able to generate complex enough structures to allow for open-ended adaptability to diverse conditions. This is primarily concerned with the rules to which elements of this substrate, e.g., organic molecules, must adhere to when they interact. In other words, we need to find a substrate capable of forming complex adaptive systems.
- Find a way to create the three parts required by a cellular system (copier, constructor, genotype) from the substrate. This part can have many levels of abstraction, ranging from representations of the self-organization from which these processes emerge to a designed organismal structure that assumes them to be present in some predetermined way.
- Find some encoding system for the structure of organisms.
- Find representations for the interactions of organisms with their environment and each other from which natural selection can emerge.

Agent-based models are a natural way to model biological evolution, given that they intrinsically are able to represent finite populations of entities acting in arbitrary ways. It is important to distinguish approaches that allow natural selection to emerge on the basis of artificial ecological interactions from those that assume natural selection to act in a certain way, i.e., for which some notion of fitness is an intrinsic modeling property. While the latter approach makes a system easier to analyze the former allows for a better understanding of mechanisms that underly natural selection, even if they arise in a very simple system compared to nature.

John von Neumann succeeded in creating an artificial abstract representation of natural organisms using cellular automata [von Neumann, 1966], which however is very complex and computationally demanding. Thus, even a functional representation of the basic requirements for open-ended evolution is no guarantee that the system will show a range of phenomena comparable to natural systems. We will take this point up again below in subsection 8.2.2 and subsection 8.2.3.

An interesting variation is represented by artificial life simulations in which a domesticated form of computer-virus-like entities has been employed, e.g., the platforms Tierra [Cisternino, 2021] and Avida [Adami et al., 1994; Lenski, Ofria, et al., 2003], each of which shows a wide array of phenomena also found in natural populations, in particular parasitism and the evolution of complex structures.

8.2.2 Difficulty of analytical modeling

Model building is usually thought to pursue the goal of prediction, i.e., a model can be parameterized according to natural observation and for the corresponding initial condition will produce a trajectory consistent with its natural counterpart. Such modeling is usually associated with analytical approaches. This power and generality however can only be achieved for comparatively simple systems where there are few agents or simple interactions. Indeed, there is little hope for an analytical "theory of everything" in complexity or biological evolution that could be formulated as a set of equations, as we know it from physics, owing to their nature: Their networked, dynamic, and self-modifying dynamics largely preclude fundamental, predictive analytical theories, and successful modeling of subprocesses like adaptive dynamics or population genetics notwithstanding. As the last few chapters have attempted to show, there is also considerable uncertainty about the phenomenological breadth and causal relationships of processes in biology, with observations being difficult and time-intensive, and experiments in the lab often altering the system into unnatural states [Sultan, 2015].

Indeed, what sets biological evolution and complex adaptive systems apart from physical systems is that they are not dominated by the natural laws that underly their building blocks (as is the case with systems that are studied in, e.g., statistical physics). Rather, they unfold in the typically large space of self-organized, emergent structures and processes they are capable to unfold into on the basis of these laws. Thus, a classical reductionist approach that tries to reduce their phenomenology onto a set of single behavioral laws their elements must adhere to will not tell us much about the unfolding of the system. Combined with their nonlinearity [S. A. Levin, 2002] and the fact that for biological evolution, we only have incomplete records of a single trajectory, this pretty much limits analytical modeling to cases where the dynamics is strongly reducible.

8.2.3 A switch to an engineering approach in complex systems research

As an alternative to classical modeling, an "engineering approach" can be used to help gain an understanding of the nature of such systems. This approach consists of the creation of artificial representations of natural systems that do not emulate them in every detail, but aim to follow a selected subset of their observed dynamical rules in an abstract setting. From the unfolding of these systems, we then can learn how certain processes are able to interact under given conditions. Comparison with natural observations or other artificial systems will then allow to discard or retain candidates for natural process assortments. Computer simulations and agent-based models are uniquely suited for these tasks because we have complete control over them and they are comparatively easy to manufacture, change and explore. In other words, this approach engages in a simplified form of reverse engineering. Other than computer models, increasingly "artificial life" is used to study biological processes [Ebrahimkhani and M. Levin, 2021].

While similar to classical empirical research, it must be noted that this approach does not directly "question nature". Rather, by implementing the principles of the system as we understand them, we can test our theories and identify gaps in our knowledge in the absence of having a full instance of the system to experiment on. This modeling approach is labeled as "engineering" because of its pragmatical nature – processes or causal structures that are producing plausible or observationally confirmed results are retained and others discarded, but without a fundamental reductionist theory that guides these decisions.

Ideally, such models allow us to learn something about the structure of an evolutionary process as an abstract system, thus identifying phenomena to investigate empirically and exploring processes that may have taken place in the past but are not represented in today's biosphere or take place over too long times, e.g., the origin of life, adaptive radiations or major evolutionary transitions in individuality.

In some more detail, we must keep in mind that when building such models, we do not only design a model for the process itself, but unavoidably also the background in which it is embedded - we must make choices on the presence and nature of energy, conservation laws, and generally the interpretation of all dynamical entities and processes in the system. In a manner of speaking, we design its “laws of nature”. Therefore, the results will be influenced by these choices, and cannot be expected a priori to follow naturally observed trajectories where the underlying natural laws and capabilities for self-organization and emergence significantly differ from the natural one. *Amee* falls into this class of models, and thus has more of an explorative and descriptive than predictive character.

8.3 Petri-nets and Net rewriting systems as tools for modeling evolution and ecology

8.3.1 What is a Petri-net?

The Petri-net formalism forms the basis of *Amee* and thus will be introduced first. A Petri-net [Petri, 1962], also called place-transition-net, is a type of discrete model for distributed, concurrent systems. It takes the form of a weighted, bipartite, directed graph, in which there are two kinds of nodes:

- “Places” represent “passive” elements of the system that exhibit a state. The set of states on each place in the network gives the system state.
- “Transitions” represent “active” elements of the system. They transform the state of the places they are linked to.

Petri-nets can be used as a model of computation [Petri, 1962; Zaitsev, 2012]. In this thesis, however, Petri-nets are used as the underlying formalism of a model of developing, evolving self-replicators that have the capacity to influence their environment.

Following Reisig [2013], for discrete state variables, the state of a place can mathematically be modeled by a multiset over some basic set of entities, which are called “tokens”, which are drawn from a set that is initially chosen for the Petri-Net in question. Places and transitions are linked via arcs, which carry weights. These weights in turn are represented by multisets over the same basic set of tokens as the place states. Henceforth, the state of a place, i.e., its token content, shall be called “marking”.

More formally, a Petri net is specified as follows: Be U a set, with the set of multisets over U called $\mathcal{M}$. A Petri Net (“PN”) is a tuple $((P, T, F, M, W), M_0)$. The tuple (P, T, F, M, W) has the following properties:

- P a finite, countable set of places.
- T a finite, countable set of transitions.
- Because Petri nets are bipartite, it holds that $P \cap T = \emptyset$.
- $F \subseteq (P \times T) \cup (T \times P)$ a relation that describes the set of arcs, each of which links either a place to a transition or a transition to a place, but never a place to a place or a transition to a transition.

- $M : P \to \mathcal{M} : p \in P :\mapsto M(p) \in \mathcal{M}$, a function that assigns a marking to a place.
- $W : F \to \mathcal{M} : f \in F :\mapsto W(p,t) \in \mathcal{M}$, a function that assigns a weight to an arc.

$M_0 \subset M(P)$ is called the "initial marking" of the Petri net. The subtuple (P, T, F) is called the "net structure" of the PN^2, with F being called the "flow relation".

8.3.2 Mathematical formulation of Petri Nets

In the following, some specialized terminology is needed, which shall quickly be introduced here. For this chapter, we assume the markings and weights employed to be real numbers, but the system can readily be extended to any kind of multi-set marking [Reisig, 2013]. Be $((P, T, F, M, W), M_0)$ a Petri net and $x \in T \cup P$:

- The set $\bullet x = \{y \in T \cup P : (y, x) \in F\}$ is called the "pre-set" of x.
- The set $x \bullet = \{y \in T \cup P : (x, y) \in F\}$ is called the "post-set" of x.
- A transition t is "enabled" if and only if $\forall p \in \bullet t : M(p) \geq W(p,t)$. This condition will become important later in the formulation of *Amee*.
- Given index sets I, J that index P, T respectively, the marking of the net can be represented by a vector $(\mathbf{m})_i = M(p_i)\, \forall p_i \in P$.
- Given such indexing, the flow relation can be represented by a matrix (the "adjacency matrix") of the system, which has $|P|$ rows and $|T|$ columns and the elements of which are the sum of the weights of arcs that connect place p_i and transition t_j:

$$(\mathsf{A}_{i,j}) = \sum W(p_i, t_j) \tag{8.3.1}$$

It is convenient to decompose A into two matrices, the input matrix I and the output matrix O of the system:

$$(\mathsf{I})_{ij} = \begin{cases} W(p_i, t_j) \ \text{iff}\ p_i \in \bullet t_j \\ 0 \ \text{otherwise} \end{cases} \tag{8.3.2}$$

$$(\mathsf{O})_{ij} = \begin{cases} W(p_i, t_j) \ \text{iff} p_i \in t_j \bullet \\ 0 \ \text{otherwise} \end{cases} \tag{8.3.3}$$

- A transition t_j is enabled if and only if the marking of all places p_k in its pre-set $\bullet t_j$ are greater or equal to the weights of the arcs that link p_k to t. The status of the $|T|$ transitions in the net can also be represented by a vector $\mathbf{e}$:
- $$(\mathbf{e})_j = \begin{cases} 1 \ \text{iff}\ \forall p \in \bullet t_j : M(p) \geq W(p, t_j) \\ 0 \ \text{otherwise} \end{cases} \tag{8.3.4}$$
- A transition t_j that is enabled can "fire". When a transition fires, it reduces the marking of all places in its pre-set by the weights of the respective adjacent arcs and correspondingly increases the marking of all places in its post-set.

---

[2] Petri-net.

- The firing of at least one transition comprises a "step" of the PN and transforms the marking of the net $\mathbf{m}_t$ into another marking $\mathbf{m}_{t+1}$. This comprises the dynamics of the PN and can be formulated as:

  $$\mathbf{m}_{t+1} = (\mathsf{O} - \mathsf{I})\,\mathbf{e}_t + \mathbf{m}_t \tag{8.3.5}$$

  with matrix-vector arithmetic being defined as usual.

- Usually, the index τ corresponds to time and the net dynamics represent the system's temporal evolution. Note that multiple transitions can fire at once, in keeping with the intention of modeling distributed, concurrent systems. As the net progresses through different steps, different transitions become enabled, so $\mathbf{e}$ changes with time.

- We define the input interface (output interface) of a Petri net as the set of places that have no incoming (outgoing) transitions. They can be represented by boolean vectors $\mathbf{in}, \mathbf{out}$, computed as

  $$(\mathbf{in})_i = \{i \in P | \mathsf{O}_{ij} = 0 \;\forall\, j\} \tag{8.3.6}$$

  $$(\mathbf{out})_i = \{i \in P | \mathsf{I}_{ij} = 0 \;\forall\, j\} \tag{8.3.7}$$

  $$\tag{8.3.8}$$

  A PN for which at least one of these sets is non-empty we call "half-open" when both are non-empty we call it "open".

A final marking is a marking for which no further transitions are enabled. Finally, Petri-nets are concurrent, therefore multiple transitions can be enabled at the same time (see Equation 8.3.5).

8.3.3 Petri-Nets with inhibitor arcs

There are various extensions of the basic Petri-net formalism, of which only one shall be discussed here. This extension adds a second kind of arc to the net that can connect only places to transitions, not vice versa, and which acts as a logical negative: Only when the attached place has a marking of zero will the adjacent transition be enabled, irrespective of any other conditions active on it. Such arcs are called "inhibitor arcs", and endow Petri-Nets with additional functionality. Petri-Nets with inhibitor arcs can be shown to be equivalent to Turing machines [Zaitsev and Li, 2018]. We formally indicate an inhibitor arc by a weight of -1 (or an equivalently unique symbol for weights not represented as real numbers) in the input matrix. For simplicity of notation, we assume real-valued markings and weights here, understanding that the formalism is readily extended to multisets. The vector $\mathbf{e}$ for IPN[3] therefore changes to:

$$(\mathsf{e})_j = \begin{cases} 1 \text{ iff } \forall p \in \bullet t_j : M(p) \geq W(p, t_j) \;\wedge W(p, t_j) \;\neq -1 \\ 0 \text{ otherwise} \end{cases} \tag{8.3.9}$$

---

[3] Petri-net with inhibitor arcs.

Accordingly, we need to modify Equation 8.3.5 in order to make sure that the system will correctly ignore the inhibitor arcs when computing the change in marking. To this end, we introduce the matrix N as

$$(\mathsf{N})_{ij} = \begin{cases} 1 \text{ iff } (\mathsf{I})\,(i,j) \neq -1 \\ 0 \text{ otherwise} \end{cases} \tag{8.3.10}$$

Because inhibitor arcs only ever lead from a place to a transition, we only need to be concerned with the input matrix I. The matrix N tells us if a particular input arc is an inhibitor arc or not. With this, the update equation for the marking of the network becomes:

$$\mathbf{m}_{t+1} = (\mathsf{O} - \mathsf{I} \odot \mathsf{N})\,\mathbf{e}_t + \mathbf{m}_t \tag{8.3.11}$$

with $\mathsf{A} \odot \mathsf{B}$ denoting the elementwise product of the two matrices A and B. In this way, the -1 entries in the input matrix I are eliminated from the computation, while the effect of the inhibitor arcs is built into the vector $\mathbf{e}$.

8.3.4 What is a Net rewriting system?

Net rewriting systems form the highest layer of dynamics in *Amee*, while PNs form the middle layer.

Net rewriting systems are a combination of the concepts of Petri-nets and graph rewriting systems, as presented by Llorens, Oliver, et al. [2004]. The presentation below is a somewhat shorter version of theirs but follows it closely otherwise. The idea is that a Petri-net is transformed into another Petri-net by the successive application of transformation rules. These map a subset of elements of the original net to a set of new elements which together with the unchanged elements of the original form a new Petri-net.

Mathematically, a net rewriting system $\mathcal{N}$ is a pair $(\mathcal{R}, (\Gamma_0, M_0))$, with $\mathcal{R} = \{r_1, r_2, \ldots r_m\}$ a finite set of rewriting rules and (Γ_0, M_0) a marked Petri-net as defined in subsection 8.3.1. A rewriting rule r is a pair $(L, R, \tau, \bullet\tau, \tau\bullet)$ with the following properties:

- $L = (P_L, T_L, F_L)\,, R = (P_R, T_R, F_R)$ are Petri-nets, being called the left-hand side and right-hand side of r.
- $\tau \subseteq (P_L \times T_L) \cup (P_R \times T_R)$ is a binary relation that relates places in L to places in R and transitions in L to transitions in R. τ is called the *transfer relation* of r.
- The subrelations $\bullet\tau \subseteq \tau, \tau\bullet \subseteq \tau$ are called the *input interface transfer relation* and *output interface transfer relation*, respectively. These are used to fix the connections between the old part of the net and the replaced part.

The *configuration* of an NRS[4] is a Petri-net $\Gamma = (P, T, F)$, and a *state* is a marked Petri-net (Γ, M), with the state (Γ_0, M_0) representing the initial state. We call an *event* either the firing of a transition $t \in T$ or a rewriting rule $r \in \mathcal{R}$.

In order to proceed, we need one more definition: Be $\Gamma = (P, T, F)\,, \Gamma' = (P', T', F')$ two Petri-nets. We call Γ' a full embedding of Γ if and only if there exists an injective map $f : P \cup T \to P' \cup T'$ with $f(P) \subseteq P', f(T) \subseteq T'$ such that for $x, y \in P \cup T : F\,(x, y) = F'\,(f(x), f(y))$. The image of f is called a full subnet of Γ.

With this, we can define a *step* of a Net rewriting system as

---

[4] Net-rewriting system

- The firing of a transition in the state (Γ, M) as presented in subsection 8.3.1.
- The replacement of the subnet L with the subnet R.

The latter part proceeds as follows. Be (Γ, M) a state of the NRS $\mathcal{N} = (\mathcal{R}, (\Gamma, M))$ as described above, and $r \in \mathcal{R} = (L, R, \tau, \bullet\tau, \tau\bullet)$ a rewriting rule.

- Find a full embedding Γ' of L in Γ with $f : L \to \Gamma$ as defined above, such that $\forall x \notin L, y \in L$: $x \in \bullet f(y) : (x, y) \in \bullet\tau$ and $x \in f(y)\bullet : (y, x) \in \tau\bullet$.
- Replace the subnet L with R such that the net Γ is turned into the net Γ_+, with $x, y \in P \cup T$. We write $x \in R$ as shorthand for $x \in P_R \cup T_R$.

$$P_+ = (P \setminus P_L) \cup P_R \tag{8.3.12}$$

$$T_+ = (T \setminus T_L) \cup T_R \tag{8.3.13}$$

$$F_+(x, y) = \begin{cases} F(x, y) \text{ if } x, y \notin R \\ F_R(x, y) \text{ if } x, y \in R \\ \sum_{z \in \bullet\tau y} F(x, f(z)) \text{ if } x \notin R \ \wedge \ y \in R \\ \sum_{z \in \tau\bullet x} F(f(z), y) \text{ if } x \in R \ \wedge \ y \notin R \end{cases} \tag{8.3.14}$$

where the latter two cases describe the connection between replaced elements and the rest of the net.

We used the following notation for relations $\tau\bullet, \bullet\tau$:

- $\tau \bullet x = \{y \in R | (x, y) \in \bullet\tau\}$
- $x \bullet \tau = \{y \in R | (y, x) \in \bullet\tau\}$

Finally, the marking of the new places is given by the markings of the old places:

$$M_+(p) = \begin{cases} M(p) \text{ if } p \notin R \\ \sum_{q \in \tau p} M(f(q)) \text{ if } p \in R \end{cases} \tag{8.3.15}$$

They further expand this system [Llorens and Oliver, 2004] by augmenting the rewriting rule r with a set of control places $C \subseteq P_L$ and a control marking $\mathbb{M} : p \in C \mapsto M(p)$: $r = (L, R, \tau, \bullet\tau, \tau\bullet, C, \mathbb{M})$. The control marking gives the minimum marking for places in C such that the rule is enabled. So the conditions for a rewriting rule to be enabled (first bullet point above) are expanded with the additional condition that $\forall p \in C : M(f(p)) \geq \mathbb{M}(p)$, i.e., for all places in the control set the current marking must be greater or equal to the control marking. Control markings are useful for modeling industrial processes for example, but are otherwise ignored here.

8.3.5 Basic building blocks and NRS setup in *Amee*

Organisms and the environment are modeled as (half-) open IPNs that process matter and energy, with the former being interpreted as the metabolism network and the latter as part of a resource-consumer network with abiotic elements, as they occur in nature. Organisms form a “focal population” which forms a node in this network by virtue of their members consuming and excreting resources. Organisms additionally are endowed with rewriting rules, that allow for dynamic changes to their habitat's IPN, but also to their own metabolism IPN. Therefore, building blocks must be found that can represent NRS and IPN. Because of the conceptual symmetry of organisms and environment as (half-) open IPNs for matter and energy, we can base both on the same fundamental building blocks, which are shown in Figure 8.1. These building blocks form thus the lowest level of organization that comprises *Amee*.

Fundamental building blocks Building blocks are minimal elements from which Petri-nets can be created by successively attaching them to each other, see Figure 8.2. We follow the scheme outlined in Chunikhin [2019] in the choice of these building blocks. We observe from Equation 8.3.11, that every entry in the input matrix I describes either an arc that goes from a place to a transition (symbolized as $\bigcirc \rightarrow \blacksquare$, with $\bigcirc$ representing a place and a $\blacksquare$ representing a transition), or an inhibitor arc, symbolized as $\bigcirc \multimap \blacksquare$. Analogously, every entry in the output matrix O represents an arc that goes from a transition to a place (represented by $\blacksquare \rightarrow \bigcirc$). Generalizing this, each entry in I or O can be a sum of such elements, except for inhibitor arcs, which always keep their value irrespective of added or subtracted elements (similar to multiplying by zero). We call these elements, $\bigcirc \rightarrow \blacksquare$, $\blacksquare \rightarrow \bigcirc$, and $\bigcirc \multimap \blacksquare$ "elementary symbols".

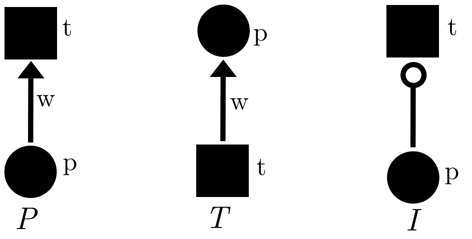


Figure 8.1: Fundamental building blocks ("elementary symbols") for an IPN are defined by the places p and transitions t an arc with weight w connects. These can be place-transition (left, P), transition-place (middle, T), or place-inhibitor-transition (right, I, note the circle and missing weight w to indicate the inhibitor property). From these, all possible IPNs for a given set of places and transitions and a given range of weights and markings can be constructed. These building blocks can be augmented with a marking m that is assigned to each place.

Therefore, for an alphabet of place labels $P = \{1, 2, ...N_P\}$ and transition labels $T = \{1, 2, 3, ...N_T\}$ the matrices I, O are in the vector space of matrices $\mathrm{Mat}_{N_P \times N_T}(W)$ with W being the set from which the weights are drawn. While W is in general a multiset, we still limit ourselves to real-valued weights for simplicity without loss of generality. We introduce symbols for each of the elementary symbols in order to formally represent them:

- $\bigcirc \rightarrow \blacksquare$: $P_p^t(w)$, $p \in P, t \in T$ representing the place and transitions (i.e., the matrix coordinates) and w the weight (i.e., the matrix entry).
- $\bigcirc \multimap \blacksquare$: I_p^t with the weight always being -1 or another flag that indicates an inhibitor arc.
- $\blacksquare \rightarrow \bigcirc$: $T_p^t(w)$.

For a set of such symbols, we represent each element in the corresponding matrices I, O as

$$
\begin{aligned}
(\mathsf{I})_{pt} &= \begin{cases} \sum_k P_p^t(w)_k \text{ if no } \ I \text{ symbol involved} \\ -1 \text{ otherwise} \end{cases} \\
(\mathsf{O})_{pt} &= \sum_k T_p^t(w)_k
\end{aligned} \tag{8.3.16}
$$

where the sum is meant to sum up the weights w_k of each symbol with place label p and transition label t.

Similarly, we can add the initial marking M_0 to the symbols: $P_p^t(w,m)$, $T_p^t(w,m)$, $I_p^t(m)$, such that the initial marking for each place can be constructed as

$$
M_0(p) = \sum_k P_p^{t_i}(w_i, m_k) + T_p^{t_j}(w_j, m_k) + I_p^{t_l}(m_k) \tag{8.3.17}
$$

where the sum this time is meant to add up the markings m_k of each symbol in the same way it has been done for weights above.

Construction of IPNs from elementary symbols Therefore, an arbitrary IPN over node labels $P \times T$ with weights in W can be represented as a string of elementary symbols $G_{P,T}$, and the matrices can be constructed according to Equation 8.3.16 and the initial marking according to Equation 8.3.17.

In order to simplify notation, we can put the "kind" of the symbol, $K \in \{P, T, I\}$ into the symbols themselves and only use the parentheses notation for everything, and use S as shorthand for all symbols, such that they are represented uniformly as $S(K, p, t, w, m)$. This allows us to formulate an algorithm to create the net (Γ, m_0) from $G_{P,T}$, described below in Algorithm 1. The process is shown in Figure 8.2 schematically.

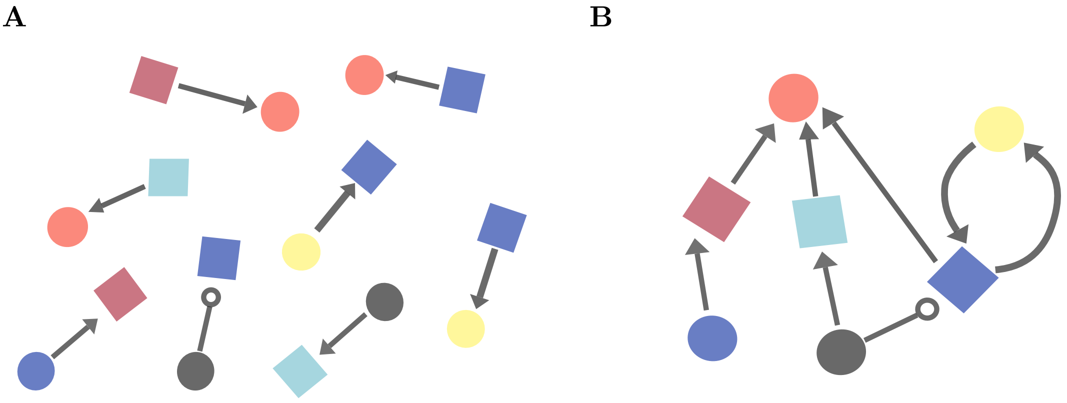


Figure 8.2: **A**: Individual building blocks of IPNs, with their place- and transition-labels color-coded, and weights suppressed for clarity. **B**: The net assembling from the building blocks by matching place and transition labels. Blue and grey places form the input interface, and orange place forms the output interface of the net.

Construction of NRSs from elementary symbols Finally, we need to add encodings for the rewriting rules in order to arrive at an encoding of an NRS and not just of an IPN. Each rule is labeled by a rule label $r \in X \subset \mathbb{N}$. As discussed above, a rewriting rule consists of two subnets L and R and a relation τ that encodes the replacement of L with

Algorithm 1 Creation of an IPN from an encoding string

```
$\mathsf{I}, \mathsf{N}, \mathsf{O} \leftarrow 0 \in \mathrm{Mat}_{|P|\times|T|}(W)$    ▷ Allocate variables for Equation 8.3.11.
$\mathsf{m} \leftarrow 0 \in \mathrm{Mat}_{|P|\times 1}$
$\mathsf{e} \leftarrow 0 \in \mathrm{Mat}_{|T|\times 1}$
                    ▷ Loop over encoding, use Equation 8.3.16 and Equation 8.3.17
for $s$ in $G$ do
    if $s.k = P$ then                                  ▷ Input matrix, normal symbol
        $(I)_{s.p,s.t} \leftarrow (I)_{s.p,s.t} + s.w$
    else
        if $s.k = T$ then                              ▷ Output matrix, normal symbol
            $(O)_{s.p,s.t} \leftarrow (O)_{s.p,s.t} + s.w$
        else                                           ▷ Input matrix, inhibitor symbol
            $(I)_{s.p,s.t} \leftarrow -1$
        end if
    end if
    $m_{s.p} \leftarrow m_{s.p} + s.m$                 ▷ initial marking vector
end for
                    ▷ Compute the indicator matrix for inhibitor arcs
for $p \in P$ do
    for $t \in T$ do
        if $I_{p,t} = -1$ then
            $\mathsf{N}_{p,t} = 0$
        else
            $\mathsf{N}_{p,t} = 1$
        end if
    end for
end for
```

R. Finally, there can be control places C and control markings $\mathbb{M}$. In order to encode rewriting rules into the encoding string of the IPN, we add the following elements to the symbols P, T, I discussed above:

- A rewriting rule label r.
- A string ρ of encoding symbols as introduced above that represent the symbols with which to replace the element they belong to.
- A control marking m_c which represents a part of the marking the encoded place needs to exhibit for the rule r to be enabled. This is added for completeness, only, though we will not make use of it in any simulations.

Therefore, we end up with symbols:

- $P_p^t(w, m, r, \rho, m_c)$
- $I_p^t(m, r, \rho, m_c)$
- $T_p^t(w, m, r, \rho, m_c)$

From a string G made up of symbols $S(p, t, w, m, r, \rho, m_c)$, we can construct the rewriting rule r by the following schema presented in Algorithm 2.

Algorithm 2 Creation of a NRS from an encoding string

- Find all symbols in the string G that contain the label r as a rewriting rule label. These constitute the string G_r.
- Loop over G_r. Build the net L from the symbols $s \in G_r$ as described in the algorithm above.
- Build the net R by concatenating all strings ρ_s into a string ϱ, and then apply aforementioned algorithm on string ϱ.
- Build the relation τ by associating each place $s.p$ encountered in G_r with the places contained in the symbols making up $s.\rho$, and the same with the transitions. In that way, we end up with $\tau = \{(s.p_1, \{s^\rho.pi...\}), ...\}$ and likewise for the transitions.
- The interface relations are determined analogously by checking which places and transitions in the original net $(\Gamma, m_0$ are connecting to places and transitions in the net R.
- The set of control places C is given by $C = \{s.p | s.m_c \neq 0\}$
- The control marking for each $c \in C$ is given by summing up the contributions of the individual symbols: $m_c = \sum_{\{s \,|\, s.p=c\}} s.m_c$.

This system, while complex, comprises the foundation of the model. It allows on the one hand to encode an NRS into a string-like system that can serve as the genome of a virtual organism, and on the other to use it for encoding NRS or IPN as initial conditions for the environment for instance, or for encoding resources that are processes by the metabolism of the organisms or the habitat network.

Secondly, the algorithm described above gives an automatic, universal construction system that can turn any encoding into a corresponding NRS or IPN. In this way, if we interpret the individual symbols indicated by P, T, I as molecules, we obtain a form of artificial chemistry if we apply it to an arbitrary collection of these elements that follows its own rules that are independent of the evolutionary dynamics that unfolds on organisms, but which influence them. The complexity of the resulting IPNs or NRSs depends on the size of the node label set, i.e., on $|P|$, $|T|$, and the number of symbols in the encoding string. Too high densities of inhibitor symbols will create non-runnable nets, however.

8.4 Principles and architecture of *Amee*

The model presented here has three levels of representation that correspond to three organizational layers of natural biological systems. Two of these are built on the formalism of Petri-nets and net rewriting systems, while the third one is a heuristic representation of individual actions in an ecosystem. As a guiding principle, we try to build a system that addresses the four challenges laid out above in subsection 8.2.1. This leads to the following basic concepts.

8.4.1 The lower layer – concepts underlying resource processing and reproduction

Given that organisms are built on top of Petri-nets with rewriting rules a natural way of reproduction is the rebuilding of the metabolism NRS from net symbols shown in Figure 8.1. Therefore, the process of ontogenesis is in its fundamental form given by the NRS assembly process, much like the process of protein synthesis is given by the tRNA-amnio-acid correspondence and the self-organization inherent to protein folding. The necessary elementary symbols to do that must be acquired from the environment. Thus, resource processing in metabolism and environment is implemented as the processing, via IPNs, of sets of elementary symbols $w_k = \{K_1, K_2, \dots K_R\}$, which we can think of as defining an IPN each. A multiset of such elementary symbol sets constitutes the marking of both the environment- and the metabolism network, $W = (w_k)$.

This elementary symbol dynamics comprise the first and lowest organizational layer of *Amee*. Where the organism-environment layer comprises ecology, we can think of this layer as a very simplistic form of artificial chemistry. It is important to note that no new resources can be produced, they can only be transformed into each other by steps of an organism- or environment IPN.

8.4.2 The middle layer – Organismal and environmental architecture

Organisms are determined by their metabolism and resource acquisition properties First, we identify metabolism and resource acquisition as the most important property of natural organisms besides their ability to self-reproduce. As a consequence, any other phenotypic traits primarily act to support resource acquisition through ecological interaction. Reproduction, while the defining property of biological systems, is generally deeply nested in an organism's makeup and thus, on a mechanistic level, is not subject to evolution anymore [Rocha, 2001]. Evolution is consequently assumed to act on resource acquisition and survival and not on the fundamental makeup of organisms. This allows us to hardcode their architecture, i.e., the number and kinds of ecological interactions, while keeping their parameters evolvable.

Organisms as complex adaptive processing systems for resources that support self-reproduction Organisms in nature are complex adaptive systems, as the previous chapters strove to show. While there is no hope of reproducing their complexity and adaptivity *in silico*, their nature as CAS has implications for evolution that arguably should not be ignored, and hence, we should strive for a system that shows such properties in principle. In order to do their complex adaptive nature justice, but keeping to the reduction of organisms to self-reproducing resource processors, we envisage them as resource processing networks that can change their own structure over the course of their life. This, in turn, we interpret as development and phenotypic plasticity.

An NRS is thus a natural choice of structure for modeling organisms because they fulfill these requirements naturally through their rewriting rules. The underlying IPN employs multi-sets of resources (yet to be defined) as marking and allows for resources to be modified during metabolism, such that organisms do not need to extract the resources that they need directly from the environment, but can transform what they acquire into a form that is more useful for reproduction. We note that this principle is not limited to resource processing, but can be applied to the processing of environmental signals too, should such a process be desired. Thus, the jump to learning systems is in principle possible in the model without too large architectural changes, although the mathematical system presented in subsection 8.3.1 would have to be changed accordingly.

Organisms not only can change their structure through rewriting in response to their own states, but also in response to environmental signals. This can modify the organism's resource processing, and thus plasticity and development can come under selection and depend on dynamically arising environmental states. The same is true for niche construction, i.e., rewriting of the habitat network, which can depend on environmental- but also on organismal states. Thus, Scheiner, Barfield, et al.'s symmetry between plasticity and niche construction is established on a mechanistic level. Moreover, this IPN-based processing system does naturally allow for waste excretion and thus for the formation of ecological resource-consumer networks. Natural selection arises when these organisms compete for finite resources, modified by rewriting or environmental heterogeneity, or other influences which are represented in the literature by separate models.

Organism's genomes are encoded using elementary symbols Because the phenotype of an organism is comprised of an NRS, it is natural to use the encoding of this NRS in a set of elementary symbols as the genome of the organism. Each symbol therein plays the role of a larger unit of genetic information that can have an independent effect on the phenotype when it changes and is thus not the analogon of a base or codon, but rather of a whole gene or GRN motif that interacts in a given way with other such symbols to produce the phenotype.

Environment as a dynamic material processing system Much like organisms, ecosystems can be seen as networks of nodes that transform material from one configuration into another and links that represent flows of matter and energy between these nodes. Nodes in this network can be biotic and abiotic, as previously mentioned. In nature, the primary energy source for the vast majority of ecosystems is solar radiation, which is made available by photosynthesizers in the form of organic chemical structures to secondary consumers, and which in turn form a trophic web that transports matter and energy through different organisms until part of it is transformed by detritivores into organic compounds that again benefit the photosynthesizers or sustain other ecosystem subnets. Abiological reaction systems exist in parallel and in interaction with them, exemplified by weathering or the transport of sediments in rivers and by wind, or many abiotic chemical processes. This establishes a conceptual symmetry between environment and organisms in that both can be simplified into resource processing networks. Organisms, or populations, are embedded in this network as dynamic elements, not only as passive objects that are "run" through an equally passive environmental filter.

Two simplifications are made, however: First, the self-adaptivity of the embedding ecosystem is ignored, i.e., it is a pure IPN, not an NRS, and does not "rewrite itself". We thereby assume that the evolution of any non-focal populations in the ecosystem has no fundamental effect on the focal population's interactions with it, or that the focal population encompasses all organisms in the ecosystem we investigate. More to the point, we focus on the evolution of the single focal population, ignoring environmental reactivity or coevolution with other populations.

Second, we ignore the first layer of resource processing, i.e., primary producers, and assume that the environment IPN we focus on starts with a fixed, but possibly initially not fully utilized, finite set of resources that it processes. The nodes of the environmental IPN thus represent different elements in the ecosystem in which resources accumulate, e.g., morphological- or ecospecies, but also inorganic things like mineral deposits, water supply, and many more.

Organisms in this system acquire resources by connecting to nodes of the environment network, extracting some resources from them, and processing them through their metabolism network, as mentioned before.

Organisms-environment interactions are not unidirectional Finally, the environment is not static but can be influenced by organismal action, i.e., not only can the resource availability change over time through consumption and waste excretion, but the network structure and size can change over time through organismal action. This is implemented by allowing net rewriting rules of organisms to apply to the environmental networks. This is a departure from the mathematical formulation presented in subsection 8.3.4, in that the network Γ_0 is exchanged, depending on the context, to either the organism metabolism IPN or the environmental resource IPN.

Thus, organisms can modify not only the local frequencies of resources, but also the processing system itself, and thus the nature of their local environment. This modification can be envisaged as, e.g., the burrowing and bulldozing of animals in sediment, the transformation of atmospheric composition or ecosystem character through photosynthesis or beaver dam buildings, and other examples of organismal modification. Consequently, the changes in the environment by organismal rewriting are experienced by other organisms, too, including the offspring of the original actor. Therefore, this environmental rewriting constitutes ecological inheritance. It must be noted that these effects do not need to be adaptive for the organism in question. Rather, they can have any conceivable effect on the ability of organisms to obtain resources and reproduce, and the emergence of any advantages is subject to evolution. This organisms-as-NRS-environment-as-IPN system forms the middle of the three conceptual layers that comprise *Amee*.

8.4.3 The upper layer – evolution from ecologial interaction

Overview of model properties *Amee* is a spatial, agent-based model. Space is represented as a collection of point-like cells connected to form a two-dimensional, rectangular grid. Each cell is connected to neighbors via the Moore-neighborhood (8-neighborhood) or the von-Neumann-neighborhood (4-neighborhood). Cells are envisaged as representing extended habitats, not individual spatial points, the dynamics of which is represented by half-open IPNs that transform matter (and energy), i.e., resources, between different configurations according to Equation 8.3.11. Individual cells get resources from the outside, i.e., from a part of the resource processing system that is not modeled explicitly. Their resource networks do not have connections amongst each other, however.

Organisms are, as mentioned, envisaged as being comprised of an NRS which contains rewriting rules for changing habitat IPNs and their own metabolism IPN. Aside from this, organisms have a number of fixed processes in which they can interact with the rest of the ecosystem. These interactions exclusively involve the environment and one focal organism, such that any interactions between organisms are mediated by their environmental impacts. Classical trophic interactions like predator-prey or parasite-host are thus not included in the current implementation. The currently implemented processes are resource acquisition and metabolization, asexual reproduction and modification of themselves and their environment during their lifetime via rewriting. Organisms can die at random or through starvation. Each organism exists on a single cell at a time, but multiple organisms can coexist on the same cell. While the number of habitat cells is fixed

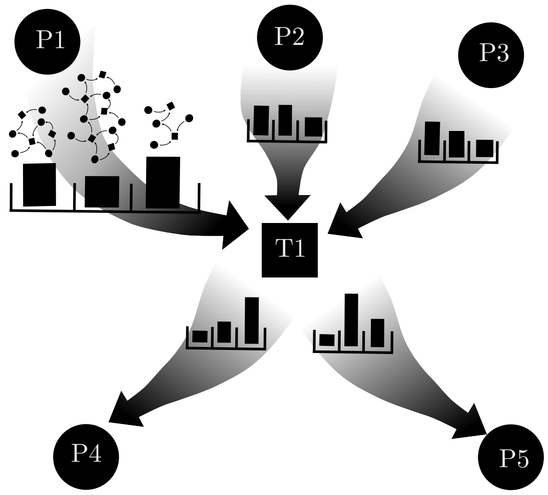


Figure 8.3: Basic resource processing across a transition in an IPN. Each resource is a set of basic building blocks, which we can think of as a net fragment, represented as a multi-set, the multiplicity of which are shown as black bars. A transition $T1$ with pre-set $\{P1, P2, P3\}$ is enabled if each place in the pre-set has at least as much marking (not shown) as the weight to the transition indicates. The transition then changes the marking on the places in its post-set $\{P4, P5\}$ by adding the weights on the respective arcs to their marking. This process can change the proportions of the multiplicities.

for the simulation duration, the number of organisms that live on each cell and in the entire system is not. Therefore, the populationsize of the system and its distribution among the habitat cells can develop freely and is affected by the evolutionary trajectory.

Energy and resource processing in a habitat As said, each cell on the grid is made up of a half-open IPN. The marking of this IPN is made up of a multi-set of sets of elementary symbols, each of which we call a "resource", as are the weights of its arcs. We can think of each resource as an IPN fragment by virtue of the assembly process in Figure 8.2, i.e., as an IPN without any further properties that would guarantee that it actually could be run.

Each habitat-IPN receives a fixed number of these resources per timestep on its input interface, and this number can be fixed for all possible resources or vary between them. However, no new resource networks beyond the initial set can be produced. Rather, their proportions can be changed (see Figure 8.3), i.e., the multiplicities of the multi-sets can change across transitions. This change in proportions also can mean that resources get produced in the post-set of a transition that have a marking of zero in the pre-set places, such that, while no completely new resources can be produced, not all have to be active on each place at the same time.

Because there are multiple resources in general, we can think of the net as consisting of multiple layers, each for one resource, which are coupled by virtue of Equation 8.3.9, i.e., either all layers run or all none do.

Energy conservation Each basic symbol shown in Figure 8.1 carries a certain amount of energy e, which can depend on the kind of symbol, such that each net fragment carries

$$E(F) = \sum_{F_P} e_P + \sum_{F_T} e_T + \sum_{F_I} e_I \tag{8.4.1}$$

where F_X is the set of P, T or I symbols in the resource respectively. Energy conservation is assured by taking into account the corresponding columns $O_{i,t}$ and $I_{i,t}$ and demanding that

$$\sum_i E(O_{i,t}) \leq \sum_i E(I_{i,t}) \tag{8.4.2}$$

for each transition T. In words, the sum of energy in the input weights of a transition must be greater or equal to the sum of the energy in the output weights of this transition. Because rewriting can change the networks, this is assured in two ways: First, any rewriting change that violates energy conservation is rejected, thus assuring that only consistent rewriting can take effect. Secondly, upon reproduction or initialization, the networks are checked for energy-conservation violation, and if one is found, the output $O_{i,t}$ is successively reduced until energy conservation is assured.

This procedure allows in principle for genomes that produce nets that violate energy conservation to stay in the population, thus creating genetic variation that would be physically forbidden in nature. However, because energy conservation is not built into the network assembly algorithm, such a post-processing approach must be taken in order to maintain conservation laws. This is a compromise with respect to computational efficiency, given that energy conservation is a difficult condition to impose in the algorithm that generates initial conditions.

Secondly, it must be noted that while energy conservation is observed, material conservation is not, again a compromise taken for simplicity. While Equation 8.4.2 could be easily extended to maintain conservation of the number of P, T and I symbols or any combination of symbol properties, the maintenance would greatly complicate the initial condition creation and its checking would be computationally expensive, and thus has been ignored here. Therefore, the number of symbols representing specific nodes or building block kinds can change arbitrarily across transitions. That being said, because symbols have been interpreted as whole genes in the genome discussion before and not as codons or bases, we can take a similar interpretation here and interpret individual symbols in the resources as molecules or larger physical entities that react with each other across transitions to produce a small, finite set of larger conglomerates.

Cell dynamics Cell dynamics is simple and consists of the influx of resources onto the input interface of each cell's IPN, followed by the running of the net according to Equation 8.3.11. The environment net thereby runs until a maximum number of iterations is reached or the net marking does not change anymore. The maximum number of runs is necessary because there is no criterium in the algorithm that generates cell initial conditions (see subsection 8.5.1) that would result in an architecture that converges each time to a final marking. This is another compromise taken for simplicity, at the price of an additional model parameter. Without it, the cell network could either run forever or for a prohibitively long time.

After the net has run, resources will be distributed in different configurations on different places on the net (see Figure 8.4), which yields environmental heterogeneity within a habitat, i.e., it allows, in principle, for different organisms to adapt to extract resources from different places and process them according to their availability. When some resources are not extracted by organisms, they will further be processed upon the next run of the cell network, which can yield temporal heterogeneity in resource availability.

Organism metabolism The first, and most important, process for organisms is resource acquisition and metabolism. This consists of four steps.

First, the labels of the places in the input interface of the metabolism net of the focal organism are considered, and the corresponding labels in the cell network are identified. Then, following the same concept as in the algorithm for building networks from building blocks, the organism metabolism net is connected to the cell's network by identifying the matching places in the input interface of the organism and in the cell network. Note that

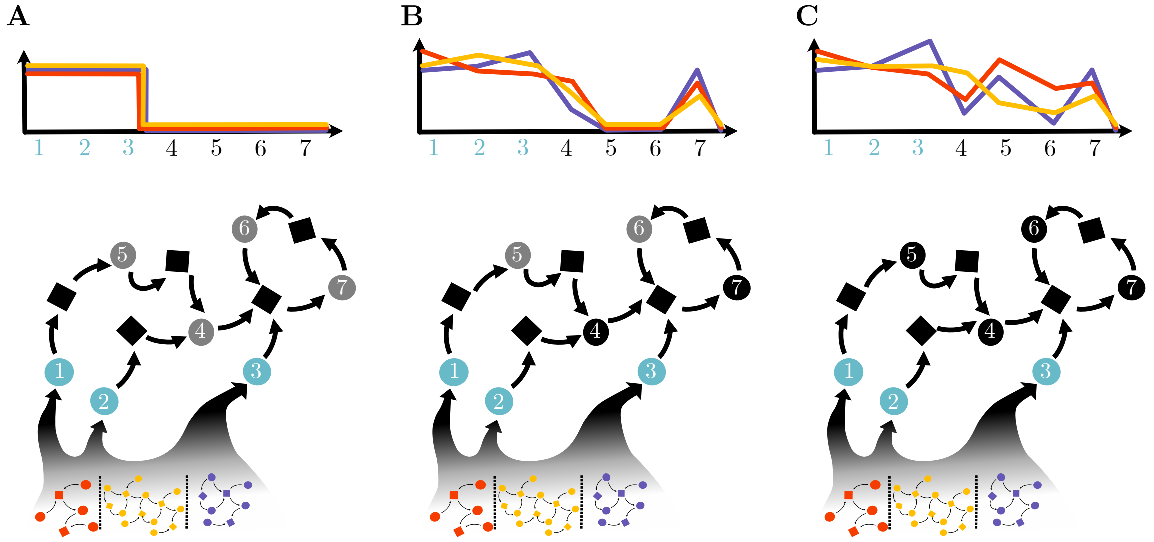


Figure 8.4: Visualization of the dynamics of the cell resource network over time. Cell input interface in light blue, labeled 1,2,3, unmarked places in grey, marked places in black. Three different resources colored red, yellow and purple are processed. The diagram at the top of each panel shows the marking of the net for each resource for different timesteps. Weights have been suppressed for clarity. **A:** Initially, the cell's IPN is unmarked. An external resource influx, here homogeneous across all resources, is applied each timestep. **B:** After the net runs, more places are marked. Some markings may stay on the input interface places because outgoing transitions become disabled at some point. Other transitions might have higher thresholds (e.g., those that lead into places 5, 6) and might not have been enabled yet **C:** Eventually, resources accumulate enough to enable some or all transitions, while others may get disabled again. The available resources on the net are thus heterogeneous across places and resources, providing a structured, dynamic niche space to which organisms can adapt.

not all of the input interface places have to be connected, any non-empty subset suffices. Failure to connect, however, will result in failure to acquire resources, precluding resource acquisition. Likewise, if organisms do not have input *and* output interface, this step will fail and the metabolism function will stop without resource uptake for the organism. This is because the output interface of the metabolism net provides the places from which resources are taken into the organism's reservoir in the end, such that their absence does preclude resource acquisition, too.

After connecting the metabolism network to the cell network, the second step consists of considering the weights that connect the input interface places of the organism to the rest of its metabolism net, i.e., the sum of the entries of the input matrix $\sum_j \mathsf{I}_{p,j}$ for $p \in \mathsf{in}$. These are compared with the marking m_p of the corresponding cell places, and the marking of the organism's input interface places M_p is changed as follows:

$$M_p = \begin{cases} M_p + \sum_j \mathsf{I}_{p,j} \text{ if } m_p >= \sum_j \mathsf{I}_{p,j} \\ M_p + m_p \text{ else} \end{cases} \tag{8.4.3}$$

$$m_p = \begin{cases} m_p - \sum_j \mathsf{I}_{p,j} \text{ if } m_p >= \sum_j \mathsf{I}_{p,j} \\ 0 \text{ else} \end{cases} \tag{8.4.4}$$

Third, the organism's metabolism IPN is run. This happens in much the same way as for the cell's resource IPN, but in contrast to cells, organisms that run their metabolism are guaranteed to have an output interface, i.e., places where no further processing takes place and resources accumulate. Like the cell network, this runs until the marking of the net does not change anymore or an upper limit of steps is hit.

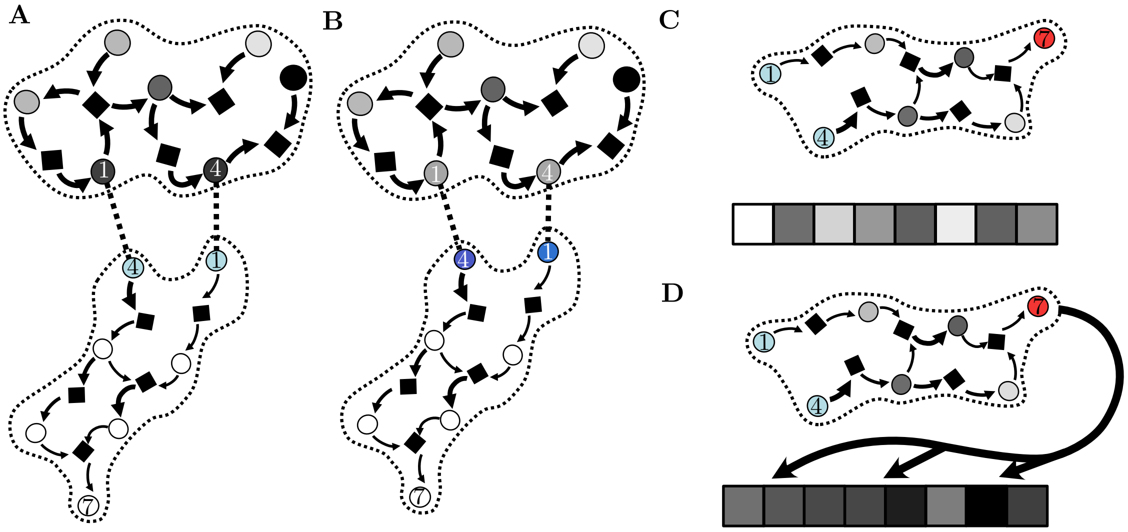


Figure 8.5: Schematic representation of the metabolism of an organism. Markings on places are shown in greyscale (cell) and blues (organism) with increasing markings from light to dark. **A**: An organism (lower) connects to the cell network (upper) via its interface places. Their place labels indicate which places connect to which. **B**: Marking is transferred onto the interface places of the organism from the cell (note the lighter coloring of the cell's places). Its output place (7) is not marked at this stage. **C**: Organsim's resource reservoir is given as grey boxes. The organism's metabolism processes the marking on the input interface until the net reached a final marking or a given upper limit of steps is reached. **D**: The marking on the output place (7) is dissolved into building blocks and stored in the organism's resource reservoir, increasing the supply available for reproduction or rewriting.

Finally, the marking of the output places of the metabolism net is considered. Each resource is considered separately and is broken down into individual building blocks which are then stored in the organism according to a predefined scheme. For example, we can set the system up to require organisms to only provide building blocks with matching place labels in the right numbers to reproduce instead of the full set of properties of the symbols (compare Figure 8.4.3 for details). For each building block gained, the organism also gains energy but also has to pay an energy cost for the acquisition, i.e., digestion, of this symbol. In this way, organisms gather material that they can use to reproduce their phenotype.

The output interface of the organism not only is the endpoint of metabolism but is also connected to the cell network through the same mechanism as the input interface. The set of connected output places forms the `waste_places` of the organisms. Finally, whatever

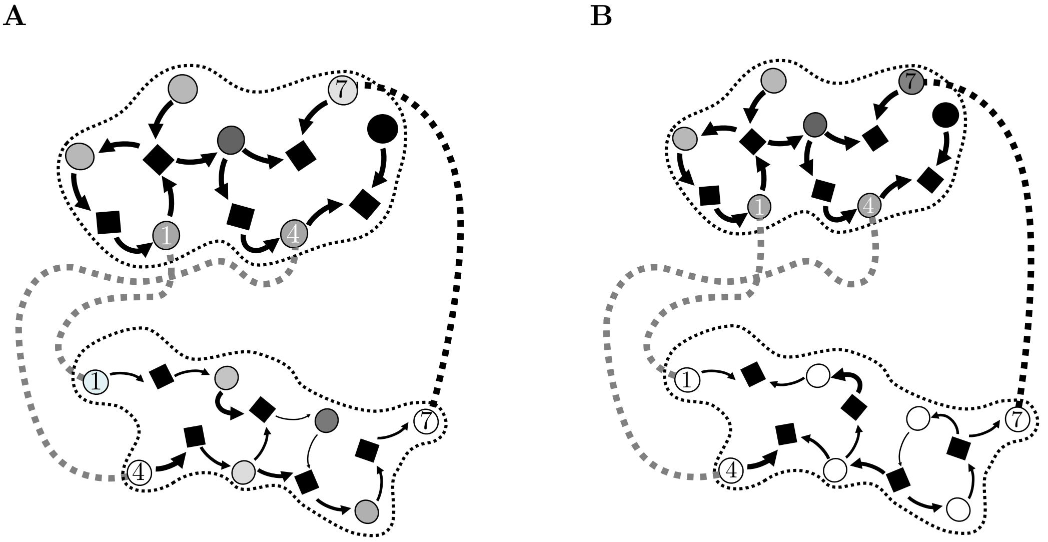


Figure 8.6: **A**: Analogously to the input interface (places 1, and 4), the output interface (place 7) of an organism (lower network) is also connected to the cell network (compare Figure 8.5). **B**: If after the run of the organism's metabolism places of its IPN still contain resources, these are uniformly distributed onto the cell's places that correspond to the organism's output places, such that their resource content increases (lighter to darker shades of grey.).

marking is left on the metabolism net is excreted uniformly onto the `waste_places`, such that unused resources are returned to the environment again, but in a possibly degraded, or at least transformed form. Therefore, organisms do not just consume resources from the environment but also change their availability through transformation and waste excretion, which may elicit adaptation in other populations.

Organisms "pay" upkeep for their own structure in the form of energy per timestep, which is subtracted at this stage, too.

Rewriting of networks The phenotype's main component is a net rewriting system. The principle of rewriting, i.e., motif replacement, has been explained in section 8.3. However, the system employed by organisms differs from this scheme in two ways. First, instead of simple replacement of motifs, it uses addition and subtraction of arc weights in addition to their creation (when they have not been present before) or removal (when one has a weight of zero), which allows for frequency-dependent rewriting effects: The effects of rewriting rules of multiple organisms can add up and create a larger effect over time if the rewriting rule leaves its target motif intact. In nature, many effects organisms have on their environment are frequency dependent, e.g., waste excretion and resource usage, but in particular also burrowing or building of structure, which can, if done by a larger population, transform an ecosystem to significant degrees.

The second difference is that the presence of the network motif to be rewritten is checked elementwise, and the organism's *and* the environment's IPNs taken into account when checking if a rewriting rule is enabled or not. Thus, the network motif that serves as the target for rewriting can be spread out over the environment as well as the organism instead of being tied only to one. In this way, the rewriting system, on which development, niche construction and plasticity are based, incorporates environmental states.

Organisms can have multiple rewriting rules, which are labeled with numbers $r \in \mathbb{N}$. The number of possible rules is a user-defined parameter R. Rules can differ from each other in size and composition, i.e., in the size of their left- and right-hand sides.

Each rule belongs to one of three possible rewriting types: "Plasticity", "Niche construction" and "Development". These are assigned intervals $R_X, X \in \{$ `Plasticity`, `Niche construction`, `Development` $\}$ such that $R = R_{\texttt{Plasticity}} \cup R_{\texttt{Niche construction}} \cup R_{\texttt{Development}} = [1, R] \subset \mathbb{N}$. These intervals need not be fully filled from the beginning, but the user can decide how many rules in each interval there should be for the initial organism, and the rest then can evolve later. Again, to make this a user-defined parameter is a choice, and the number of rewriting rules could be derived from some organism structure if desired. The ordering of the labels r in each interval R_X also gives the order in which rules are activated and checked if they can be applied.

Finally, the right-hand side of rule r can contain all or a part of the elements in the left-hand side of another rule r'. Alternatively, two rewriting rules can have the same elements in their left-hand side such that they rewrite the same subnet. In this way, rewriting rules can depend on each other, such that the application of one rewriting rule can enable or preclude the application of another one.

Additionally, a maximum number of symbols each element ρ of a genome symbol can encode is given as a user-defined parameter, too. This length of each ρ elemnet in a gene can evolve freely later and is only given for the initial condition. Therefore, the amount of genetic information in the rewriting systems is subject to evolution. With the number of allowed rewriting symbols per genome symbol and the limit on the number of rules, the initial complexity of the rewriting rules can be determined. Allowing many different rules with few allowed rewriting symbols yields simple rules that replace, e.g., one or two network elements with a handful of new ones. In contrast, allowing relatively few rules but with a large number of rewriting symbols in each gene can produce very complicated rewriting rules and a system in which the genetic information in the rewriting rules exceeds the genetic information used to build the metabolism network substantially. However, while the former case has correspondingly easy conditions to enable rewriting, the latter's are much more complex and thus harder to fulfill.

If a cell-IPN is rewritten, this comprises a way for organisms to interact, because the rewriting by one type of organism can, in principle, enable or preclude rewriting rules of other organisms. Theoretically, this can lead to a rich phenomenology in which organisms facilitate or compete for rewriting opportunities in their environment, or, in other words, engage in antagonistic activity.

Each rewriting rule that is applied incurs costs in the form of energy required per replaced symbol and in the form of elementary symbols that are needed to effect the structural change in the target network. Rewriting is thus a costly investment which, however, can have evolutionary advantages by modifying resource– and energy flows in the environment or within the organism's metabolism. When it applies to the environmental net, antagonistic, cooperative or neutral ecological interactions are possible, too. Rewriting can also have different effects on different timescales or habitats, e.g., with phenotypic plasticity being helpful on one cell but being harmful on another, or environmental rewriting helping in the short term by making more resources available, but being harmful in the long run by degrading the ecosystem's ability to process matter and energy.

It must be noted that rewriting in our model is irreversible unless a rule exists that specifically reverts the effect of another one. While this can evolve in principle, it appears exceedingly unlikely. Thus, phenotypic plasticity in this model is more akin to developmen-

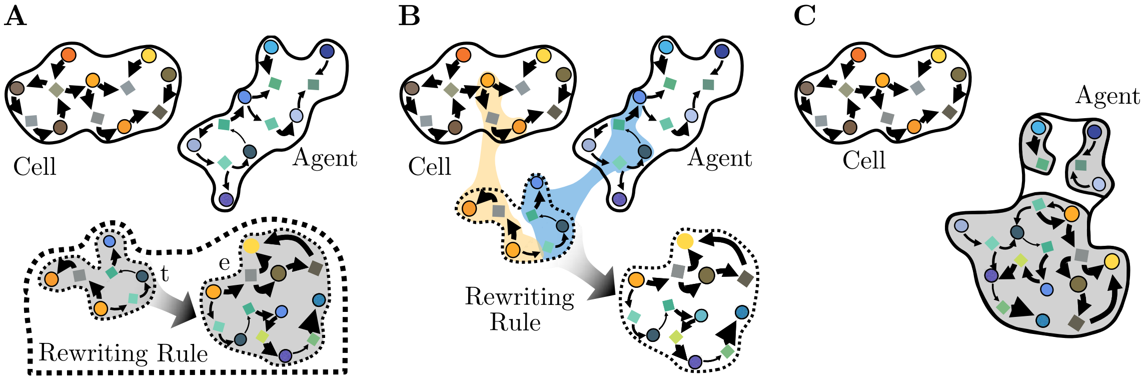


Figure 8.7: Illustration of rewriting of an organism's metabolism network. Places are shown as circles, and transitions are shown as squares. Organisms have nodes labeled blue or green, while cell nodes are shown orange, grey to brown. Here, we focus on the rewriting of an organism. Cell rewriting functioning analogously. The current IPNs or cells and organisms are bounded by solid lines, rewriting rules by dashed ones. **A**: Organism rewriting rules, like the one shown here, can mix node labels of organisms and cells. The arrow links the target network and the effect network of rewriting. **B**: Whether a network can be rewritten is determined by checking if the nodes in the target network are found in the cell- or organism IPN (shaded areas in blue and orange). Both are relevant at the same time for this search, such that environmental and organismal structures provide signals that initiate rewriting. A rewriting rule can become active if all nodes of the target net are found in either the organism network, the cell network, or both. **C**: Rewriting then replaces or adds the part of the target net that pertains to, in this case, the organism (blue area in the middle panel), with the complete effect network. For organisms, this changes their metabolism, for cells, this changes the resource processing. New elements are included, elements that are not in the effect network are removed, and those that exist in target and effect are added up together. Rewriting can fracture the original net (different grey areas), reconnect its elements, change weights, or change the interfaces, depending on the topology. How large the rewriting targets and effects can be depends on system parameters.

tal plasticity, in that it changes the developmental trajectory of the organism, instead of making minor changes that appear and disappear depending on environmental conditions. That being said, the successive application of different rewriting rules to the metabolism network can produce changes over time that resemble natural phenotypic plasticity given that there is no limit to their architecture and complexity.

Organism reproduction When an organism has enough resources to construct its own IPN *and* for the application of its developmental rewriting rules, and it has enough energy, it can reproduce. Organisms only produce one offspring at a time. This is a modeling choice made to bring reproduction in accordance with natural asexual reproduction like cell division or budding. Different criteria can be applied to measure what "enough resources" means, which yields different model complexities. The simplest version is to count only how many symbols of each type P, T, I are needed to reproduce the metabolism IPN, such that only three reservoirs for building blocks need to be maintained. The next more complex possibility is to count the needed place labels p, for which as many reservoirs as there are different places in the network and its development rules are needed Third, the full topology i.e., the k, p, t combinations can be used, which can easily $\mathcal{O}(100)$ variables that must be filled in order to reproduce, depending on the system parameters and genome length. Throughout this work, we use the place-label-based version. Reproduction happens in three basic steps:

Genome copying First, the genome of the parent is copied. There are five different mutation types that can happen, associated with three probabilities per gene. The new genome is built by walking over each symbol in the old genome, going one by one through all mutation types, drawing a random number $p \in [0, 1]$ and executing the mutation on the symbol if $p \leq p_x$, and finally putting the new symbol into the new genome. The five mutation kinds that can happen are:

- Substitution mutation: With probability p_{sub} a symbol is chosen for substitution. A single element in the symbol (e.g. place- or transition label, weight, or rewriting rule label) is chosen at random and modified, with everything else staying the same. In case the weight w or rewriting content ρ is hit, a single element is modified in the former case, while in the latter, the substitution function is applied to a random symbol in the ρ. This particular choice was deemed appropriate, given that each elementary symbol in the genome can roughly be related to a gene in a natural genome, for which base substitutions can also affect different parts with different effects.
- Duplication mutation: With probability p_{dup}, a symbol is inserted twice into the new genome. This effectively doubles the arc weight it corresponds to, such that duplication is a way to strengthen a path through a network. However, this can also have the opposite effect, e.g., because the pre-set of an affected transition must provide enough resources to run this strengthened arc.
- Deletion mutation: With probability p_{dup} a symbol is skipped and therefore deleted from the new genome. This can delete or weaken an arc.
- Crossover to a rewriting rule: A gene can cross over from the genome into a random rewriting rule with probability p_{cross} by being not copied into the new genome, but into the rewriting specification ρ of an existing gene.
- Crossover from a rewriting rule into the genome with probability p_{cross}, erasing a part of the rewriting rule in the process but adding a new gene to the metabolism IPN. This generates an entry in the genome with no rewriting because symbols within the rewriting specification ρ of a symbol have none of their own. However, later crossover mutations can change this again.

Because the mutations are applied per gene, a single gene symbol can change through multiple kinds of mutations at once. Mutations are, however, independent of each other, such that it is comparatively unlikely that more than one mutation happens to any one symbol that is copied.

Organism construction Organisms are constructed from their genome as discussed in subsection 8.3.5.

As a first step in this process, the genome is split into two sub-genomes at a user-prescribed length l_m, such that two symbol sets emerge. The first is used to build the movement propensity (see Figure 8.4.3), and the second is used to build the metabolism and the rewriting system of the organism. The same principle can be applied to build all other parameters of an organism that currently have to be defined by the user, should this be desired.

The node-label set used for the construction of the organisms and the habitat can be different, i.e., while they must overlap due to the way the metabolism function is implemented, they need not be equal. The exact sets used are model parameters and are determined by the user. The genome consequently can contain symbols that pertain purely to the environment, i.e., create environmental rewriting rules, but have no impact on the metabolism IPN of the organism. This is achieved through the introduction of overlapping, but not identical sets of node labels for the environment and organism networks. For example, a symbol $(P, p_x, t_y, w, r, \rho, m_c)$ with x, y not part of the set of node labels that refers to organisms will, if it is part of the organism's genome, only create or contribute to a rewriting rule r that pertains to the habitat's resource network but will not affect the organisms' metabolism in any way.

In that way, organisms can rewrite networks that have a differing structure than their own metabolism IPNs, because symbols that represent nodes outside of the organismal node label range take part in the building of rewriting rules but not in the creation of the metabolism IPN. As a consequence, when the metabolism of organisms is created, the genome must be filtered such that only those genes take part that have node labels in the organismal node label set. The filtering of the genome according to the prescribed node label range is called "splicing" here and is the first step of organism creation.

Next, the metabolism IPN is built from the spliced genome, while the rewriting rules are created from the complete genome. Then, auxiliary parameters like initial energy and recording variables for age and number of offspring are initialized.

Finally, the developmental rewriting rules are applied (see Figure 8.4.3). As their application can depend on the presence or absence of environmental structures, the final phenotype of an organism is, in general, dependent on the environment.

Phenotypic variation The creation of a new organism from a mutated genome can yield a number of phenotypic differences. First, the metabolism's input interface places can change, thus making the organism derive resources from other environmental elements. This also includes the possibility of tweaking which resources are drawn from the environment in what quantities by mutating the weights that connect the input interface to the rest of the metabolism net. Additionally, the structure of the metabolism network may change in total over time to produce the specific resources organisms need to reproduce more quickly, with fewer input resources or from different ones. This evolution is of particular interest if the node-label set used in the simulation is complex and there are many resources of which each would be insufficient for successful reproduction.

Another kind of phenotypic variation is the evolution of rewriting rules, which can impact vital metabolic or environmental restructuring. In particular, the evolution of new rewriting rules or the change of existing ones can change the efficiency of the metabolism network or its input or output interfaces, thus allowing for niche shifts based on developmental and phenotypic plasticity. Evolution of rewriting, because it uses the same kind of resources as reproduction, can represent a tradeoff against lifetime reproductive success, such that we expect different strategies to evolve over time.

Finally, movement propensity can change, too, via mutations of the symbols in the genome interval that's used to build it. Therefore, different kinds of organisms can pursue different strategies of mobility between habitats.

Developmental rewriting The last step in phenotype creation is the application of the developmental rewriting rules. In contrast to the rules associated with plasticity and niche construction, the material cost for the developmental rules is supplied by the parent organism. Developmental rewriting is thus “free” for the offspring, but costly for the parent. Development, as with all rewriting, does integrate environmental elements into the process of checking rule applicability. In this way, development includes environmental signals such that environmentally conditioned developmental plasticity can come about. This rewriting can be vital for the functioning of the offspring because it can transform an otherwise non-functional metabolism net into one that is functional by relinking parts of the net that allow for it to run, or by providing the necessary interfaces. Therefore, harmful genetic variation in the metabolism architecture part of a genome can be offset by genetic variation in rewriting.

It must be noted that developmental rewriting is distinguished from other rewriting rule applications just by its timing during organism creation, not by its impact, which is normally used to distinguish phenotypic from developmental plasticity.

Cost subtraction Finally, after the newborn organism is created, the material costs are subtracted from the parent’s material pool, as is the energy cost. The latter is divided into two parts, first a plain energy cost for reproduction, and second, a batch of energy that is imparted to the offspring to serve as its “starting” package. Both values are system parameters.

Movement Movement is probabilistic, with the movement probability for an organism being given by

$$p_{move} = \exp\left[-\mu \frac{e}{m_{move}}\right] \tag{8.4.5}$$

where m_{move} is an evolvable quantity that is given by the sum over the first n place labels in the genome of the organism according to

$$m_{move} = \sum_{i=1}^{n} \frac{p_i}{\max(p)} \tag{8.4.6}$$

where $\max(p)$ is the maximum possible placelabel. The n symbols this is derived from are the part that is split from the metabolism genome as explained in Figure 8.4.3. The parameter μ, in Equation 8.4.5 is a user-defined system parameter that sets a baseline mobility, and e is the organism’s energy storage. Organisms with high energy, and thus a good nutritional status, are less likely to move away from their current habitat, but a high movement propensity m_{move} can offset this effect.

This ad hoc formula allows for a population to evolve mobility over time. After p_{move} is computed, a random number between 0 and 1 is chosen, and if it’s smaller than p_{move}, the organism chooses a random cell in the Moore-neighborhood of its current cell and relocates there. Thus, spatial location, but also mobility, become dimensions of the niche space of the system.

Movement incurs an energy cost e_m for the organism, which is a system parameter chosen by the user.

Phenotypic plasticity Although the term “phenotypic plasticity” is often loosely used to include developmental plasticity, here, it only refers to changes during lifetime and only to those that occur after the construction process of the focal organism has been completed. Phenotypic plasticity rules are applied in much the same way as development,

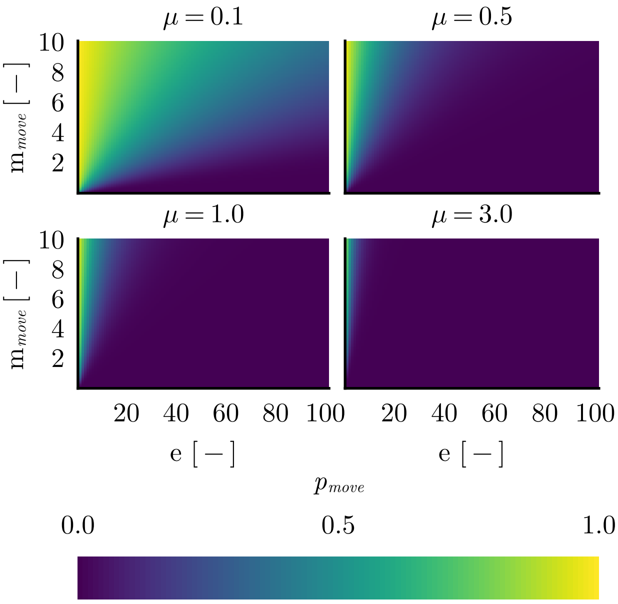


Figure 8.8: Movement probabilities according to Equation 8.4.5 for different organism movement propensities m_{move} and energies e and different paramters μ.

i.e., the system looks for the target element's presence in the environment- or metabolism IPN, but only the metabolism IPN is affected by the rewriting. Phenotypic plasticity can happen at different points during the lifetime of the organism, and consequently affect its further ecological behavior. Plasticity rewriting can only occur if the organism has enough energy and material building blocks to execute at least one of its plasticity rules. As mentioned, the distinction between phenotypic- and developmental plasticity in *Amee* is artificial in the sense that they are only distinguished by their timing, and not by their effect.

Niche construction and decay Niche construction functions in the same way as plasticity, but the environmental IPN is affected instead of the metabolism net of the organism. In contrast to plasticity and development, these changes potentially affect all organisms in the population and can be frequency dependent, i.e., the more organisms modify the same environmental motif, the stronger the effect can become, because rewriting can be additive instead of purely replacement based, as mentioned before. Niche construction rewriting can only occur if the organism has enough energy and material building blocks to execute at least one of its niche construction rules.

Repeated rewriting can be precluded or enabled by the rewriting activity of other organisms, such that this implementation of niche construction allows for reciprocal scaffolding of the environmental effects of different rewriting rules carried by different organisms. Thus, the niche construction of one organism can benefit or harm other organisms. This effect also extends into plasticity and development, because a rewritten motif can provide environmental signals that allow for previously disabled development- or plasticity rules to become enabled. Therefore, not only does it enable a form of environmental interaction beyond competition for resources (see Krakauer et al. [2009] or Chisholm et al. [2018] for an extensive discussion of such effects), it also includes the possibility of the long-term biogenic transformation of a habitat.

In passing, we note that, in nature, organismal effects typically decay over time, even though this may take much longer than the effector's lifetime. In the model presented here, this effect has been ignored for simplicity, in the same way as phenotypic plasticity is irreversible, too. It is thus assumed that whatever structural influence organisms exert on their environment has a lifetime that is longer than the simulation time. In nature, such an assumption may correspond to coral reefs, the creation of forest clearings, or similar long-term effects, but the creation of burrows or beaver dams may be a less suitable example, as they typically become non-functional after one or a few generations of not being maintained. While the long-term maintenance of modifications in *Amee* provides perfect ecological inheritance over arbitrary times in principle, it must be noted that network motifs, after being rewritten, can be modified again, because the rewriting of one organism can enable rewriting by others or elicit evolutionary responses in other's environmental rewriting rules. Thus, while no effect decays again *if it is left alone*, neither do they need to stay the same over time. With this effect, ecological inheritance becomes imperfect again. A decay algorithm for environmental influences can be implemented to relax this assumption, but its presentation and investigation have been forgone here to reduce complexity.

Death Organisms die probabilistically, dependent on their energy supply:

$$p_{death} = \begin{cases} \min\left(\exp\left[-d\left(e/l\right)\right] + d_0, 1.0\right) & \text{if } e > l \\ 1 \text{ else} \end{cases} \tag{8.4.7}$$

where d represents a scaling parameter that is fixed for the simulation, e is the organism's energy, l is the living-cost parameter (the upkeep mentioned in Figure 8.4.3) each organism has to pay each timestep, and d_0 is another user-defined parameter that makes sure organisms retain a finite death probability per timestep. Each timestep when an organism is updated, Equation 8.4.7 is evaluated for its current energy supply, and a random number in $r \in [0, 1]$ is chosen. if $r < p_{death}$, then the organism will be removed from the population.

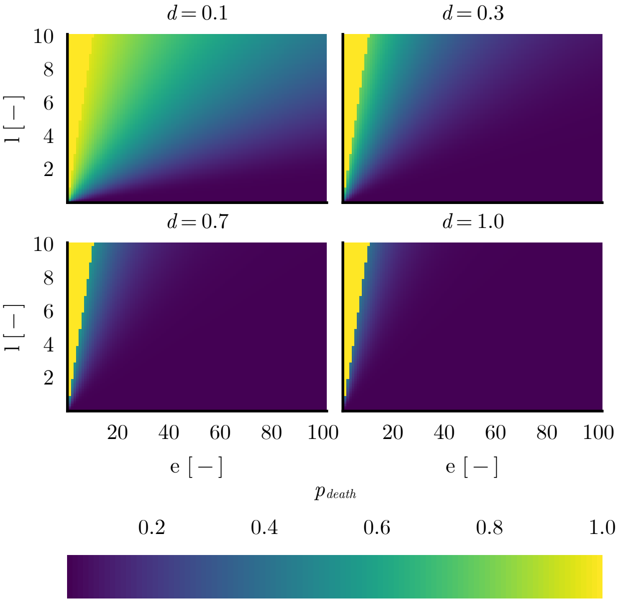


Figure 8.9: Death probability according to Equation 8.4.7 for different scaling parameters d with a fixed offset $d_0 = 0.05$. With rising living costs and a lower energy supply, the death probability for an organism rises, emulating starvation or the effects of malnutrition. The intensity of this effect can be modulated by the scaling factor d, which is a model parameter. If energy e of an organism is less than the living cost, it dies always, which creates the area with $p_{death} = 1$.

8.5 Simulation setup and algorithm

8.5.1 Initial conditions

Initial conditions are found by a genetic algorightm The initial IPNs or NRSs for cells and agents, as well as the resources, are created using a genetic algorithm [Mitchell, 1998], that optimizes an initial population of randomly generated NRS or IPN encodings (genomes) to fulfill a number of conditions that are geared towards allowing initial organisms to survive and thrive under given simulation parameters. This relies on the concepts presented in subsection 8.3.5 and the idea that in the same way that organisms can be constructed from their genomes, the IPN of the habitat can, too. The degree to which the respective entities fulfill the required conditions is measured by a hand-crafted fitness function for each type of entity, i.e., habitat and organism, which maps the performance of an entity with respect to these conditions onto a scalar f which is called “fitness”.

For each entity, organism and habitat, the algorithm employs n populations with N net encodings each, which are subject to random exchange between populations and selection on survival based on the fitness function chosen for each case. N thereby is constant and in each step, a fraction of the total population is stochastically removed proportional to their fitness f. Organisms reproduce sexually and are selected for reproduction with a probability that is proportional to f. During reproduction, a new encoding is created by recombining the two parent encodings symmetrically, i.e., the probability for a symbol at place i of the new encoding to originate from one or the other parent is 0.5 each.

Exchange between populations is implemented by randomly exchanging p percent of two randomly chosen populations at each step. Through manipulation of this parameter, the algorithm can be made to evolve multiple more or less isolated lineages with different gene pools and possibly different parts of the solution.

Additionally, mutations can happen by inserting a new, randomly created symbol, deleting a random symbol, or replacing a random symbol, each with probability m. The number of organisms removed per step, s, can also be varied to allow for, e.g., faster but less explorative optimization. In essence, this parameter governs the strength of the selection the populations are subject to. The system runs until it has found a viable solution according to the given criteria or until it hits a maximum runtime, in which case the simulations break off without running further because no viable initial condition could be found.

The fitness f is a scalar that is computed by a weighted sum of measures for how well an entity fulfills a set of requirements. These depend on whether an initial condition shall be found for a cell, an organism or a resource. In general, f is given by

$$f = \left(\sum w_i * \phi_i\right)^{1.5} \tag{8.5.1}$$

where w_i is the weight assigned to a measure ϕ_i. Taking the sum to the power of 1.5 allows for a greater range of fitness values which in turn allows for more fine-grained optimization by more rapidly filtering out low-fitness organisms without having to rely on a large s that in one step would greatly reduce genetic diversity. Furthermore, in combination with w_i it allows for encodings that fulfill multiple different requirements, but not all of them, to survive and thrive in the population, thereby allowing them to recombine into a full solution.

In the case of the agent, this fitness function is a combination of the measures for up to nine conditions that must be fulfilled to be a viable solution, and which are weighted individually. These conditions include the existence of input- and output interfaces, a net increase in reproductively useful symbols over k metabolism runs with k being an algorithm parameter, and successful development, plasticity and niche construction if these are activated in the simulation. Furthermore, a minimum number of steps of the metabolism network is required until it converges, to preclude the system from producing always the most simple initial condition.

For cell networks, the conditions include the existence of input interfaces and the requirement that when the net runs, a fraction n_p of its places are left with some marking on them, to provide a sufficiently diverse background ecosystem to which the focal population can adapt. In total, a candidate cell IPN must fulfill three different conditions.

For resources, the sole condition is that the set of resources that is fed into each cell via the resource influx is such that it provides all potentially needed resources overall, i.e., all place labels are present in the processed multi-sets.

The parameters governing the genetic algorithm are listed in Table 8.1.

Parameter	**Meaning**
exchange_proportion	Percentage of exchanged elements
m	Mutation rate
m_without_selection	Mutation rate when selection is off
n	Number of populations
optimization_rounds	Maximum runtime of the algorithm
N	populationsize
s	Percentage of elements removed each round
rng_seed	Initial seed for the random number generator
stop_selection_every	Number of steps for which selection runs until it is interrupted to allow the population to generate genetic variation again.
stop_selection_for	Interval for which selection is interrupted.
length_altering_mutations	Whether insertion and deletion mutations are enabled

Table 8.1: Parameters for specifying the genetic algorithm run.

It must be noted that for the organism, which is a far more difficult optimization problem to solve, the mentioned nine conditions are partially redundant: For example, the presence of input- and output-interface of the metabolism network are a prerequisite for a successful metabolism run, but both, successful metabolism run and interface presence, are included as relevant quantities into the fitness function Equation 8.5.1. Including such redundancies creates a more fine-grained fitness landscape in which the organism encoding can evolve, which allows for the successive generation of traits that can recombine to form a solution, instead of requiring a solution to evolve in a few big steps.

It must be noted that it is not guaranteed that the algorithm finds a solution for any combination of parameters, and finding a good compromise of efficiency and effectiveness can require tests by trial and error. In order to avoid premature degeneration of the population, selection can be disabled after a user-defined number of steps to allow for an increase of variation after it has been reduced by a period of natural selection.

Setup of resources and net building blocks The parameters that define resources are given in Table 8.2. Each symbol in a resource, IPN or NRS-encoding is initially generated randomly, i.e., an element of the encoding of a net building block is drawn at random from a uniform distribution, except for the symbol kind, where each kind has its own probability in order to ensure that nets build from a pool of symbols are not, e.g., filled with inhibitor arcs that would preclude them from running.

Each symbol type has an associated energy value. Resources can have numbers of symbols in $[S_{lower}, S_{upper}]$, such that resources can be heterogeneous in composition and length. However, the `homogeneous_resources` parameter can make them all equal in length and composition, in which case a random length in $[S_{lower}, S_{upper}]$ is chosen, a single resource is generated and all resources are filled with this sequence. The node labels for the places and transitions of each symbol are drawn from the union of the labels given for cells and agents, respectively, such that all possible compositions of networks and rewriting rules can be created.

Parameter	**Meaning**
E_P	Energy of a P symbol
E_T	Energy of a T symbol
E_I	Energy of a I symbol
homogeneous_resources	Whether all resources should be the same or not
S_{lower}	Min. length of resource sequence
S_{upper}	Max. length of resource sequence
p_P	probability to draw a P symbol when creating a symbol sequence
p_T	probability to draw a T symbol when creating a symbol sequence
p_I	probability to draw a I symbol when creating a symbol sequence

Table 8.2: Parameters for specifying the resources and building blocks for the NRSs and IPNs in the simulation.

Setup of cells and organisms The parameters used for organism and cell creation are listed in Table 8.3 and Table 8.4.

After the genetic algorithm produces a net encoding for a cell that fulfills the given requirements, each cell of the grid is initialized with this encoding, such that the grid is homogeneous. This applies also to resource influx and resource capacity, and all other cell parameters. Environmental heterogeneity can be implemented if desired.

Then, the genetic algorithm is employed to find a single organism genome that fulfills the requirements for organisms, from which a single organism is created that is put onto a random cell in the grid. The simulation is therefore set up to include an initial colonization phase.

8.5.2 Simulation algorithm

On the choice of update algorithms in models of complexity and evolution The issues arising from the choice of time discretization schemes in numerical ODE solvers are well

Parameter	**Meaning**
M_{max}	Maximum marking that can be put onto input-interface places
encoding_size	Initial number of elements for the linear representation of the resource network.
γ_M	Number of resource instances put onto the input interface places per resource and timestep
fill_start	index of the first resource to be utilized.
fill_end	index of the last resource to be utilized.
I_N	Node label interval for cell IPN
I_W	Allowed weight interval per symbol
i_{max}	Max. number of steps for a single IPN run
seed	Random number seed for initial genetic algorithm
neighborhood	Cell neighborhood. “Moore” or “vonNeumann”.
extent	Grid size in x- and y- direction
periodic	Whether the grid is periodic or not

Table 8.3: Parameters for specifying the initial habitat.

known. In computer models of complex and complex adaptive systems, the focus is not on accuracy, because no known, analytical function is discretized.

Rather, the question arises whether the particular algorithmic choices made during implementation will systematically influence the results. For cellular automata, *synchronous* and *asynchronous* update schemes have been introduced, which differ in the way in which the state of the automaton's cells are updated: In a synchronous update scheme, the new state of all cells is calculated first and then the states of all cells are updated simultaneously, such that no interaction between an updated state on cell A and an old state on cell B can occur, while in an asynchronous scheme, each cell is updated immediately and such interactions can potentially occur. In general, the choice of an update method determines how information on the state of a particular element of the simulation propagates through the system. Complex adaptive systems, and evolutionary systems in particular, are generally sensitive to details of this flow, as are all nonlinear systems. In addition to this intrinsic sensitivity, evolutionary systems are by nature stochastic. Because of the independence of mutation from their effect, they have to be modeled by some stochastic process, with the mutation rates involved being model parameters. Therefore, it seems unavoidable that a quantitative dependency on the choice of update scheme, and implementation details in general, will be present in a model for an evolutionary system.

In nature, all the different functions that are executed in evolving systems, while often ordered into some cyclic process, are carried out by all members of the population in parallel, with synchronization between them often being a major evolutionary outcome, e.g., mating or migration. In computational terms, this constitutes a form of massively parallel processing without synchronization. Indeed, the ensuing conflicts for resource access that occur in such a system are a main source of natural selection. Aside from this, these functions are typically bound to conditions, e.g., reproduction is only attempted when enough resources are available or the general conditions are favorable, and a habitat is only left if it is depleted or otherwise shifts into an unfavorable state.

Parameter	**Meaning**
G	Number of basic symbols in the initial genome
i_{max}	Maximum number of steps the metabolism can run before aborting
μ	movement scaling parameter
l	living cost
n_G	percentage of the genome used to build m_{move}
d	death probability scaling
d_0	death probability offset
fill_start	index of the first resource to be utilized.
fill_end	index of the last resource to be utilized.
e_{digest}	digestion cost per acquired symbol
e_m	movement cost
$e_{offspring}$	energy imparted to offspring
e_r	energy cost for reproduction
p_{sub}	Substitution mutation probability
p_{dup}	Duplication mutation probability
p_{cross}	Crossover mutation probability
I_W	Weight intervals for symbols for metabolism IPN
I_N	node label interval for places and transitions
I_{dev}	Interval for developmental rules
$N_{0,dev}$	Initial number of developmental rules
e_{dev}	Energy costs per symbol for development
I_{constr}	Interval for nicht construction rules
$N_{0,constr}$	Initial number of nicht construction rules
e_{constr}	Energy costs per symbol for nicht construction
I_{plast}	Interval for phenot. plasticity rules
$N_{0,plast}$	Initial number of phenot. plasticity rules
e_{plast}	Energy costs per symbol for phenot. plasticity
R	Range of possible rule labels
L_R	Size range for rewriting interval each symbol can have
r	Interval of resources that can initially be used. May not encompass all possible resources.
rng_seed	Seed used for drawing random numbers in the init. genetic algorithm.

Table 8.4: Parameters for specifying organisms.

Therefore, all functions an organism can carry out are split into a "condition" that must be fulfilled for the function to become active and a "propagator" that changes the state of the organism. For example, the movement condition computes the probability p_{move} according to Equation 8.4.5 and checks if the movement should happen or not, and if it does, the movement propagator will then put the organism onto a random neighboring cell. All the possible functions are organized into pairs of a propagator together with its corresponding condition, which results in the following list of pairs:

- 1: Movement
- 2: Plasticity rewriting
- 3: Niche construction
- 4: Metabolism
- 5: Reproduction
- 6: Death

For a computer model like the one discussed here, the unsynchronized parallelism of nature is difficult to implement directly. Therefore, three different versions of the ordering of organism functions have been implemented, which can be chosen by the user.

- Fully synchronized: Each organism executes the same function at the same time, e.g., the entire population executes movement, then metabolism, and so on. The execution order is given by the list of condition-propagator pairs listed above.
- Fully randomized: Each organism draws a random number in $[1, 6]$ each timestep and executes the respective function from the list above.
- Evolvable: Each organism carries an evolvable list of integers that corresponds to an index in the list of propagators above. Death is excluded and is checked in each step to avoid a situation where organisms get rid of the death function by evolution, thus becoming immortal. This lifecycle is an independent part of the phenotype and is not derived from the genome. Rather, it is initialized to the list given above and may evolve via substitution of indices, but not by deletion or duplication. This choice has been made to avoid a situation where organisms would evolve ever more reproduction elements in their lifecycle without having to give up any other function for it. It is, however, easy to relax and constitutes a possible future research direction. Additionally, the lifecycle could, in principle, be derived from the genome in the same way the movement parameter m_{move} is, which, however, is a minor correction. Organisms cyclically go through their lifecycle vector and execute one function each time they are chosen to update.

For cells, the situation is much simpler because they only have a single propagator (resource influx and processing), such that the aforementioned possibilities of ordering the update functions all reduce to a single one.

Time-stepping algorithm of the model To arrive at a functional system, the method of choosing an organism or cell for updating must be explained. In cellular automata, the entire population of cells is traversed in each timestep, in deterministic automata even in the same order. Here, we follow a different approach, with two goals. First, we again wish to avoid systematic influences on the created trajectories by avoiding hardcoded orderings wherever possible. For instance, traversing a population of organisms on a given cell always in the same order would result in the first organism accessing always a full resource reservoir, while the last one will always encounter a depleted one, given it accesses the same resources as the others. Therefore, the order in which organisms are updated must be randomized.

Second, we wish for timesteps to have a unique, definite meaning that only depends on system parameters, but not on dynamic system variables, like populationsize.

If we randomly choose a single organism to be updated each iteration of the model with probability $1/N$ with N being the populationsize, then the frequency of being updated for any single organism would be stretched for large populations and compressed for small ones, which essentially means that the system's effective timestep, defined for instance as the average time until every organism has activated all its possible actions at least once, would depend on the populationsize N.

In order to avoid that, we follow Schönfisch and Roos [1999] in decoupling the flow of time from system variables and binding it to a parameter τ_O. The same holds for environmental cells with a parameter τ_E. Each of these two variables defines an exponential distribution

$$p_i(x) = \frac{1}{\tau_i} \exp\left(\frac{x}{\tau_i}\right) \tag{8.5.2}$$

which is employed to stochastically generate a timestep at which each organism and each cell, respectively, is scheduled to be updated next. Once this timestep has been reached and the organism or cell has been updated, a new number is drawn via Equation 8.5.2, and is added to the current timestep, which gives the next timestep the focal entity is scheduled to be updated.

Given a uniform random choice of activity each time an organism is updated in that way, and a fixed number of activities n_O, the average time until an organism has executed all its activities is given by $n_O \tau_O$, and respectively for the environment cells. Here, the environment cells have only one activity which is their resource processing, so their average full update time is trivially given by τ_E.

Implementation The model is implemented in C++ using the Utopia framework (see chapter 7), which provides all the necessary tools for running agent-based models. For data analysis and visualization, the Julia programming language has been employed [Bezanson et al., 2017].

Parameter	**Meaning**
seed	The random number generator's initial seed
num_steps	Number of steps the simulation runs for
stepper	Whether the lifecycle of organisms is fixed, evolvable or randomized.
additive_rewriting	Whether rewriting is additive or replacement based only.
agent_timescale	Characteristic timescale of organism update
cell_timescale	Characteristic timescale of cell update
weight_comparison	Wether weights are included when rewriting rules are checked for applicability
resource_number	Number of possible resources
with_construct	Whether niche construction is enabled
with_development	Whether development is enabled
with_plasticity	Whether plasticity is enabled

Table 8.5: Parameters for specifying the simulation dynamics.

9. Results I: Evolution without rewriting

Because *Amee* is a complex model with a plethora of possible phenomena, a comprehensive overview of all possible parameter combinations and scenarios would go beyond what is possible to present in one work. Rather, we focus on a number of basic scenarios in order to showcase the model's capabilities and highlight specific directions for future research. The approach is thereby more explorative and descriptive, giving an overview of *Amee*'s possible phenomenology, rather than diving deep into individual analysis.

9.1 Baseline 1 – Single-resource single-cell systems without rewriting

Basic setup As a baseline, we choose the most simple configuration of the system, i.e., a non-spatial simulation with a single resource and no rewriting, e.g., phenotypic adaptation is only possible via genetic change. Habitat and organism metabolism can have more than one node, however. The niche space is thus restricted, and the population is forced into competition for resources. The only significant niche space dimension is thus the different places of the environmental net that organisms can access, such that evolution can be expected to be focused on the metabolism network and its input interface. The resource restriction will later be lifted, and multiple resources will be introduced. The basic parameters for the creation of the resource composition, the habitat, the simulation algorithm, and the organisms are given in Table 9.1 and Table 9.2, respectively. Parameters shown in chapter 8 but not given here are varied over different scenarios and given as described, or are of no consequence for the current scenario (e.g., neighborhood for a grid with a single cell). The parameters have been chosen to produce a simple scenario in order to allow for tractability of results.

On the role of rewriting rules when rewriting is off Rewriting rules are encoded in symbol sequences that each genome symbol can carry (see subsection 8.3.5). Because the crossover mutation mechanism exists, the genetic information that normally constitutes rewriting rules still plays a role even when no rewriting is allowed, because it provides a form of genetic variation that can become active via crossover mutations. Therefore, each organism carries an active and inactive gene pool, with the latter playing a larger role in higher crossover mutation rates. Indeed, if the rule size parameter L_R is larger than one, this inactive gene pool can become substantially bigger than the active one, potentially contributing significantly to the adaptability of the population. We can liken this to "cryptic" genetic variation that has been discussed in chapter 5, although there, cryptic genetic variation represents genetic variants that are normally neutral but can come under selection via plasticity or changed environmental conditions, whereas, here, the genetic information in the rewriting rules can become active via mutations.

Parameter	**Value**
E_I	0.02
E_P	0.02
E_T	0.02
p_I	0.05
p_P	0.475
p_T	0.475
S_{lower}	70
S_{upper}	71
h_S	true

Parameter	**Value**
M_{max}	25600
γ_M	5120
$I_{N,Habitat}$	$[0, 10]$
$I_{W,Habitat}$	$[0, 256]$
i_{max}	25
extent	$[1, 1]$
periodic	false

Table 9.1: **Left**: Resource parameters for the baseline simulation. **Right**: Habitat parameters for the baseline simulation.

9.1.1 Initial conditions and the influence of randomness

Basic expectations and hypotheses Because evolution, or complex constructive systems in general, are nonlinear, their trajectories exhibit sensitive dependency on initial conditions and small disturbances. Given the conditional character of the state propagation functions in *Amee* and the underlying complexity of the IPN/NRS systems, we can expect this to be the case here, too.

Moreover, evolution is uniquely sensitive to singular events through its dependency on mutations that appear in a single organism. The carriers of these, by chance and through selection, can rise in frequency to make up a significant part of the population and therefore change the trajectory of the entire system (compare chapter 3).

The initial organism with which a simulation is started is also not unique for a given set of parameters but is dependent on the trajectory of the genetic algorithm employed to generate it. In other words, the optimization problem that is posed to the genetic algorithm has no unique solution. Therefore, we expect that the trajectory of any given simulation will crucially depend on the initial condition with which it is started, and therefore, on the random number sequence that the simulation is generated with. Consequently, an ensemble of simulations is necessary in general. Comparing this with nature, we have to take into account, especially with regard to extinctions, that all the natural species, colonization of new habitats and generally evolutionary systems we find are just those that survived until observation. Others might have existed at some point, but are invisible to us. Evolution thus is susceptible to founder effects, e.g., in a colonization event, the traits or genetic variants which the original colonizers carry. Thus, we expect this behavior from *Amee*, too.

Moreover, organisms are evaluated by the initial genetic algorithm in isolation, while the actual simulation produces a population of interacting agents. Consequently, the successful creation of an initial condition does not guarantee that a population derived from this organism will survive in the long term. Indeed, a population could deplete a local resource to a degree that precludes a successful run of metabolism networks at some point, thus increasing mortality among a nascent population, possibly to the point of extinction. Finally, simple random death may, in early stages where populationsizes are small, also produce extinction or preclude the fixation of advantageous mutations because of the offset in d_0 in Equation 8.4.7.

Parameter	Value
i_{max}	20
l	1.0
d	0.2
d_0	0.05
e_{digest}	0.002
$e_{offspring}$	5.0
e_r	1.0
$I_{W,organism}$	$[0, 10]$
$I_{N,organism}$	$[0, 10]$
R	$[0, 30]$
I_{dev}	$[0, 10]$
$N_{0,dev}$	5
I_{constr}	$[10, 20]$
$N_{0,constr}$	5
I_{plast}	$[20, 30]$
$N_{0,plast}$	5
R	$[0, 30]$
L_R	$[0, 3]$
r	$[0, 1]$
p_{sub}	10^{-4}
p_{dup}	10^{-4}
p_{cross}	10^{-4}

Parameter	Value
num_steps	25000
stepper	randomized
agent_timescale	1
cell_timescale	1
resource_number	1
with_construct	false
with_development	false
with_plasticity	false

Table 9.2: **Left**: Organism parameters for the baseline simulation. **Right**: Simulation parameters for the baseline simulation.

In order to get an overview of the potential impact of variability in the random number sequence specifically, we investigate two scenarios, one where the random seed[1] of the simulation is fixed for all simulation runs while a different random seed is used each time for the generation of the initial condition (scenario B), and the opposite layout where the RNG seed for the generation of the initial conditions is fixed, but the seed for the simulation is variable (scenario A). The set of parameters to build the initial organism are shown in Table 9.4.

Scenario A We start out with Scenario A, where the same initial condition is used for all simulations, i.e., the same parameters and random number seed value are used in the genetic algorithm that generates resources, cells, and the initial organism. The simulation itself, however, uses a random number generator with different seeds. Each simulation was run for 1000 timesteps in order to focus on the expected initial divergence of the individual trajectories.

In Figure 9.1 we find that different random number sequences for the simulation make for a gradual, but rapid, divergence of the different simulations, even though initial conditions are the same. This is because different random numbers bring about changes in the sequence of mutations that arise in the system, and likewise in the death function, such that the pool of genetic variation is subject to RNG-dependent variation, which in turn influences the further trajectory of the system. The disturbances such random events

---

[1] The initial value for a random number generator algorithm.

Parameter	Value
encoding_size	150
exchange_proportion	0.05
fill_end	1
fill_start	0
min_iterations	2
m	0.3
m_without_select.	0.5
num_population	10
optimization_rounds	600
place_coverage	0.8
N	300
s	0.4
rng_seed	varying
stop_selection_every	50
stop_selection_for	5
length_altering_mut.	true
asexual_intervals	[[0, 3],]

Parameter	Value
exchange_prop.	0.1
m	0.01
m_without_select.	0.5
num_population	2
optim. rounds	200
N	1000
s	0.1
rng_seed	varying
homog._res.	true

Table 9.3: **Left**: Parameters for the initial genetic algorithm (GA) that generates the cell network for the baseline simulation. **Right**: Parameters for the initial GA that generates resources for the baseline simulation.

bring about need not be minor, especially not for low population numbers. Rather, the emergence of a single mutant can give rise to a lineage that increases the populationsize of the system significantly, as seen by the upper curves in Figure 9.1.

Moreover, the exponential rise in populationsize observed in some of them, including the overshoots and subsequent breakdown to a lower population is consistent with mutant organisms being able to exploit other places in the habitat network that have previously accumulated resources, which are then consumed until the influx of resources onto these places can no longer sustain the consumer population.

This, then, leads to a breakdown of the population and a possibly oscillatory behavior, which, however, can be modified by the flow behavior of the habitat's resource network, compare Equation 3.5.5. We will discuss this effect below in more detail.

Finally, simulations seem to group somewhat around a populationsize of ≈ 50, from which some rise to higher rates as seen before. While not confirmed in this section this behavior is consistent with a population in which no interface-changing mutations occur, such that the low-level grouping represents the equilibrium population for the initial input-interface.

Extinction, when it occurs, does happen quickly and does not occur, for the admittedly small number of timesteps, once a population has established itself. Possible reasons have been mentioned above, and are not further discussed here. It must be noted, however, that extinctions, which are "by chance" in this scenario, are not rare (five out of twenty for the sample used here). This could possibly be remedied by improving the genetic algorithm that generates the initial conditions further where the cause is a failure of the organism-habitat pair to function for a population, or by choosing a different initialization scheme where a population is present from the beginning, thus making random extinction for very low population numbers in early phases unlikely. Indeed, in Figure 9.2, we find

Parameter	**Scenario A**	**Scenario B**
asexual_intervals	none	none
energy	50.0	50.0
exchange_proportion	0.05	0.05
fill_end	1	1
fill_start	0	0
G	100	100
i_{max}	2	2
mutation_rate	0.1	0.1
mutation_rate_no_selection	0.3	0.3
num_population	15	15
number_of_runs	5	5
optimization_rounds	500	500
N	750	750
s_asexual	0.75	0.75
s	0.15	0.15
stop_selection_every	200	200
stop_selection_for	10	10
with_length_altering_mut.	true	true
initial_num_agents	1	1
seed	7244785, 58803214, 23881181, 62223675, 85200019, 93735805, 63221134, 31791985, 61638410, 24898659, 1444181, 8540650, 23615654, 49594325, 84361426, 37588962, 4050723, 89848885, 38816287, 87921488	4863152
init_seed	4863152	7244785, 58803214, 23881181, 62223675, 85200019, 93735805, 63221134, 31791985, 61638410, 24898659, 1444181, 8540650, 23615654, 49594325, 84361426, 37588962, 4050723, 89848885, 38816287, 87921488

Table 9.4: Parameters for the initial condition production for organisms in the baseline simulation. The middle column represents a simulation where only the time development has different RNG seeds but initial conditions are always equal, and the right column represents a run where the time development has a fixed RNG seed but the initial conditions vary, with all other parameters being the same.

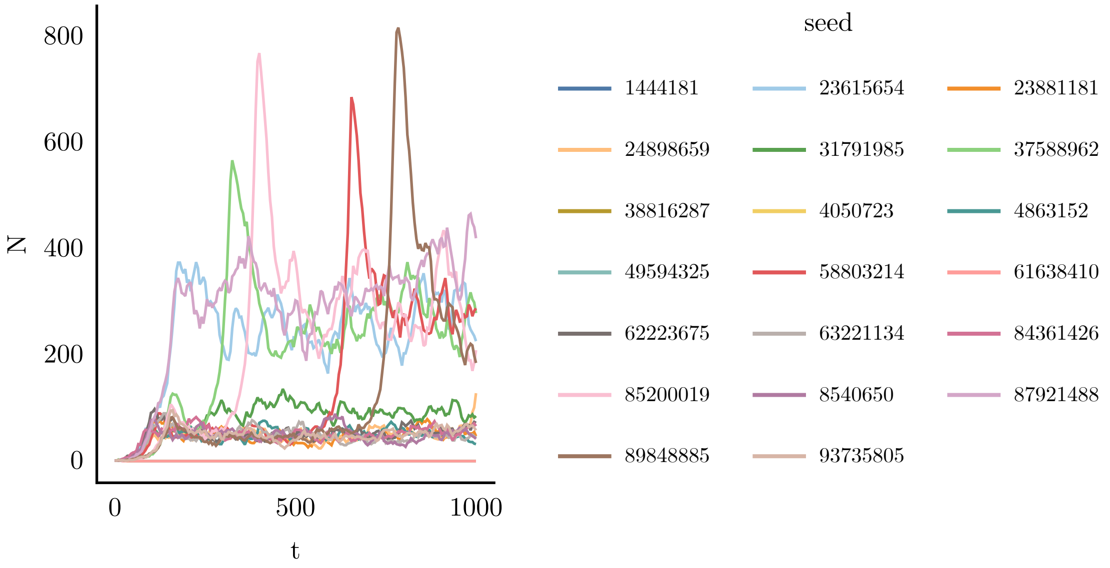


Figure 9.1: Populationsize over time for different RNG seeds for the simulation for scenario B, run for 1000 timesteps. The RNG seeds for the GA that generates the initial organism, cell, and resources are the same for all simulations, such that they all start with the same initial condition. Colors represent different RNG seed values. The different trajectories diverge quickly, although an initial phase with mostly very similar behavior exists. Exponential increases in populationsize with overshoots, well known from population dynamics, are consistent with the occurrence of mutations that allow access to different mesh places, the resource reservoir of which is then depleted until a dynamic equilibrium between resource influx and consumption is reached. Qualitative differences are observed, in that some populations go extinct, while others reach very high populationsizes compared to others. There appears to be a grouping of simulations around a populationsize of 50, which appears to be the equilibrium populationsize for the initial niche, while another grouping appears at roughly 300 which, however, is reached at different points.

that the populations that do go extinct are those that fail to reproduce reliably and that they die out somewhere between 15 and 70 timesteps, which is roughly consistent with a baseline death rate of 0.05 as used in the present simulations.

Scenario B For the second scenario, where the initial condition is changed by changing the random sequence with which the underlying genetic algorithm finds a cell- and organism structure that is deemed functional to start the simulation, we expect the following: While the random number sequence of the population itself will be fixed, the structure of the organisms and habitats may vary significantly. Consequently, initial organisms may take up resources differently because of different metabolism structures, and while all the random numbers that factor into probabilistic decisions like mutation and death are the same for all simulations, the individual properties that can provide respective conditions, e.g., the organism's energy- or material reservoir, are not. Therefore, the behavior of the simulations for low population numbers should be effectively determined by the initial condition, such that we should see significant differences in the population's behavior between different initial conditions even there.

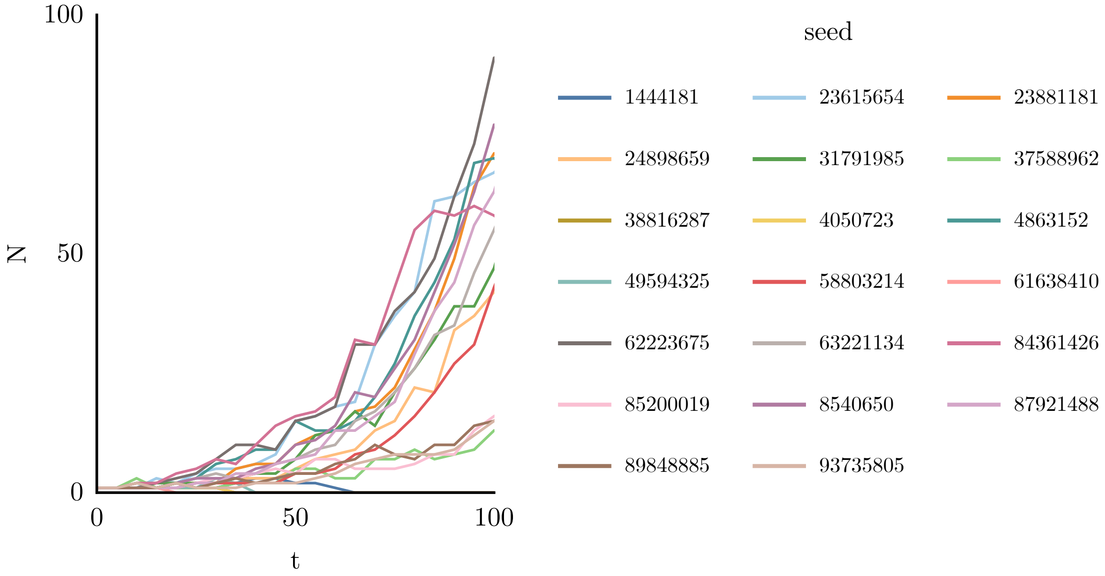


Figure 9.2: Same as in Figure 9.1, but zoomed in to the first 75 timesteps. Different RNG seeds, i.e., different sequences of random numbers, lead to a gradual divergence of the trajectories, as is expected from a nonlinear system into which disturbances are introduced. It must be noted, however, that disturbances for such low population numbers need not be minor in the sense the term is usually used in nonlinear dynamics, but, because of the conditional nature of many organismal state propagators (mutation, reproduction, death), these disturbances can significantly alter the population state, which is consistent with the nature of evolution. Consequently, the divergence between different trajectories occurs quickly. Not all extinctions are visible due to the curves overlaying each other at zero.

Similar considerations apply to the cell's resource network, which is also created without considering a population that feeds on the resources it processes. Therefore, continuous depletion of the resources in some of its places by a population of consumer organisms can hinder future runs of the cell's resource processing network, thereby hindering resource uptake from other places, which is ultimately harmful to the population. Likewise, individual places can gain resources via waste excretion by organisms, which in turn may disable crucial transitions in the cell's resource network if they are connected by an inhibitor arc. Therefore, the different structures that different RNG seeds can bring about in the initial condition can yield qualitative differences in the behavior.

In Figure 9.5, we find that while the phenomenology is generally similar to Scenario A, the behavior is far more variable. Such differences can be explained by differences in structure between cell- and organism-IPNs for different GA seeds. More deeply, this is because evolution is based on imperfect inheritance and is highly path-dependent within lineages. For example, a mutation that in one genome would not cause any substantial difference may, in another one, produces significant behavioral changes. Therefore, a fixed sequence of mutations that occurs sequentially in consecutive offspring may have very different effects depending on the structure of the mutated genomes.

More importantly, different initial conditions generally produce different input interfaces for the organisms and different flow patterns through the habitat network. This influences the resource uptake such that initial organisms which take up and digest resources more

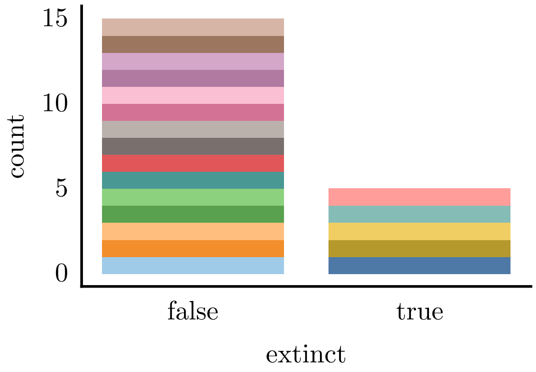


Figure 9.3: Barplot showing for which seeds the population went extinct, using the same color coding as in Figure 9.1. Extinctions do occur rather frequently (five out of twenty).

quickly than others will generally give rise to more quickly growing populations than others, although both are capable of survival and reproduction. Even without further mutation of the interfaces, these differences explain different equilibrium populationsizes of the system by virtue of the finite available material and the mentioned differences in usage patterns. Indeed, Figure 9.8 shows that the input interfaces are rather different between initial metabolism IPNs, although some overlap between individual pairs exists. Because the initial conditions are different, we find a faster divergence of the population curves for different seeds in Figure 9.5, with a closer look revealing some early extinctions, too.

Thus, when the system is initialized with only one individual that can give rise to a colonizing population, the initial condition is of paramount importance for the system's behavior, such that especially early extinctions are associated with different initial configurations and a lot less with disturbances brought about by different outcomes of probabilistic events during the early steps of the population. That being said, the sample shown here is rather limited, and a larger set of seeds should be expected to show chance extinctions in scenario A or a surviving population for some of the extinction cases discovered in Figure 9.5.

The system, all in all, shows the behavior we expect from natural colonizing populations, namely founder effects, a high influence of stochasticity on the early unfolding of the system with a sensitivity to initial conditions and minor disturbances that is the hallmark of nonlinear dynamical systems. While consistent with nature, this property makes it difficult to operate as a scientific experiment and forces us into considering ensembles of trajectories. We cannot know a-priori how well these ensembles represent the phenomenological scope of the system, because there is no closed, analytical expression that gives us information about attractors or bifurcations, and the creation of a bifurcation diagram for all parameters of the system is computationally prohibitively expensive and would additionally intractable due to its high dimensionality.

With regard to the results presented in this subsection, we are additionally forced to choose between a constant initial condition for possible parameter sweeps or constant simulation RNGs, or none of it. Given that extinctions as a consequence of a maladapted initial condition are rather uninteresting, we opt for the first approach in the following sections, corresponding to scenario A.

9.1.2 Ecological niches and ecological species for the baseline case

Ecological niches and ecological species in *Amee*

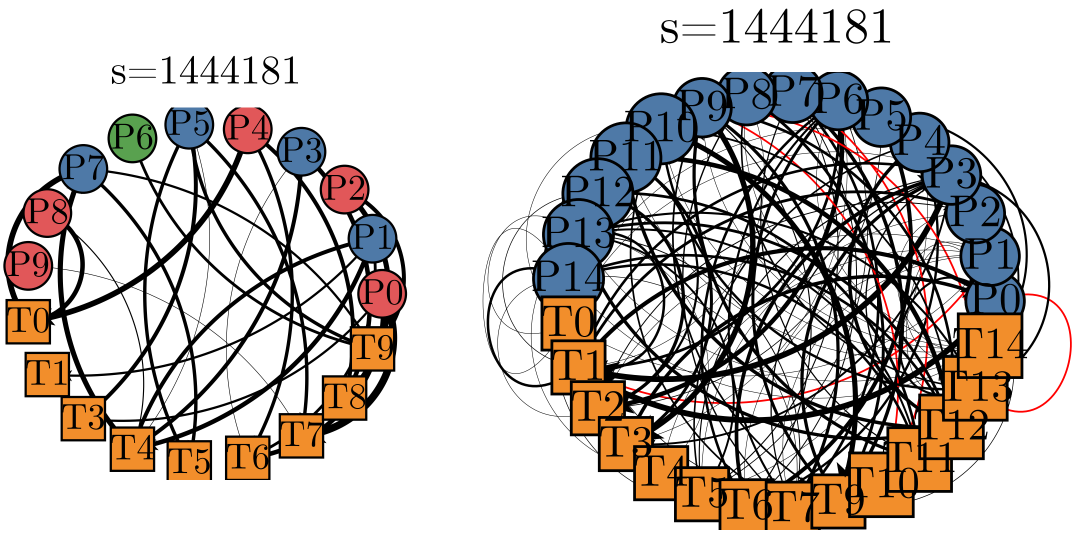


Figure 9.4: Representation of the initial IPNs of organisms and cells. IPN places are blue, transitions are organge, and edge weights are represented by linewidth, with the actual weights being suppressed for clarity. Inhibitor arcs are highlighted in red. Because all parameters for the initial condition algorithm, including the RNG seed, are the same, the initial conditions for all simulations are equal. **Left**: Metabolism IPN of the initial organism of each simulation. The input interface is given by places [0, 2, 4, 8, 9], colored red, the single output interface place 6 is colored green. All other places are intermediate. No inhibitor arcs are present, indicating that they are eliminated by the genetic algorithm. **Right**: Resource processing IPN of the cell. Without rewriting, this is constant for all times and is also constant for all simulations for the same reason the metabolism network is. The habitat's IPN has more possible node labels than the organism initially ($[0, 15]$ instead of $[0, 10]$), and is thus considerably more complex. Organisms can, over generations, adapt to the additional cell places.

Ecological niches have been discussed in section 3.1. In *Amee*, an ecological niche can be defined by the way an organism consumes resources, i.e., by the input interface of its metabolism network and the weights that connect the input places to the rest of the metabolism IPN. This represents the maximum consumption per timestep of each resource on each connected place. When rewriting is disabled, resource consumption is the only kind of interaction an organism has with its environment.

Because the markings and weights of the network are a multiset of resources, we can add an additional index r for each resource, such that the input-matrix I and the output matrix O become 3-Tensors I_{ptr}, O_{ptr} with p, t indicating place and transition, respectively. Using the IPN-formulation in subsection 8.3.3, the niche is then given by

$$N_{pr}^{fund} = \sum_{t \in p\bullet} I_{ptr},\ p \in i \tag{9.1.1}$$

where i represents the input interface of an organism's metabolism network, and $p\bullet$ represents the post-set of the place p, i.e., all the transitions t that receive resources via arcs (p, t) from the place p.

The ecological species concept defines a species as a group of organisms that exploit the same ecological niche, i.e., interact with the same set of ecosystem elements to ensure

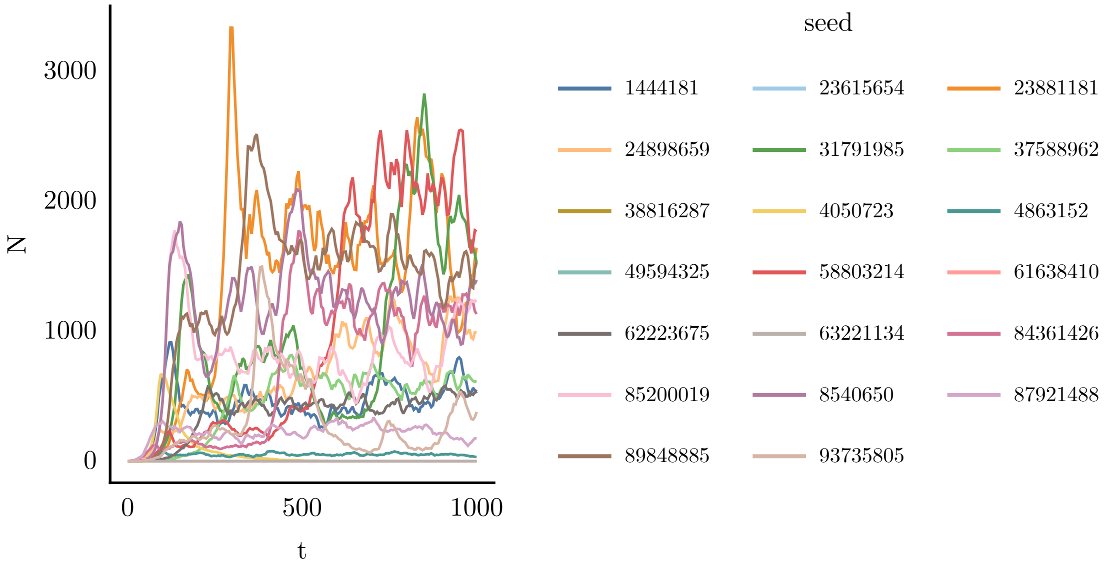


Figure 9.5: populationsize over time for different RNG seeds for the initial genetic algorithm, run for 1000 timesteps. This produces different initial conditions for the simulations, which are then run through the same simulation with the same time development RNG seed value. We find substantial differences in populationsize over time which, while generally similar to Figure 9.1, are significantly more variable, with the grouping at a populationsize of ≈ 50 being missing, and the populations are more evenly distributed across different populationsizes, which is consistent with different initial interfaces.

survival and reproduction. These groups, the concept posits, come about because niche differentiation in an ecosystem drives natural selection in such a way that the phenotypic groupings we call species are produced [Ridley, 2004] and are correlated with their ecological niche.

Realized– and fundamental ecological niches It must be noted that *Amee* naturally produces a distinction between a "realized niche" and a "fundamental niche" in analogy to Hutchinson's niche concept (compare section 3.1): Organisms' metabolism networks have each a number of input places which are defined by the metabolism IPN alone. These connect to a set of places in the habitat networks which have the same labels, but not all input places *must* connect. Rather, it is sufficient if some do. The connected places con are therefore a subset of *in* and the realized niche is given by Equation 9.1.2:

$$N_{pr}^{rel} = \sum_{t \in p\bullet} I_{ptr}, \; p \in \mathrm{con} \tag{9.1.2}$$

Because con $\subseteq$ in, multiple realized niches can exist for a single fundamental niche. We can interpret this as a form of plasticity, in analogy to how natural species exploit different resources depending on availability and other factors. It must be noted that without variation in the encountered habitat networks, however, there is exactly one realized niche for any given fundamental niche because always the same habitat places will be encountered. Therefore, we can ignore the distinction between realized and fundamental niche here because we have only a single cell and no niche construction.

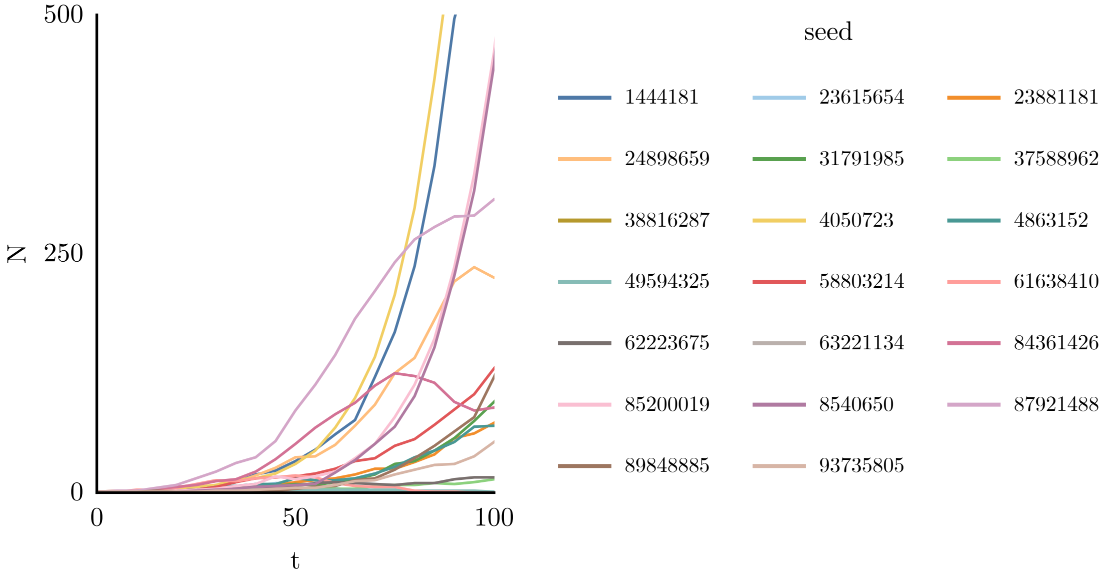


Figure 9.6: Same as in Figure 9.5, but zoomed in to the first 100 timesteps. Like in Scenario A above, divergences are quick, but individual curves diverge more strongly for early phases. This is consistent with the individual simulations having different input interfaces and thus resource usage patterns. Groupings of multiple curves appear to be largely gone.

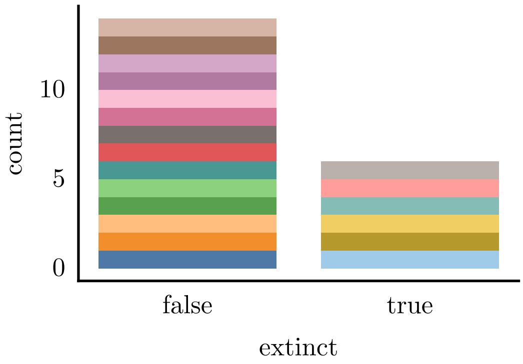
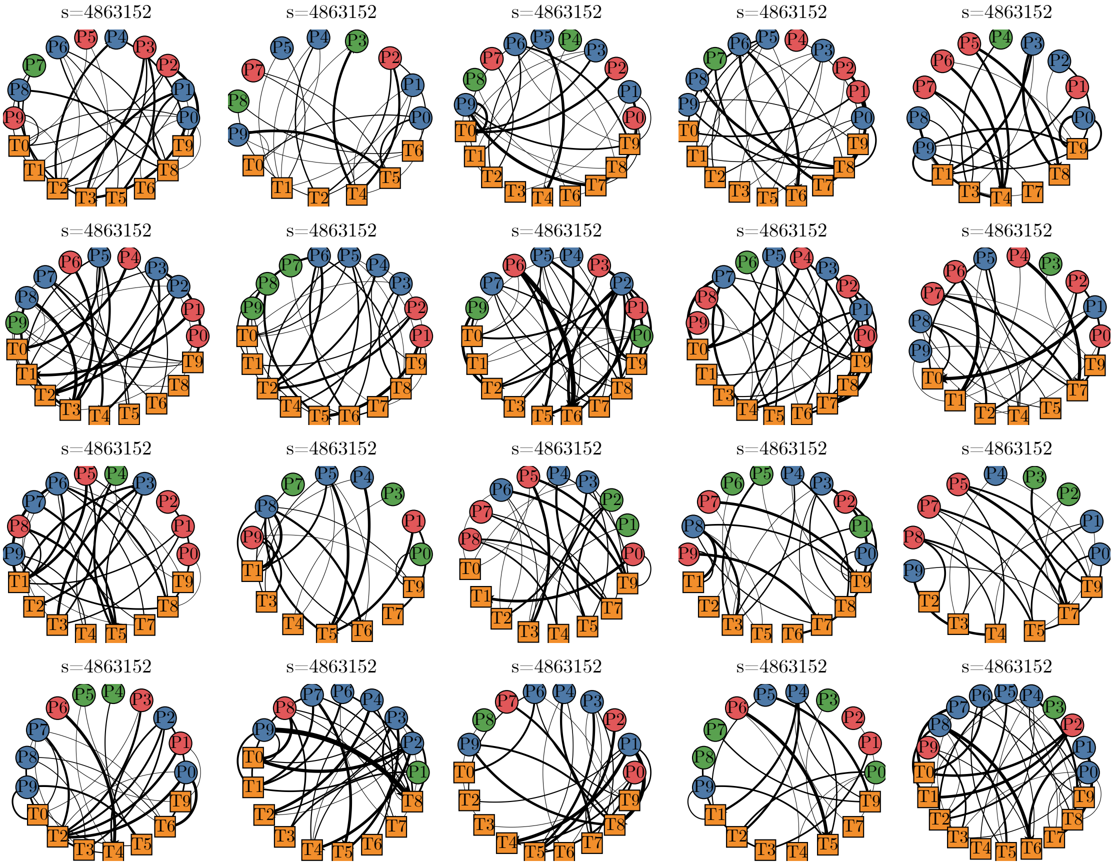


Figure 9.7: Barplot showing for which seeds the population went extinct, using the same color coding as in Figure 9.5. Extinctions do occur rather frequently (six out of twenty).

This changes in a spatially extended system with multiple heterogeneous habitats, or when niche construction is active. The latter can bring about spatiotemporal heterogeneity and thus can result in variation in the realized niche over an organism's lifetime. While this can have substantial evolutionary consequences, their deeper discussion is postponed until chapter 10.

Niche adaptation Consequently, we should expect the niches defined in Equation 9.1.1 to either diversify when the habitat is complex enough or to be dominated by a single ecological species when the niche space is restricted. Secondly, we would expect ecological niches to adapt to the resource availabilities on the places they access, i.e., if the habitat IPN has multiple resources with different abundances on different places, organisms should over time evolve to exploit this abundance pattern for their particular interface more effectively.



s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152
s=4863152

Figure 9.8: Initial metabolism networks created for different RNG seeds in the initial genetic algorithm. As can be seen, the input interface and linkage structure are quite different between all found solutions, which appears to be the cause for the greater variability in the behavior of the different simulations. Inhibitor arcs (in red) are absent, indicating that the genetic algorithm eliminates them.

This can have multiple facets: First, the pattern of exploitation itself can change, i.e., the niche N_{pr} can evolve to include interface places where resources are more abundant. Secondly, the weights, i.e., I_{ptr} can change to accommodate the resource abundance better on each connected place. Third, the metabolism network behind the input interface can evolve to produce the needed resources more effectively. Typically, we will observe all three of these concurrently, with the exact expression and relative importance of each effect being shaped by system details, i.e., the particular history of the system.

Nevertheless, given that a mutation in an interface place is potentially more disruptive than a mutation in a single weight of an arc, we can establish a hierarchy of which evolutionary changes we expect to be more prevalent than others. In particular, the evolution of the underlying metabolism without changes to the niche itself should be the most prevalent because of probability: Assuming that there are fewer places in the input interface (and outgoing arcs from them) than in the rest of the metabolism IPN, a mutation is more likely to affect a place or arc within the network than at its periphery. Variation in the metabolism's architecture on which natural selection can act is thus greater, and evolution should be faster. This includes naturally the arcs that leave input interface places towards the first transitions, such that the adaptation of the niche without changes in the

s=4863152 s=4863152 s=4863152 s=4863152
s=4863152 s=4863152 s=4863152 s=4863152
s=4863152 s=4863152 s=4863152 s=4863152
s=4863152 s=4863152 s=4863152 s=4863152
s=4863152 s=4863152 s=4863152 s=4863152

Figure 9.9: Initial habitat resource networks created with different RNG seeds in the initial genetic algorithm. As can be seen, the input interface and linkage structure are quite different between all found solutions, which appears to be the cause for the greater variability in the behavior of the different simulations.

interface, i.e., niche partitioning without changes in the ecological interactions themselves, should be prevalent, too, but, by virtue of the same argument as above, less than metabolic evolution. Finally, changes in the interface, i.e., a diversification of ecological interactions themselves, should be more difficult, because a change in input interface can be disruptive to the metabolism. Thus, variation of an established niche is more often maladaptive and more limited by virtue of probability.

For now, we limit ourselves to the baseline case without niche construction, plasticity, or development, which will modify the niche concept and the above discussion significantly.

Niche diversity for the baseline case Due to computational limitations, just three intervals for each simulation are investigated. For a runtime of 25000 timesteps, these are $[0, 50000], [10000, 15000], [20000, 25000]$. The simulation is carried out for five different RNG seeds for the time stepping scheme and is started always with the same initial condition. Other parameters are listed in Table 9.5. In order to reduce noise, ecospecies that contain fewer than 50 individuals are ignored in the analysis.

Parameter	**Value**
seed	975392623, 108320434, 59616444, 98555521, 798797005
init_seed	4863152
M_{max}	32000
γ_M	6400
$I_{W,Habitat}$	$[0, 1280]$
i_{max}	20

Table 9.5: Parameters for single-resource simulation for investigating niche evolution. All parameters that are not listed explicitly here are the same as in Table 9.1 Table 9.2 and Table 9.3.

Focusing on the number of niches N over time first in Figure 9.10 together with Figure 9.11, we find that the number of niches generally increases in the first interval as the population grows. This is a consequence of a more far-reaching process: The many reproduction events during this initial colonization phase, where the population grows quickly (compare Figure 9.11), give rise to many possibilities for mutations and thus to an initial diversification. This process does not only produce diversification in the niches but, by construction, also in the metabolism of the organisms. Therefore, this diversification process does determine the populationsize attained by providing variation in niche N and variation in metabolic efficiency: Niches determine how much resources can be taken up, but the metabolism IPNs determine how this resource uptake translates into reproduction, and selection can act on both.

We can see this process in some more detail in Figure 9.11, left column, which shows how the initial rise in total population (grey curve) is dominated by a single niche which gives rise to the majority of the population, diversifying in the process and giving rise to many new niches, which visually correlates well with the corresponding curves in Figure 9.10. However, it is apparent how some simulations produce only small populations in comparison to others, at least for the initial interval, which corresponds to low niche diversity. This behavior is consistent with the population missing key innovations that occur only later. Indeed, we find that for the last investigated interval $\Delta t = [20000, 25000]$ timesteps, the populationsizes are all significantly higher than in the first 100 timesteps or so (as found before) with the last line of Figure 9.11 showing a late occurrence of a key innovation, while for the seed 798797005, the respective key innovation must have occurred somewhere between the first and second recorded interval. Indeed, from the behavior seen until now, these innovations must be such that they allow the population to access another adaptive local maximum space that produces larger populations. This is consistent with the natural evolution of the biosphere, where, arguably on all scales of biological evolution, innovations occurred that allowed for the exploitation of new or different resources and a subsequent increase in complexity and carrying capacity of the entire system, or for more efficient use of resources, with often similar effects. Often, such adaptations are associated with accessing new portions of the niche space. On the largest scale, photosynthesis, adaptation to life on land, as free swimming organisms in the water or burrowing can be named, while on more familiar scales mutations that allow bacteria to exploit new kinds of nutrient molecules in their environment (compare chapter 4) follow a similar principle.

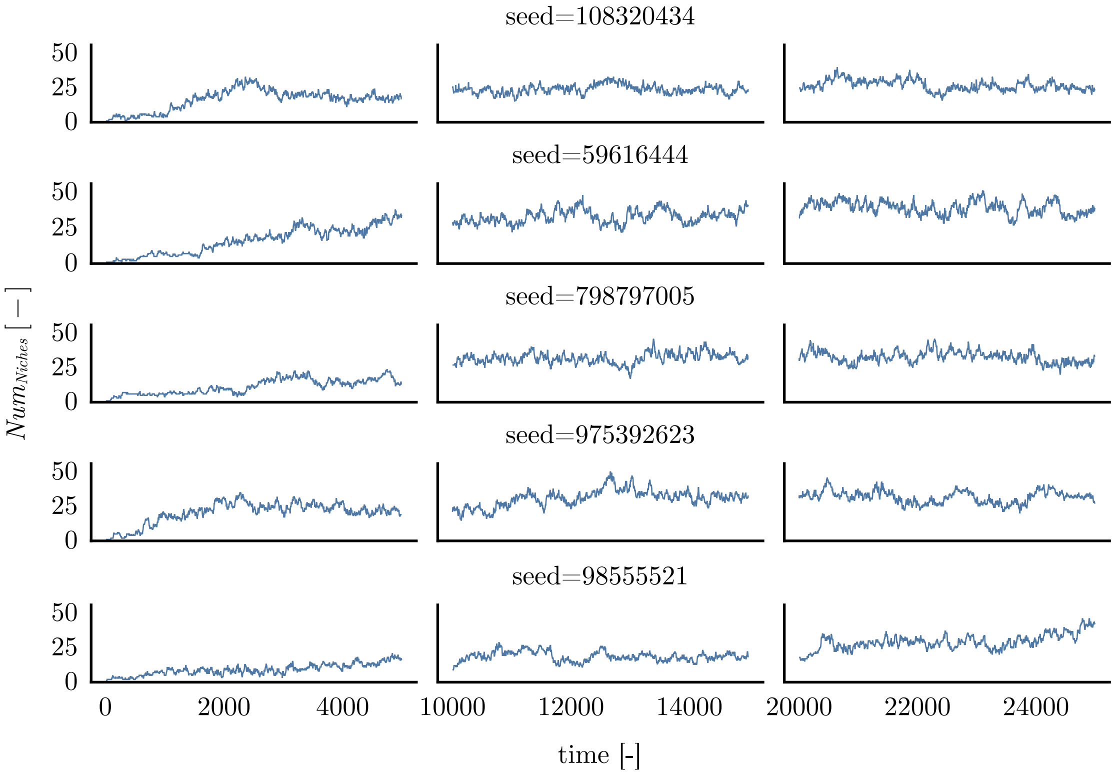


Figure 9.10: Number of different ecospecies over time for the three intervals investigated: left: timestep 0 to timestep 5000, middle: timestep 10000 to 15000, rigth: timestep 20000 to 25000. The number of niches increases initially due to a growing, diversifying population, and then attains a steady state with significant variation in both directions. While the initial increase can be attributed to diversification of niches with the growing population and consequently a growing competition for resources, the subsequent behavior is more consistent with a population whose diversity is limited by the niche space dimensionality and which undergoes ecological overturning through time, i.e., niches arise, rise in frequency in the population, give rise to new niches and eventually get replaced.

The eventual discovery of such more adaptive sections of the niche space is not guaranteed in evolving systems (unless infinite, well-mixed populations evolving for infinite times are assumed) because of evolutionary continuity, compare subsection 3.3.5.

We find also in Figure 9.11 that the originally dominant niche gives not only rise to other niches but in particular, is often more or less quickly replaced by one other niche, the populationsize of which is strongly anti-correlated with the original one, such that it is safe to assume a competitive interaction. This process is mainly seen during the first investigated interval $\Delta t = [0, 5000]$, where the niche space available to the population is not yet fully colonized and adaptation can be expected to be low, although given that in the simulation shown in the last row of Figure 9.11 some key innovation occurs late, this phenomenon can in principle occur whenever a new innovation occurs that subsequently differentiates and comes under selection. This phenomenon furthermore appears to be consistent with the assumptions of adaptive dynamics, where, a single mutation appears in a homogeneous population. This, however, is not always the case, as can be seen in the second line of Figure 9.11 where successive niche replacements overlap at roughly timestep

3000. This shows that for the true adaptive dynamics regime, the mutation rates should be further reduced.

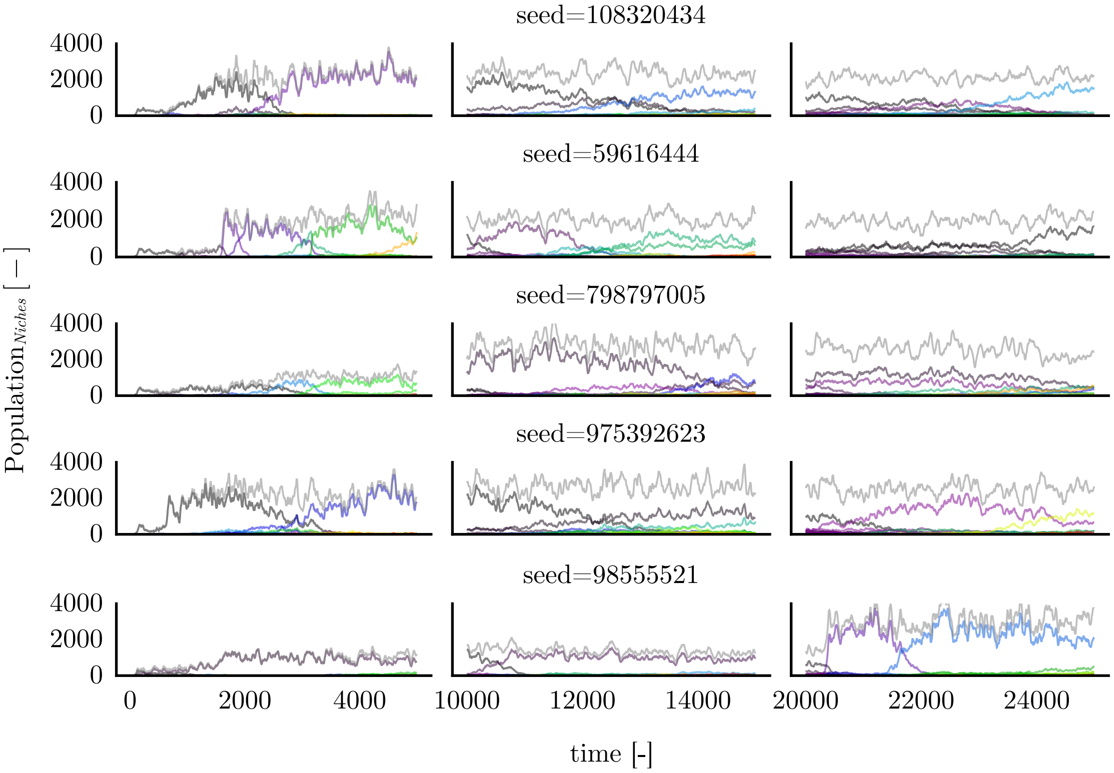


Figure 9.11: populationsize over time analogous to Figure 9.10. The total population (grey line) decays into individual populations associated with a specific niche in the sense of Equation 9.1.1. Often, a single niche (i.e., ecospecies) becomes fixed in the population, i.e., it comprises the vast majority of the population. Over time, these wild-type niche populations produce variants again, such that the population diversifies again. Adaptive-dynamics-like behavior is observed for some cases, where a dominant niche is replaced by a new dominant one (e.g., interval 20000 to 25000, seed=98555521), while at other times, the dominant niche diversifies out into multiple variants that coexist for $\mathcal{O}(1000)$ timesteps. The adaptive-dynamics-like behavior is more frequent for the initial interval, consistent with the expectation that the initial phase of the simulation is dominated by (comparatively) rapid adaptation. Comparison with Figure 9.10 shows that if lower populationsize is attained initially, diversity is lower, too, again consistent with expectations. Colors only distinguish niches. Many ecospecies only attain low populationsizes.

The assumption that this phenomenon is competitive in nature makes sense firstly because of the observed anticorrelations between the populationsizes of the niches involved. Secondly, if it was due to neutral fixation we would expect no change in its frequency over time by virtue of Equation 3.7.3, but would assume a constant rate of replacement on average. Thirdly, for the cases where key innovations occur quickly, we find that this “adaptive radiation” gives way to a regime in which successive niches replace one another more slowly or not at all anymore until the system consists no longer of a single dominant “wild-type” ecospecies but of a collection of, for the time interval considered, mostly coexisting set of niches which comprise the total population. This process is, however, dependent on the history of the system, and is more pronounced in some simulations than in others, consistent with the discussion in subsection 9.1.1. On the other hand, some ecospecies successions occur slowly and show no exponential initial growth, which would be consistent with neutral replacement, especially where there are long times of coexistence between the involved niches. However, in a noisy system, neutrality and adaptive replacement are always mixed. While any quickly occurring species replacement certainly has a strong adaptive component, *Amee* produces always a mixture of neutrality and selection.

For the limited niche space considered here, which consists of a set of fifteen IPN places on the habitat and a single resource, this phenomenology is likely driven by an asymptotic exhaustion of resources by high-impact adaptations, such that the population centers itself within a region of the available space of adaptations, which then transitions into niche differentiation such that coexistence becomes possible, with further successions being mostly driven by neutral evolution. We consequently expect that if the simulation was run several times longer, we would continue to see this asymptotic regime change from a highly competitive initial phase into a less competitive coexistence phase. This regime change is most apparent in the attainment of a more or less steady state in Figure 9.10.

Finally, because new niches come about by mutations carried by single organisms, many more ecospecies are created than have an appreciable impact on the simulation. This can be gleaned from Figure 9.11, where we see how most curves concentrate towards the bottom of the respective graph even when they exist for significant times. Indeed, due to the nature of mutations, we expect most of these to be minor variations of the dominant niches, which are continuously created and go extinct again. Due to lower populationsize, the dynamics of these is more likely to be governed by a neutral process, which should yield a more or less constant overturning of ecospecies at low population numbers, modulated by higher rates of species generation whenever an adaptive radiation occurs, due to the accompanying rise in reproduction events.

In Figure 9.12, we find indeed that the vast majority of ecospecies produce only very few individuals, with only very few attaining high total populations. This is consistent with what we see in Figure 9.11. For most of these low-population ecospecies, their fate will be likely governed more by neutral evolution than by natural selection due to low populationsizes, unless a strong negative mutation precludes metabolic activity and dooms the individual to extinction once its initial energy reservoir runs out.

Therefore, we expect a more or less constant rate of ecospecies overturn due to random mutations and random deaths, with an accumulation of ecospecies in the lifetime histogram for very low but non-zero lifetimes due to destructive mutations, which is visually consistent with the close-to-steady behavior of the number of ecospecies in Figure 9.10 for the latest investigated interval where the niche-differentiation regime has been attained.

Moreover, the scenario of a neutrally governed overall overturning rate modulated by occasional adaptive radiation events fits the lifetime histograms for ecospecies in Figure 9.13. In Figure 9.12, we find, however, that some ecospecies exist for substantially longer than the investigated interval lengths, such that we can expect all lifetimes to be present in the simulation. It has been long assumed that extinction risk is independent of species age in nature [van Valen, 1973], while data from the fossil record has shown evidence for age dependency at least for some groups [Doran et al., 2006; O. Hagen et al., 2017].

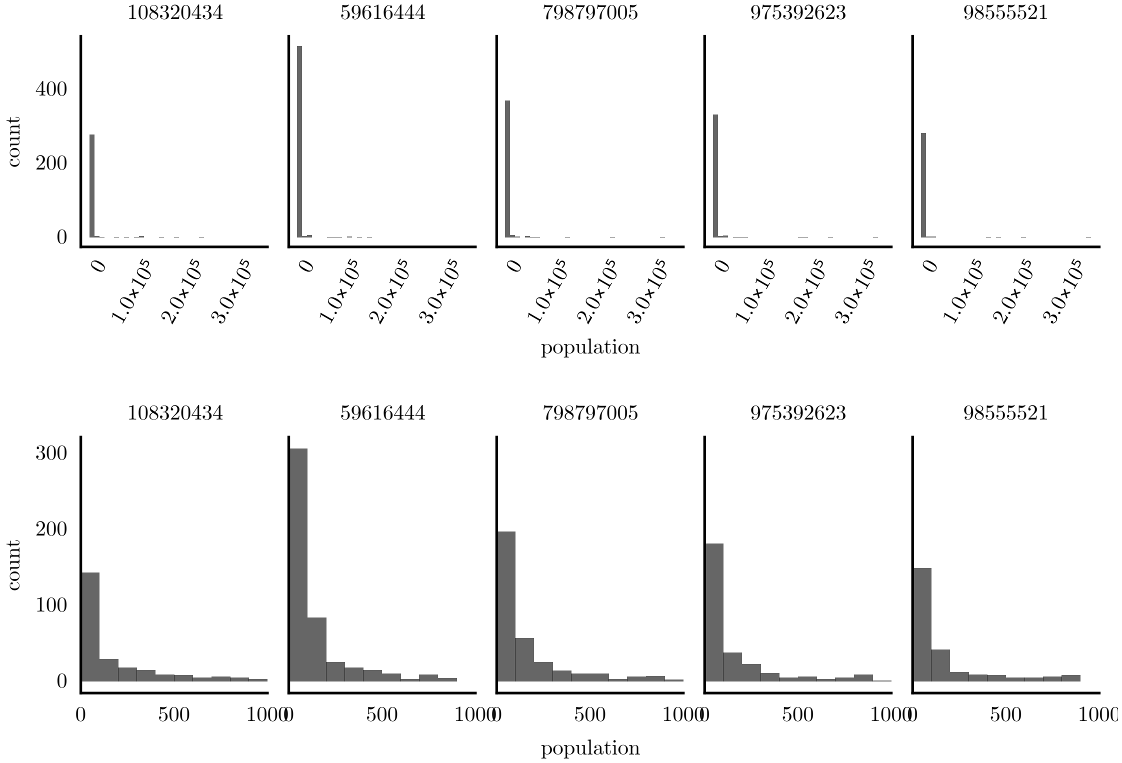


Figure 9.12: **Upper**: Histogram of the number of organisms in ecospecies over the entire course of its existence. The vast majority has low populationsize, with only a few ecospecies attaining very high numbers of individuals over the course of their existence. **Lower**: Same as above, but limited to populations between zero and 1000 organisms. The majority of the ecospecies only attain populationsizes in the double digits.

In Figure 9.14, lower row, we find that ecospecies that survive for a long time can have a high population, but many also produce overall populations of $\mathcal{O}(100)$ or up to 1000. Ecospecies with high numbers but relatively low lifetime are often associated with the first interval of investigation and thus with the initial adaptive diversification process, where a newly discovered niche often rises to dominance quickly, thus producing large overall populations before being replaced again.

Niche evolution for the baseline case

Evolution of resource utilization In light of the findings above, we next focus on the evolution of the niches themselves with respect to the hypotheses on its evolution discussed above in Equation 9.1.2. We find in Figure 9.15 that most simulations do not create diversity in the habitat places they connect to, with the possible exception of the lowermost panel in Figure 9.15, which shows the emergence of the utilization of place 9. Moreover, while the second and third panels from above (seeds 59616444 and 798797005) show an increase in the prevalence of the utilization of place 4, this variant exists in the population from the beginning, whereas the first and fourth panel (seed 108320434 and 97539262) lack this variant and never evolve it. The last simulation has a niche that utilizes places 3 and 9 in the earliest investigated interval but loses them again. The loss of place three is interesting because it represents the input interface of the cell's resource IPN. Accessing this place is not only potentially advantageous because resources could accumulate there

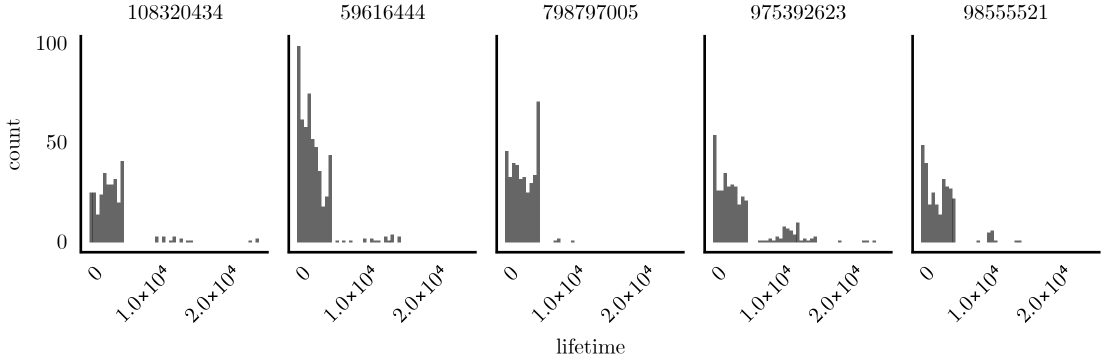


Figure 9.13: Histogram over the lifetime of ecospecies. Due to the interval-based analysis, the measured lifetime has an upper limit at 5000 timesteps. Ecospecies that attain this lifetime must be interpreted as existing for at least that long. Likewise, for higher intervals, older niches that originated before the measured interval are included. Accumulation points of lifetimes in this histogram correlate with adaptive radiation events in Figure 9.11, e.g., in the panel (59616444, [0, 4995]), where most ecospecies are only generated after the population breakthrough at timestep 2000 such that lifetimes are almost completely limited to 3000 timesteps at the most. Also, ecospecies' lifetimes tend to accumulate at very high and very low values at least when the differentiation regime has been reached, indicating that many ecospecies die very quickly due to low population, with most having essentially constant extinction rate, while some, most probably the ones dominating the population, attain very long lifetimes due to being well adapted and having high population.

due to the influx being relatively high, but it also deprives the downstream places of further resources, and consequently could potentially starve other organisms. Utilization of this place also must evolve within the niche, because due to the above reason its usage is forbidden artificially in the initial condition. Assuming the loss of niches that utilize this place is due to selection, we can only assume that the resource network is such that binding to other IPN places in the system is more advantageous, although it stays in the final configuration for two of the five simulations. Furthermore, we find that the adaptation that most likely brings about adaptive radiations and the observed increase in populationsize is the utilization of place 9, which increases in utilization over time and the emergence of which does correlates generally with the quick increase in populationsize that is accompanied by the adaptive radiation phase.

Additionally, in the last investigated interval, we find that the oldest ecospecies are generally those which have the lowest utilization of place 10, and it appears that some lose the utilization of place 5 in favor of place 9. Comparing all five simulations, we find that those which produce a quick initial increase in population and diversification all have comparatively high and more frequent utilization of this particular place when compared to, in particular, the lowermost simulation in Figure 9.15, which, likely due to a long time of evolutionary static, also features a higher diversity in place utilization. This simulation is also the only one in which place utilization disappears (first and second intervals) and later reappears (second and third panels). Given that observations in the other simulations indicate that the particular adaptation that has given rise to the greater populationsize is associated with the quantitative utilization of place 10, it is likely that the

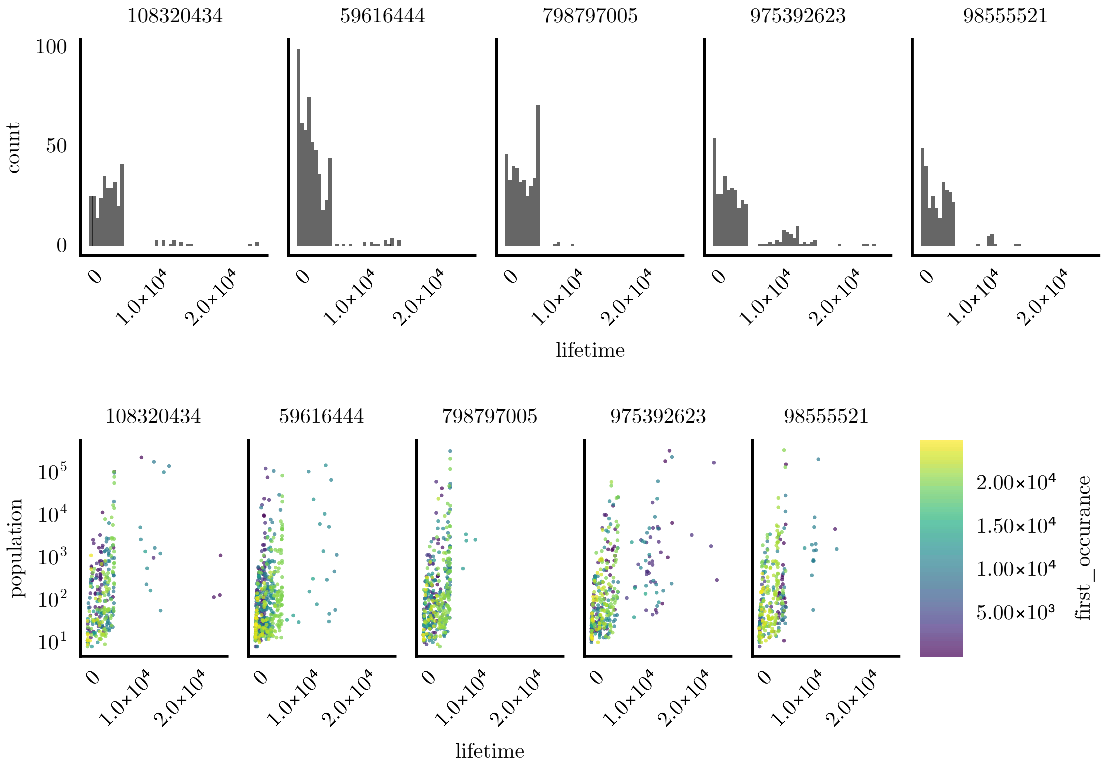


Figure 9.14: **Upper**: The data of Figure 9.13 accumulated over all intervals. While the majority of the data is limited to the length of the investigated intervals of 5000 timesteps, some ecospecies exist for long enough to transcend this interval length, with individual ecospecies existing for the entire duration of the simulation for the leftmost simulation. Hence, we can expect all possible lifetimes to be present in the simulation. **Lower**: Total population over lifetime. While the data is incomplete due to the interval-based analysis, the system shows that even ecospecies with very low total numbers can survive for a long time, although relatively rarely. Some populations with low lifetimes but high population are probably associated with initial competitive diversification where niches rise to high frequency and get replaced quickly. The sharp edge at a lifetime of 5000 is due to the length of the interval length for which data was taken.

low utilization of the place in the beginning failed to produce a sufficient fitness advantage to prevent extinction. Finally, because the habitat's resource IPN is the same in all cases, all systems converge to similar place utilizations which is determined by the topology of the former.

In summary, we find that evolution in place utilization is indeed slow or absent for the systems discussed here, as expected. Rather, the different ecospecies differentiate in the way they use the resources on the places that their ancestors utilized, often back to the very beginning of the simulation (compare the initial conditions in Figure 9.16). Therefore, the coexistence pattern we found earlier must be due to a differentiation in the pattern of access to the different places, such that, e.g., one ecospecies uses place X a bit more than place Y, while another does the reverse. This pattern is consistent with the expected niche differentiation which has, e.g., been shown to be present in many systems, one of them being experimental plant communities [Zuppinger-Dingley et al., 2014], in bat

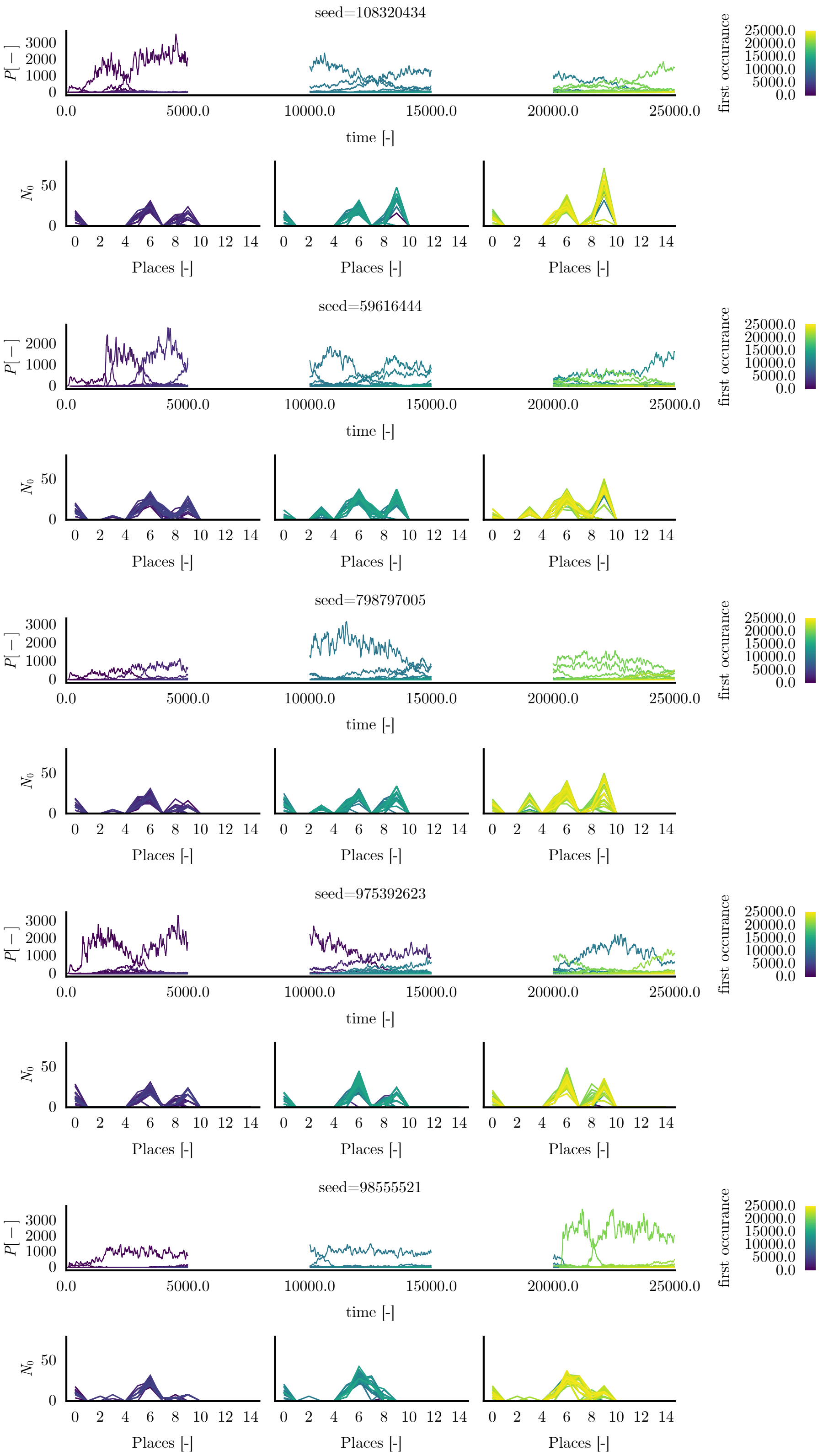


Figure 9.15: Ecospecies evolution over time. For each simulation (distinguished by random seeds), the time evolution of the different niches in each interval is plotted (top panel), and below each, the niche of each ecospecies is shown accordingly. Both are colored by time of first occurrence. As we can see, there is little overlap between the communities in each interval, and all systems tend to converge to a similar general form of ecological niche determined by the habitat IPN structure. Niche differentiation appears to happen primarily on the utilization pattern of ancestral places instead of in the utilized places themselves.

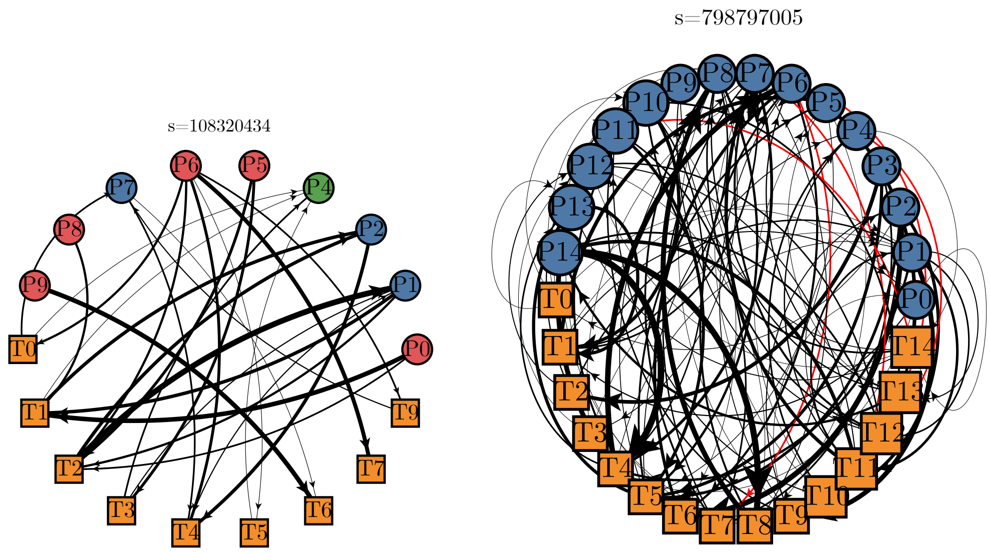


Figure 9.16: Initial conditions for all simulations presented in Equation 9.1.2, input interface colored in red, output in green. **Left**: Initial agent's metabolism IPN. **Right**: Initial resource IPN of the Habitat.

communities [Siemers and Schnitzler, 2004] or in primates [J. R. MacKinnon and K. S. MacKinnon, 1980]. Indeed, the dense distribution of the niche graphs in Figure 9.15 show seems to support this argument, further strengthened by the observation that these densely populated bands appear to be the broader the less diverse the habitat-place utilization is in the population.

Evolution of the metabolism While the resource utilization patterns appear to be in good agreement with our expectations so far, the evolution of the metabolism networks is yet an open topic. Given their networked nature, there is a plethora of possible analyses that can be applied, with research in the structure, topology and functional organization of biological metabolic networks being a large and active research topic. Indeed, tools and concepts from this field could be transferred to this model in order to further test its phenomenology, or tailor it towards simulating the evolution of such networks in a more detailed way. That being said, the field is quite highly specialized, and analysis of the metabolic networks in *Amee* is beyond the scope of this thesis. Therefore, we limit ourselves to a cursory glance, deriving conclusions from the visible network structure in a small sample of individuals.

In Figure 9.17, the metabolism of randomly chosen individuals from the largest ecospecies at timesteps 2500, 12500, and 22500 is shown, assuming that advantageous metabolic pathways should be preserved across different ecospecies over time, as indicated in Figure 9.15. Given that the connection to place 9 was of particular interest before, we focus on it first. A natural expectation would be that the connection along certain paths that efficiently link input and output would increase in weight, such that the organism can metabolize more resources in a shorter time. Indeed, this is the case in Figure 9.17 for seed 975392623, where the connection P9 $\mapsto$ T6 $\mapsto$ P3 forms an isolated cluster that rapidly produces resources on the output place. The connection from T6 to the output, however, is missing in other individuals (see seed 59616444 on the right, for example), where T6 has no outgoing connections at all. A closer examination shows that such cases

Figure 9.17: Sample networks for the ecospecies with the largest populationsize at times 2500, 12500, 22500 (from left to right). Same color conventions as in Figure 9.16. Focusing on the connections of place 9 and comparing them with Figure 9.15, we find that the connections increase over time in strength where the place is accessed. This can happen because a single-transition connection to an output place arises (e.g., for seed 975392623, right), but also in order to block "parasitic transition" (e.g., seed 798797005, see text), which would destroy resources. In such cases, the place connection can even vanish from the niche.

are relatively common. A transition that has no outgoing connection can still fire, but because all outgoing connections are effectively zero, the material that is transferred from its pre-places is lost, as is its energy[2]. We call these structures "parasitic transitions", because they often, in particular when they have input interface places in their pre-set, harm the metabolic efficiency of the organism and thus its reproductive success. Thus, it is evolutionarily advantageous to shut off these transitions or relink them into a useful pathway. While the latter possibility has been discussed before, the former can be achieved in much the same way: A transition t can only fire when $m(p) \geq w(p,t)\, \forall p \in \bullet t$ in the Petri-Net formulation used here, such that a higher weight $w(p,t)$ can effectively preclude a parasitic transition from becoming enabled, thus precluding loss of material and allowing for metabolic efficiency. Of course, a loss of the corresponding interface place is also an alternative, which can be observed for seed 59616444 in the right panel in Figure 9.17. However, in the case of the strong parasitic transition in Figure 9.17, right panel for seed 7989797005, place P9 does not appear to be part of any important path, although the similarly connected place P7 is. Therefore, the strong connection P9 $\mapsto$ T6 appears to be, in this case, a secondary switch that aids the pathway P8 $\mapsto$ T8 $\mapsto$ P7 $\mapsto$ T5 $\mapsto$ P4, and any other pathway P7 is involved in because it precludes the parasitic transition T6 from subtracting resources from it. It must be noted, however, that we only have a handful of individuals here from a population of several thousand, therefore, these explanations just give a coarse insight into the range of phenomena possible, and a more thorough study of the system would be needed for a comprehensive overview.

Alternatives to this secondary switch mechanism would include inhibitor arcs, which would disable transition T6 as long as any one of the places in its pre-set contains resources, the loss of the connection P7 $\mapsto$ T6, possible rerouting via other links or the emergence of an outgoing link from T6 to some other place, as is the case in other simulations. When discussing the "why" of such alternatives, however, it must be noted that evolution acts on variation that exists in a population, and given the architecture of the metabolism IPNs, one plausible explanation for the emergence of the switch mechanism is that this particular ecospecies' population contained variation in the P9 $\mapsto$ T6 link that lead to variation in reproductive success on which evolution could act, but lacked significant variation in the possible alternatives. Arguably, the usage of an inhibitor arc would be more efficient and probably preferable to a human engineer, yet such a particular mutation is rather unlikely and thus variation would be small or absent. Variation in edge weights can, in particular, be brought about by gene duplication, which immediately doubles a particular weight, and produce a potentially impactful variation. At the same time, because the metabolic network is complicated, we would expect that established pathways, even if suboptimal by a human designer's standards, will be preserved over time and strengthened or modified, while potentially advantageous innovations are hard to evolve in a way that benefits fitness. Thus, *Amee* represents well evolutionary continuity, the relative nature of natural selection. Its complexity aids in this because it allows for some of the, at times, confusing and counterintuitive compromises, path dependencies and limitations of natural evolution to autonomously evolve and be studied.

More generally, it stands to reason that for the case presented here, i.e., with homogeneous resources, or at least large resources with limited variability, the metabolism will evolve towards fewer steps between input and output, because reduced complexity provides for a more reliable resource supply for reproduction. Figure 9.17 provide hints towards this

---

[2] Note that this is not a violation of energy conservation as imposed here, because the system is allowed to be arbitrarily dissipative, and material conservation is not imposed in the resource networks.

because we see in the fourth panel from above (seed 975392623) how the path P9 ↦ T6 ↦ P3 is successively strengthened over time, which represents the shortest possible path between input and output interface. Consequently, we would expect the system to evolve towards such a “survival of the flattest” regime [3] The alternative, i.e., many, complex, highly inhomogeneous and small resources in a system with many possible nodes should produce more complex metabolism networks which would arguably be the more interesting case for a study of metabolism network topology.

We consider the shortest path between the input– and output interface of the metabolism networks. For a set of nodes $V = \{v_i, i = 1, \dots, N\}$ that makes up a graph g and a weight function $f : V \to \mathbb{R}$, the shortest path between nodes v_0 and v_n is the tuple $(v_0, v_1), ..., (v_{n-1}, v_n)$ for which $w = \sum_i f(v_i, v_{i+1})$ is minimal. Because we are interested in the paths for which resource transport is maximal, however, we use a modified weight function $f_I : v \mapsto 1/f(v)$. This reasoning is grounded in the expectation that the paths between interfaces where resource transport is maximal are the evolutionarily most relevant ones because they tend to dominate the resource flow through the network.

The result for the three intervals data has been recorded for can be seen in Figure 9.18 for different ecospecies, where each dot represents the mean shortest paths between interface places for an individual (the mean of all shortest paths that connect an input place to an output place), and the white dots show the mean of the population for a given ecospecies. We find that from the beginning, the mean shortest path lengths for individuals vary between two and four with a population mean (white dots in Figure 9.18) between two and three predominantly at the lower end. Non-existent paths have been excluded, such that a shortest path length of two is the minimum possible length. This shows how the most relevant paths between the input– and output interface are often short, and while there are longer dominant paths (e.g. in the middle interval, $t \in [10000, 15000]$ for seed 59616444), these ten to decrease over time again or at least show no tendency to increase the metabolic complexity.

Lifetime reproductive success and metabolic evolution In order to investigate niche differentiation further, we consider the lifetime reproductive success of the individuals belonging to individual ecospecies in Figure 9.19, Figure 9.20 and Figure 9.21.

In these figures, only ecospecies whose maximum population over all timesteps they exist for exceeds fifty individuals are shown. Additionally, all organisms which don’t reproduce during their lifetime or only exist for a single step have been filtered out to make the evolutionarily relevant part of the lifetime reproductive success distribution more apparent. Previously, we identified adaptive and neutral successions of ecospecies, with the former being mainly associated with the first interval for which data have been recorded.

Indeed, we find in Figure 9.19 that the mean lifetime reproductive success for successful ecospecies which replace others is generally higher than it is for the ecospecies they replace and that the mean lifetime reproductive success tends to increase with time, i.e., ecospecies that occur later have generally higher lifetime reproductive success than earlier ones. Moreover, we find that the increase in the mean lifetime reproductive success over time generally flattens off after a series of adaptive successions, transitioning into a plateau with fluctuations (compare Figure 9.20 and Figure 9.21). This strengthens the argument above that the slowly progressing successions in later intervals are driven mostly by neutral

---

[3] “Survival of the flattest” has originally been used to describe another evolutionary phenomenon [Sardanyés et al., 2008], but given that the IPNs discussed here proceed in a step-wise manner, it seems appropriate.

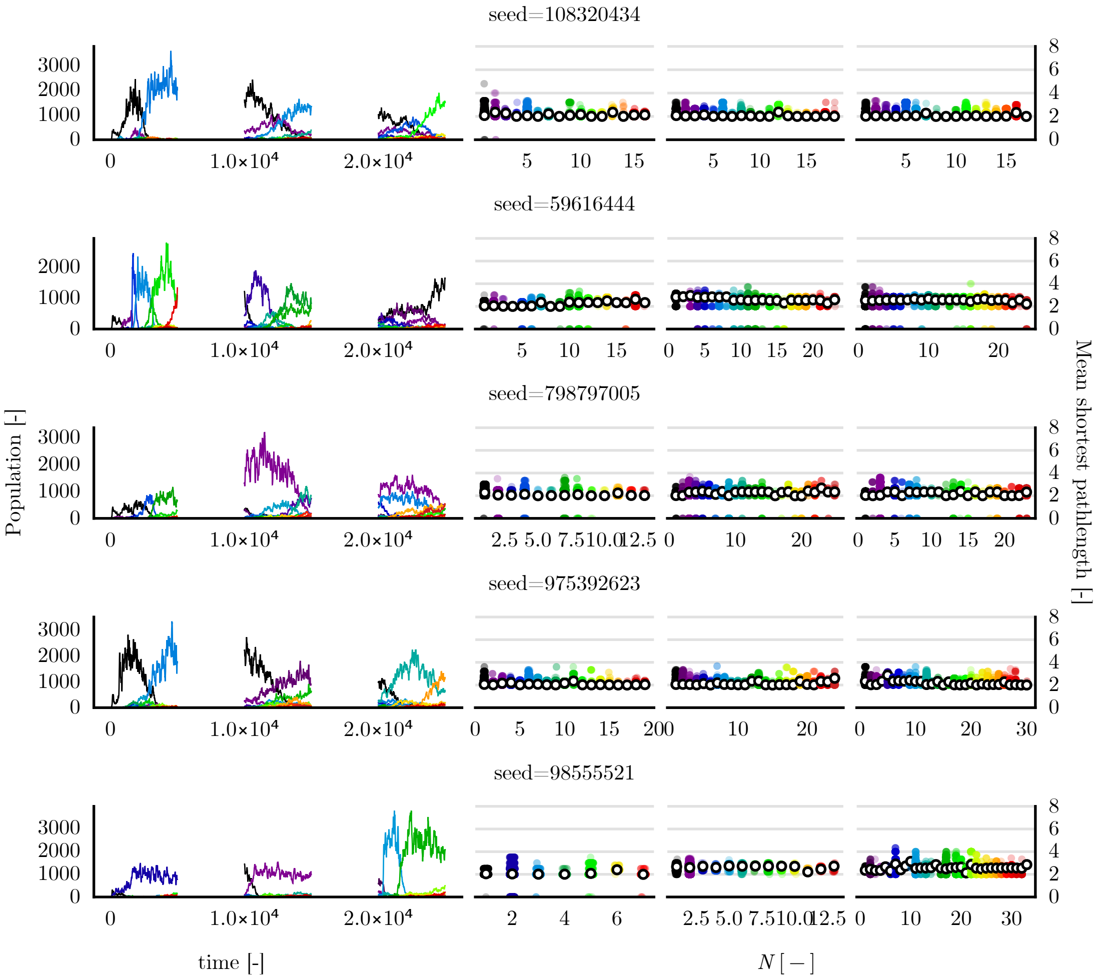


Figure 9.18: Ecospecies populations for time intervals $[0, 5000], [10000, 15000], [20000, 25000]$, with ecospecies with a peak population of less than 50 organisms filtered out for clarity. Distributions of the mean shortest path lengths per individual for each respective interval are shown to the right, with the population means shown as white circles. Colors identify ecospecies. We find no long-term trend to increase the shortest path between interfaces, i.e., no tendency to produce more complex metabolisms than are present in the beginning. An increase is evident for seed 98555521 in the final interval (rightmost panel) for the extremes of the distribution, similar to seed. This fits well with the "survival of flattest" hypothesis discussed above, although we cannot deduce from this that the shortest possible path will always become dominant eventually.

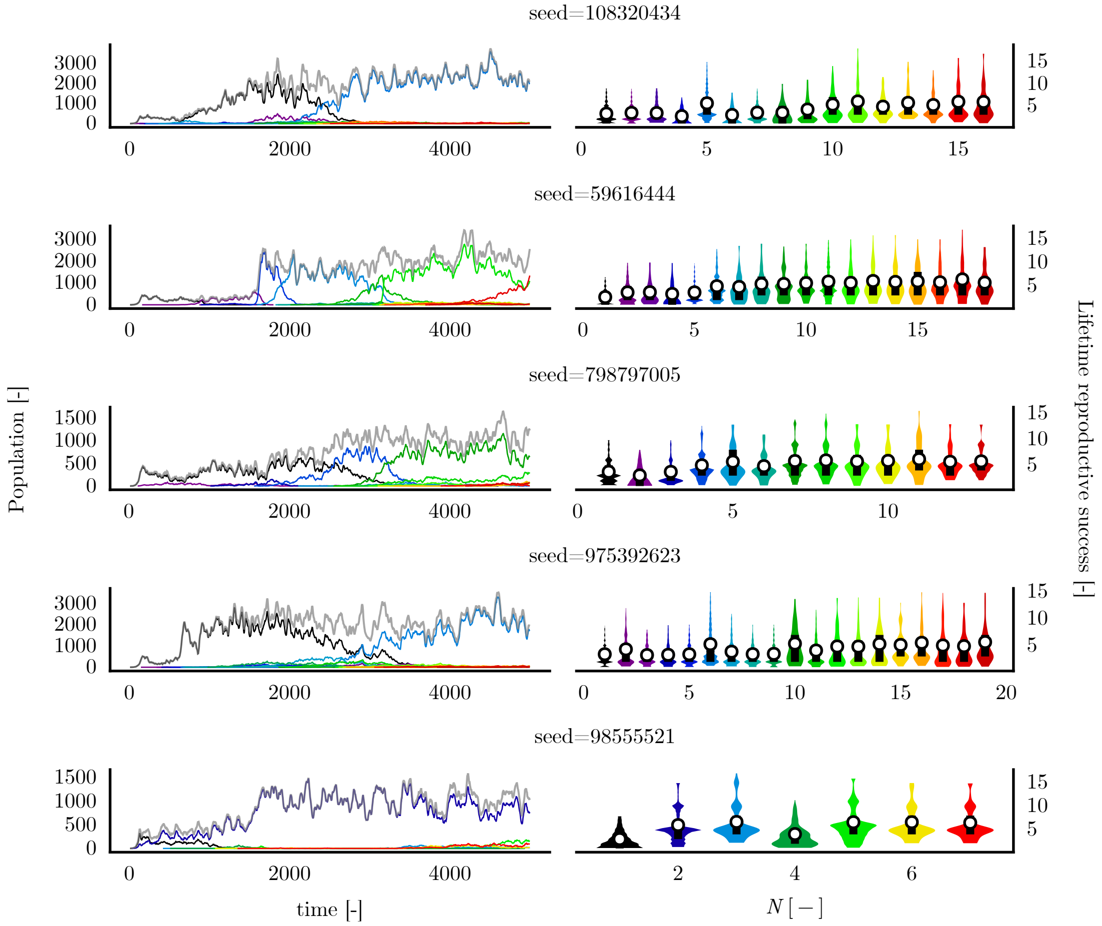


Figure 9.19: Lifetime reproductive success of ecospecies for interval $[0, 5000]$. Colors distinguish ecospecies. Ecospecies with less than a maximum population of 50 have been filtered out. **Left**: Population of each ecospecies over time. Grey line shows the total populationsize. **Right**: Violin plot of lifetime reproductive success for each ecospecies, with the mean shown as white dot and quartile distance as black bars. Violinplots extend from five to ninety-five percent quantile. We generally find an initial increase of mean lifetime reproductive success over time, which is correlated with the adaptive succession discussed before, leveling off once an adaptive phase transitions into a neutral turnover phase.

succession, while the exponential growths of some ecospecies that are mostly apparent in the first interval are driven by adaptations that confer direct fitness advantages, i.e., these successions are competitive. Moreover, because lifetime reproductive success does not decrease over time for the neutral succession phase, we can conclude that the populations are not affected by mutational meltdown, which likely will only play a role for much larger simulation times (compare Figure 3.5) for the given mutation rates.

A strange case presents itself in the lowest panel of Figure 9.21 (seed 98555521), where the newly emerging ecospecies with a much higher population start an adaptive succession phase which contains ecospecies with much lower lifetime reproductive success than the ones they replace. As mentioned before, this simulation gets “stuck” in the initially occupied subset of the nichespace, and only evolves innovations that make another part

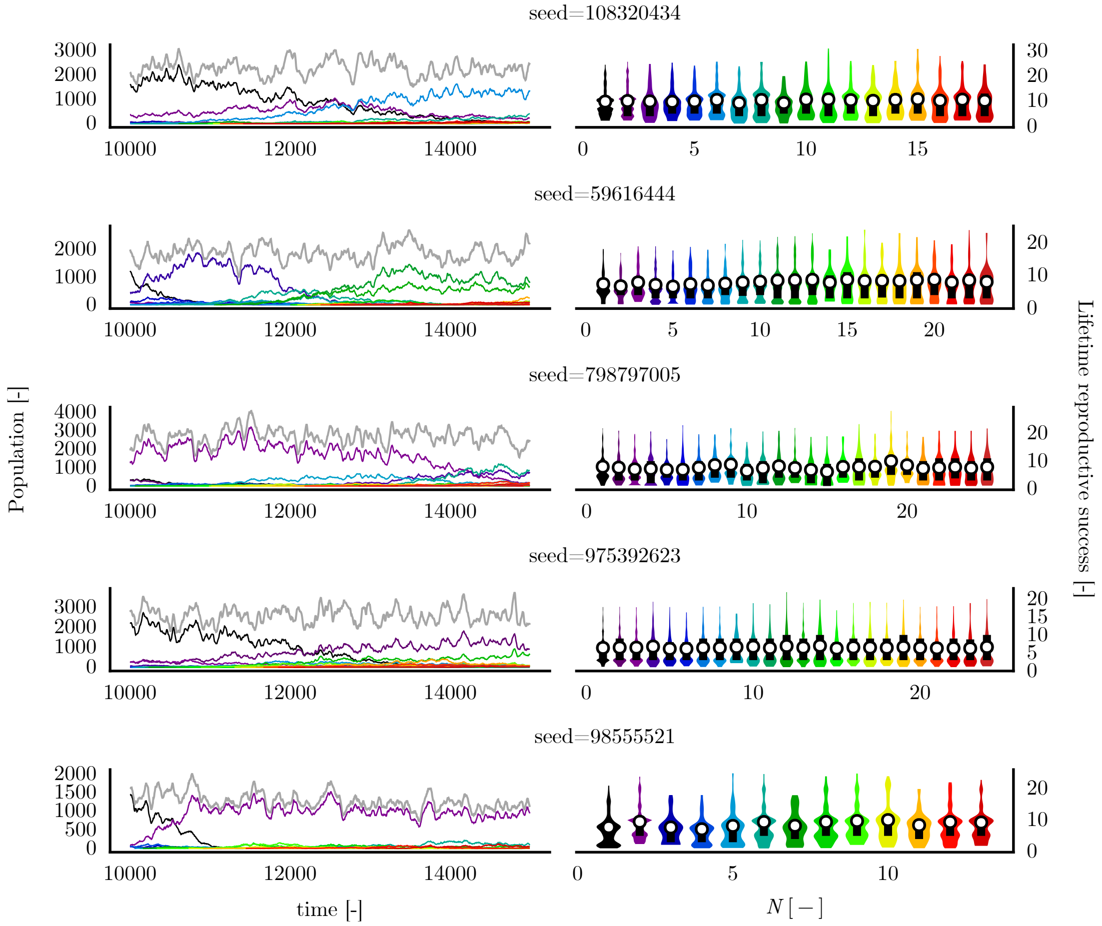


Figure 9.20: Lifetime reproductive success of ecospecies for interval $[10000, 15000]$ as in Figure 9.19. Mean lifetime reproductive success is roughly equal for all ecospecies, as is the quartile range, supporting the case for neutrally dominated ecospecies successions after the initial adaptation phase is over.

with a higher carrying capacity available later. Consistent with this, there are fewer ecospecies in this population in the first interval (Figure 9.19), and it remains in a neutral phase for most of the simulation time until $t \approx 20500$.

This behavior appears to contradict the mechanics outlined above but also shows that the evolutionary innovation that leads to the higher-population state represents a fundamental transition in ecological function, at least for a time.

A possible way for such a state to manifest is that the individual lifetime reproductive success of organisms is reduced (which in this model means that they die younger because only one offspring can be born each timestep), but more organisms reproduce in the ecospecies' population, i.e., the effective populationsize of the ecospecies is increased. This will primarily be accomplished through metabolic innovation, either directly in the niche composition of the innovating ecospecies or in the network structure of the metabolism.

In Figure 9.22, we see indeed (lowest panel, seed 98555521), that the fraction of reproductively active organisms goes up significantly once the higher carrying capacity state is reached, as can be seen from the correspondence of the colors which mark different

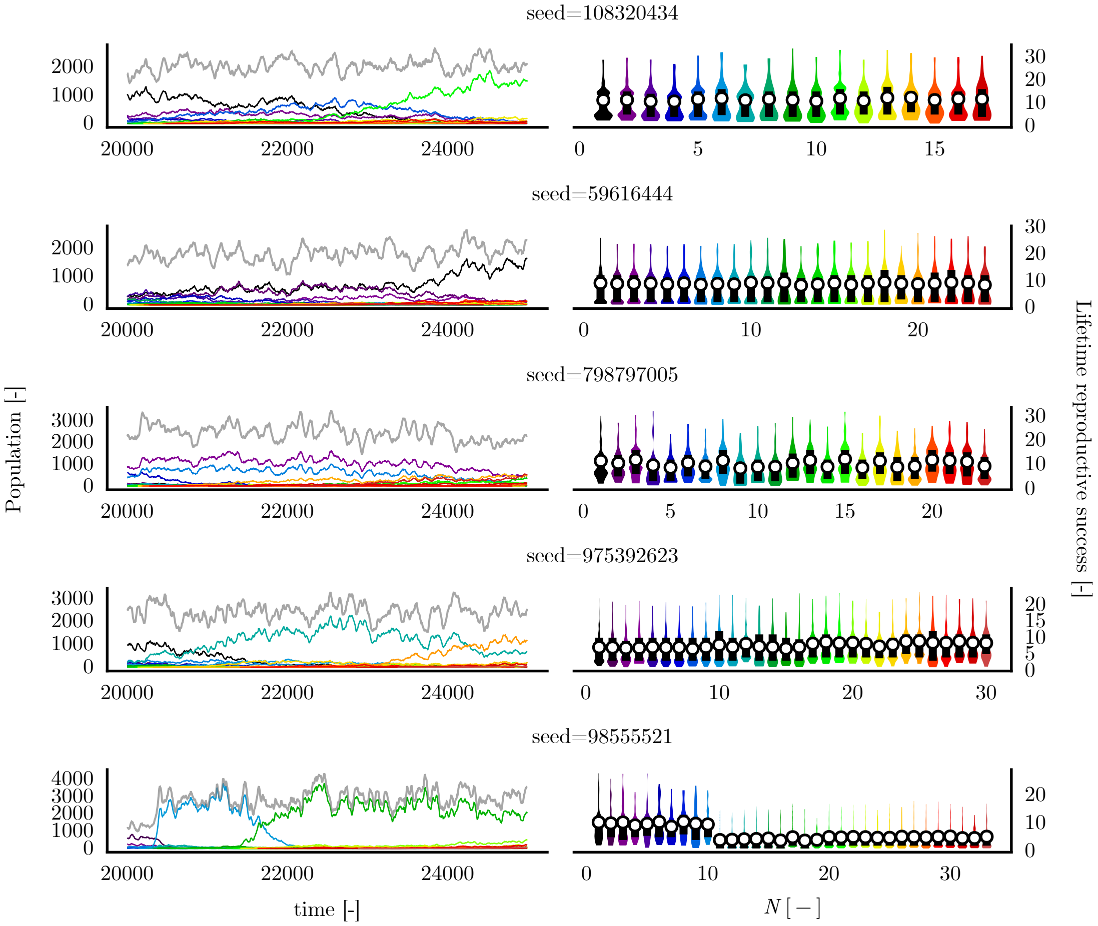


Figure 9.21: Lifetime reproductive success of ecospecies for interval $[20000, 25000]$ as in Figure 9.20 and Figure 9.19. For seed 98555521 (lowest panel), an adaptation phase is observed, characterized by the exponential population increase of the light blue ecospecies around timestep 20500. This leads to a transition to much lower reproductive success for the subsequently emerging niches.

ecospecies in Figure 9.22 and Figure 9.21, and attains values which in other simulations are only attained during the initial phase of the simulation, compare Figure 9.23. In these, as well as for the seed = 98555521 case, we find that these high effective populationsizes are not retained, but slowly converge to values between 0.15 and 0.2 with the levels of fluctuation being dependent on the simulation, but being generally small.

Furthermore, all simulations and all ecospecies, during evolutionarily neutral phases, converge to a near-universal fraction of reproducing organisms which depends on the simulation but always lies between 0.15 and 0.2. This convergence is attained as a simulation settles into a local fitness equilibrium, i.e., transitions into an ecological state dominated by neutral successions, and is only broken through evolutionary innovations that happen during the first phase of the simulation (see Figure 9.23) or when an evolutionary innovation happens later like for seed 98555521 (and probably for seed 798791005, where it is not in the recorded interval, however). This convergence mirrors the static observed for lifetime reproductive success in Figure 9.19, 9.20 and 9.21. While the convergence towards a universal fraction of reproducing organisms again shows that the

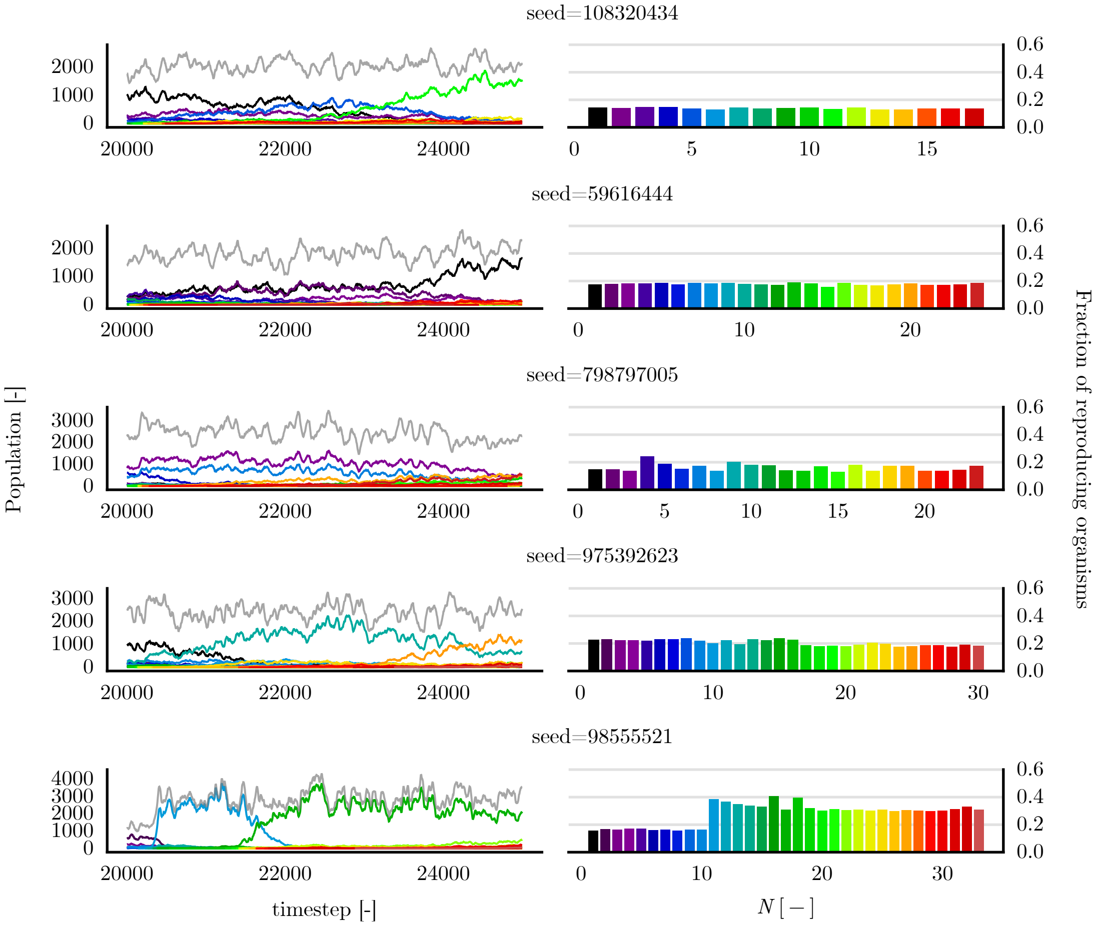


Figure 9.22: **Left**: populationsizes of each ecospecies over time for comparison like in Figure 9.19. Colors distinguish ecospecies. **Right**: Associated fraction reproductive individuals in the population of each ecospecies. This fraction goes up significantly for seed 98555521, sinking again afterward towards the level other simulations attain. The equilibrium level for the reproducing population fraction is around 0.15 to 0.2 with very limited fluctuation between species.

simulations tend towards evolutionary static, why it attains this exact value has not been investigated further here.

Given that it is similar, but not equal in all simulations makes an artifact of the simulation algorithm unlikely, and the behavior observed does not hint at further convergence of the simulations towards some truly universal value that would only be attained for longer runtimes or so. Rather, the effective populationsize seems to be a function of the competitiveness of the system, which we would expect to slightly differ between individual simulations even when they develop similar innovations and attain similar carrying capacities and niche states. A higher mutation rate should negatively influence this value because it will invariably produce more defective mutants which cannot reproduce, with the adverse being true for the baseline death rate d_0.

Summary of metabolism evolution for the baseline case We thus find that metabolic evolution is complex and intertwined with the evolution of resource utilization, with

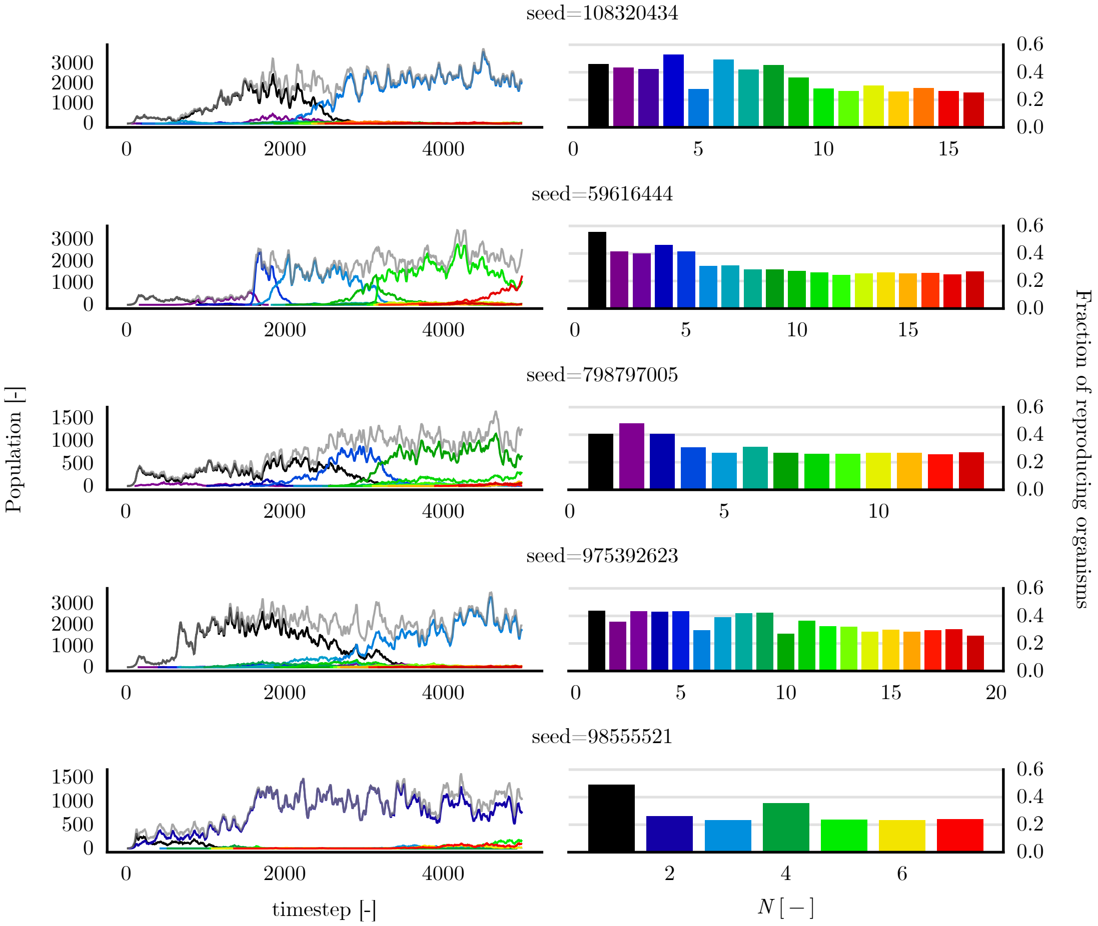


Figure 9.23: Same as Figure 9.22, but for timesteps $[0, 5000]$. We find the same phenomenon encountered for seed 98555521 in Figure 9.22 in general, which shows that the increased effective populationsize is indicative of adaptive ecospecies successions.

relinking of metabolic pathways having potentially large influences. This is somewhat similar to the behavior expected from gene regulatory networks under mutation, which are ultimately underlying metabolic networks. Figure 9.17 also allows the supposition that the evolution of habitat-place-utilization is, in fact, more prevalent than is apparent from Figure 9.15, although this has not been investigated further. The above discussion shows, moreover, that even if mutations in habitat utilization would be more common, the metabolic network that processes these resources must react evolutionarily to these new resources, which can be a complex and difficult process that therefore is slow. Simulations with larger runtimes and higher mutation rates should provide more insight here. Investigation of the lifetime reproductive success of the organisms in individual ecospecies uncovered a pronounced equilibrium behavior, with a slightly simulation-dependent fraction of reproducing individuals. Mean lifetime reproductive success and the reproductive fraction per ecospecies were mostly constant within a neutral ecological system. Interestingly, while mean lifetime reproductive success during early times shows slight increases over time, in later adaptive phases the opposite is the case, associated with a sharp increase in the reproductive fraction of each ecospecies. This hints at a transition of the system into a more competitive regime with higher population

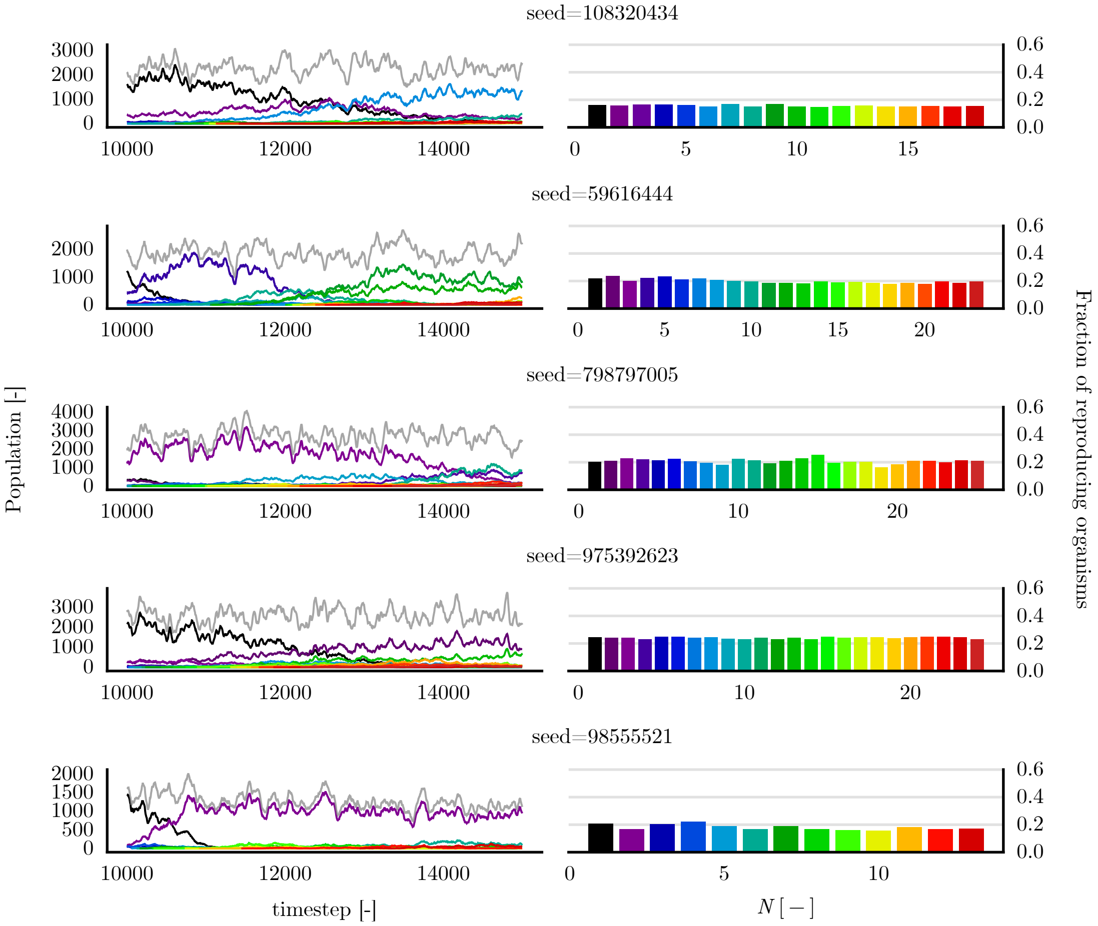


Figure 9.24: Same as Figure 9.22, but for timesteps [10000, 15000]. We find all simulations in an ecological equilibrium state dominated by neutral successions, with low fluctuations of the reproducing population fractions between ecospecies.

density. After the system passes again into the neutral succession mode, the reproductive fraction again starts to converge toward the universal value. The origin of this value has not been investigated further, although we can expect it to be tied to parameters influencing adaptiveness, like mutation– and death rates.

Finally, because organisms distribute undigested resources again onto the habitat places that correspond to their output interface, they may increase the availability of resources in these places and thus produce adaptive opportunities for other ecospecies. Much like measuring niche overlap between ecospecies can give rise to a dynamic network of competition, the overlap between input and output can give rise to a network of facilitation, which may have significant overlap with the former, but which may also benefit largely unrelated ecospecies or generally give rise to radiations that are dependent on a "keystone species" providing resources on specific habitat elements. While principally within reach, this has not been investigated further, but is held to be a promising research direction and one that holds opportunities for further development of the model.

Higher-order species groupings As can be seen from Figure 9.15, most ecospecies are highly similar, and differentiate within a certain resource utilization pattern. Therefore, the question arises whether a higher-order structure can be found that collects very similar niches into a group like it is done in phylogeny. To this end, The DBSCAN clustering algorithm [Ester et al., 1996] has been applied, using pairwise distances given by Equation 9.1.3.

$$d(i, j) = 1 - \tau(N_i, N_j) \tag{9.1.3}$$

with τ begin the Kendall-τ correlation measure. The choice to use Kendall-τ instead of another correlation measure like Spearman's or Pearson's, or a Euclidean distance, is that we are interested in a measure for the *form* of resources utilization, i.e., how similar the form of the niche-curves in Figure 9.15 is. Kendall-tau considers the number of concordant and discordant pairs of values in the niches considered, thus naturally placing the emphasis on the form of the curves, which makes it a natural choice. While other correlation measures could supposedly be used as distance measures for the DBSCAN algorithm, a Euclidean distance has no obvious meaning as a distance measure for niches because it ignores the utilization patterns over network places.

The DBSCAN algorithm interprets the provided matrix as an adjacency matrix of a graph, then uses a supplied radius for establishing a neighborhood to be considered for each point, and finally produces clusters by essentially executing a flood-fill algorithm on the graph such established. The measure given in Equation 9.1.3 is such that distances are larger for smaller Kendall-τ. Three different radii have been tried out (i.e., three different sizes of neighborhoods), and the silhouette measure s [Rousseeuw, 1987] has been used to estimate the quality of the clustering. $s \leq 1$ and a mean s for a cluster that is close to one tells that the data points in the cluster lie well within it on average, i.e., the cluster is well defined. The silhouette measure for th i_{th} datapoint in a cluster c with datapoints $c_1, \dots, c_i, \dots, c_j, \dots$ is given by

$$\begin{align} s_i &= \frac{b_i - a_i}{\max(a_i, b_i)} \tag{9.1.4} \\ a_i &= \langle d(i, c_{j \neq i}) \rangle \tag{9.1.5} \\ b_i &= \min_{k \neq c_i} b_{ik} \tag{9.1.6} \\ & \tag{9.1.7} \end{align}$$

with $\langle \bullet \rangle$ being the arithmetic mean, a_i therefore being the mean value over the distances to other points in the same cluster, and b_{ik} being the average distance from the i-th point to the k-th cluster, such that finally $s \in [0, 1]$.

While DBSCAN is a pioneer algorithm in the field that has since been developed further, it is still viable when the data structure is known well enough to choose the required parameters, and is one of the algorithms in the used software package[4] that does not make assumptions about the distance matrix.

Comparing Figure 9.25, Figure 9.26 and Figure 9.27, we find that for a large radius parameter for the DBSCAN algorithm, the algorithm substantially underfits which is corroborated by low silhouette values in Figure 9.28. Indeed, a radius of $\epsilon = 0.05$ appears to offer the best compromise of the values tested here, although more research would be needed to fine-tune the clustering. Indeed, for some simulation intervals, the silhouette values indicate that even for this value the clustering is suboptimal, which correlates with

---

[4] Clustering.jl of the Julia programming language.

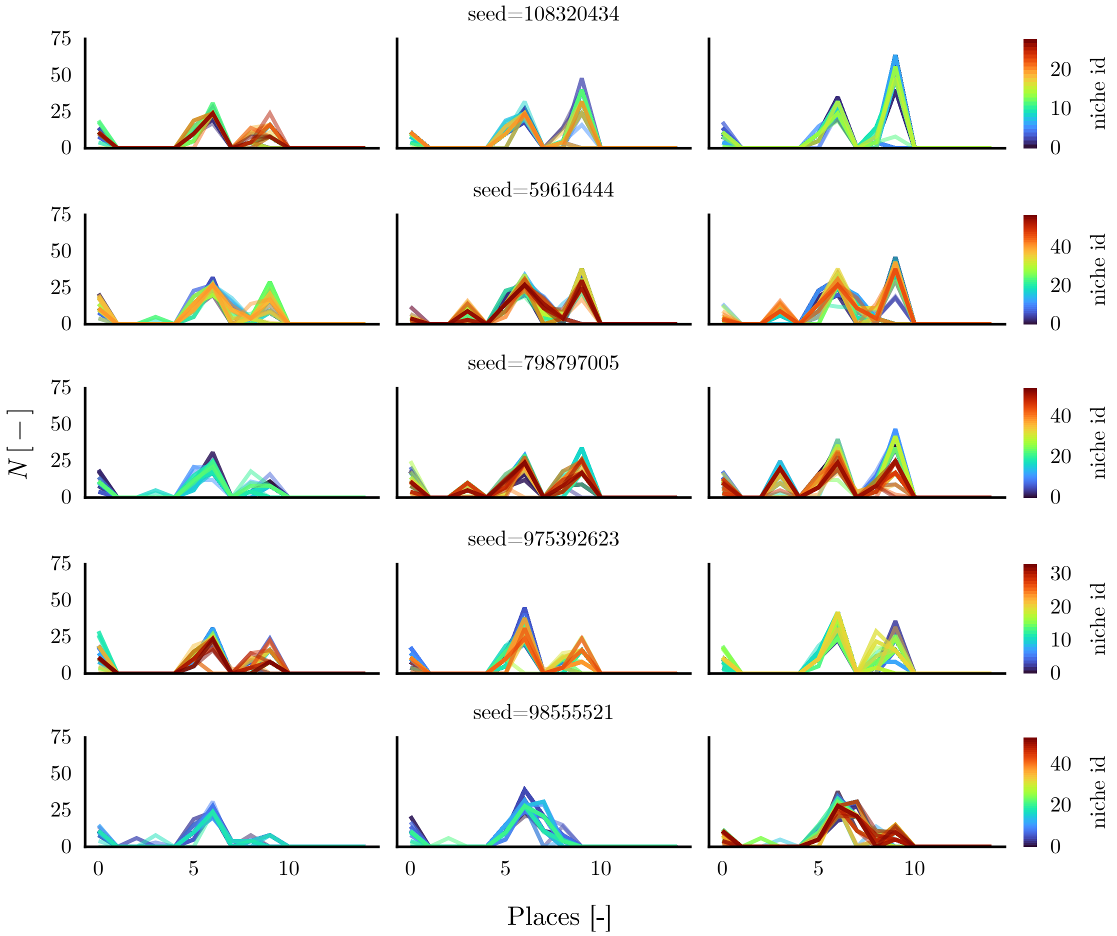


Figure 9.25: Niche clustering for a radius parameter of $\epsilon = 0.005$. The radius appears to be too small and thus the clustering overfits significantly, producing a number of clusters that is comparable to the number of ecospecies, i.e., datapoints (compare Figure 9.10).

the number of habitat utilization variants. Another possibility would be to use a different distance measure, which more accurately captures the different forms of habitat utilization that do exist. Additionally, other algorithms might be more suitable for such problems, although fine-tuning this analysis goes beyond the scope of this thesis. Even with these shortcomings, however, the results presented here appear to be promising, in that the system does produce some form of higher-order grouping of ecospecies similar to genera in biology.

Like in nature, this separation is at least somewhat dependent on the way we measure the similarity between different ecospecies, with the Kendall-τ based function being one possibility among possibly many. This harkens back to the discussion on species Figure 3.1, where the pragmatic nature of these classifications was introduced. Indeed, clustering on the basis of ecological niches observed would, in nature, probably also produce something that is correlated with physiological species, but would not map perfectly onto them.

In Figure 9.29, we find that the found ecosystem clusters live, naturally, a lot longer than individual ecospecies, as is the case in nature with genera and families when compared to species. Secondly, the adaptive radiations found in Figure 9.15 now mostly collapse into

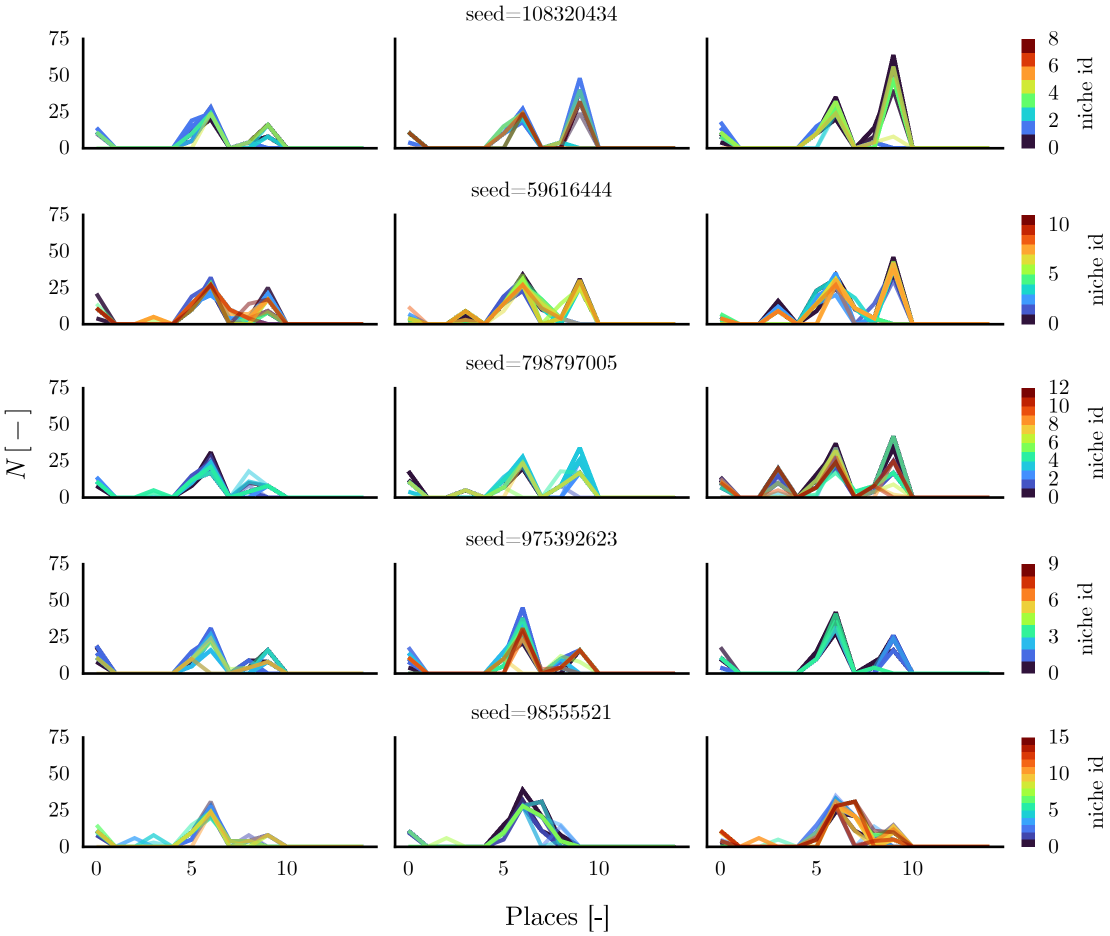


Figure 9.26: Niche clustering for a radius parameter of $\epsilon = 0.05$. Visually, the separation into different clusters according to resource utilization pattern is successful. However, Figure 9.28 reveals that the clustering is not optimal in some cases, which we identify here as those with significant variation between ecospecies, like the lower right panel for seed 98555521.

single clusters, which occasionally give rise to new higher-order ecospecies conglomerates. In more detail, it appears that the clusters that emerge early and become dominant within the first interval generally survive until the end of the simulation, with the coexistence of more than one of these clusters being reduced to one case (seed 798797005). This behavior would, in principle, allow us to build a phylogeny of the ecospecies and eco-genera in the system, and, although the methodology is far from perfect, shows that *Amee* produces such clusters naturally as a consequence of natural selection. This opens the door to the investigation of macroevolution and questions like species selection or long-term diversification. However, in order to investigate the succession of eco-genera, much longer runtimes of the simulations would be needed.

For the lowest panel in Figure 9.29, we also find that survival of low-population eco-genera into late simulation stages can still result in the production of daughter genera that then adaptively radiate, which shows that the possibility for adaptive radiation is not bound to the age of the genus that radiates. We can find similar phenomena in nature,

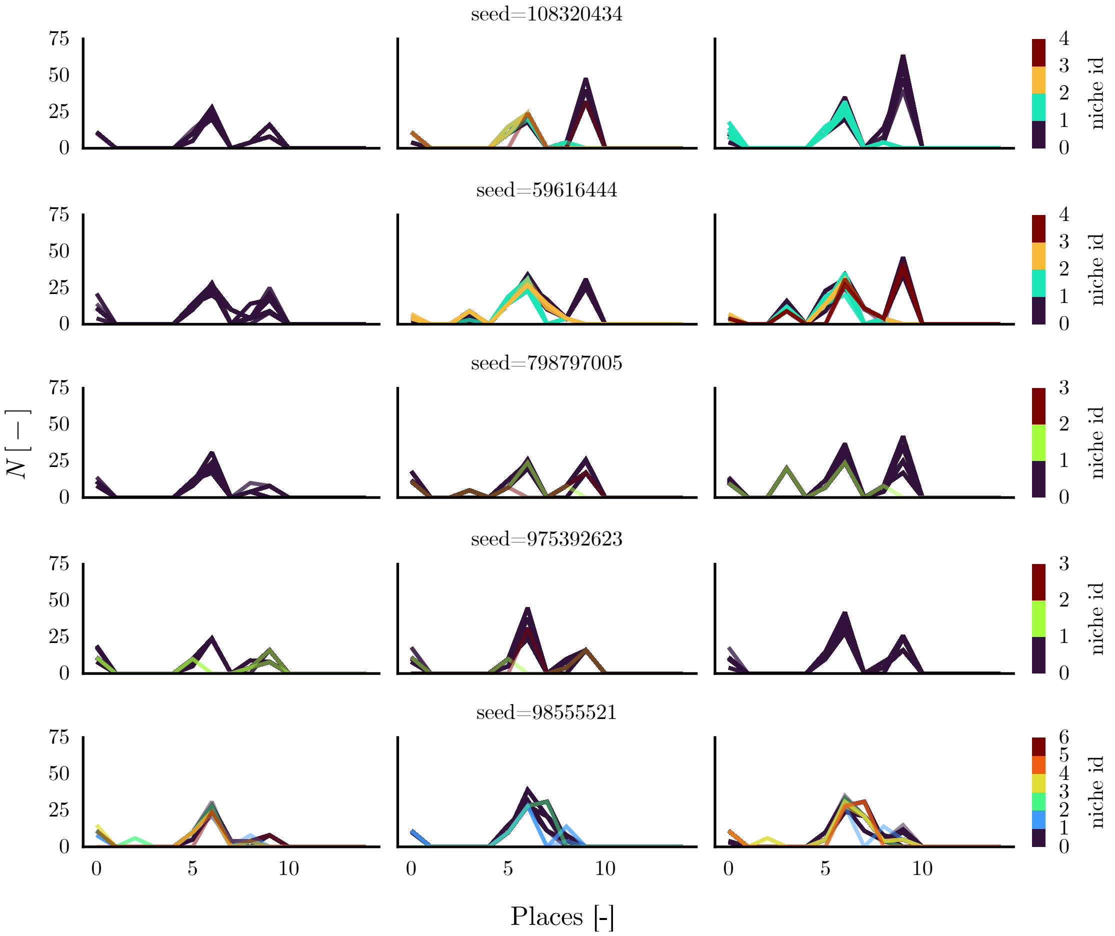


Figure 9.27: Ecospecies clustering for $\epsilon = 0.1$. The algorithm underfits for this parameter, with different habitat utilization patterns being mapped to the same cluster.

a prominent one being the rise of mammals after the K-T extinction event, although in physiological terms.

Mammals are members of the group of synapsids and thus belong to a lineage that dominated terrestrial ecosystems in the Permian age, lost most of their lineages during the end-Permian extinction event, had subsequently been largely marginalized for around 150 million years while archosaurs (in particular dinosaurs and pterosaurs) rose to dominance and, after the latter's extinction, radiated again.

All in all, while the results provide evidence that *Amee* behaves along the lines expected from natural evolution, the methodology in this chapter can be substantially improved using other methods which, ideally, do not rely on user-supplied parameters that influence the clustering. Performance improvements are also required because the current algorithm uses $\mathcal{O}(n^2)$ memory and runtime, n being the number of ecospecies. Such scaling is not sustainable when longer runtimes and larger systems are to be investigated, which are needed to investigate macroevolution or generally the long-term ecological behavior of the system.

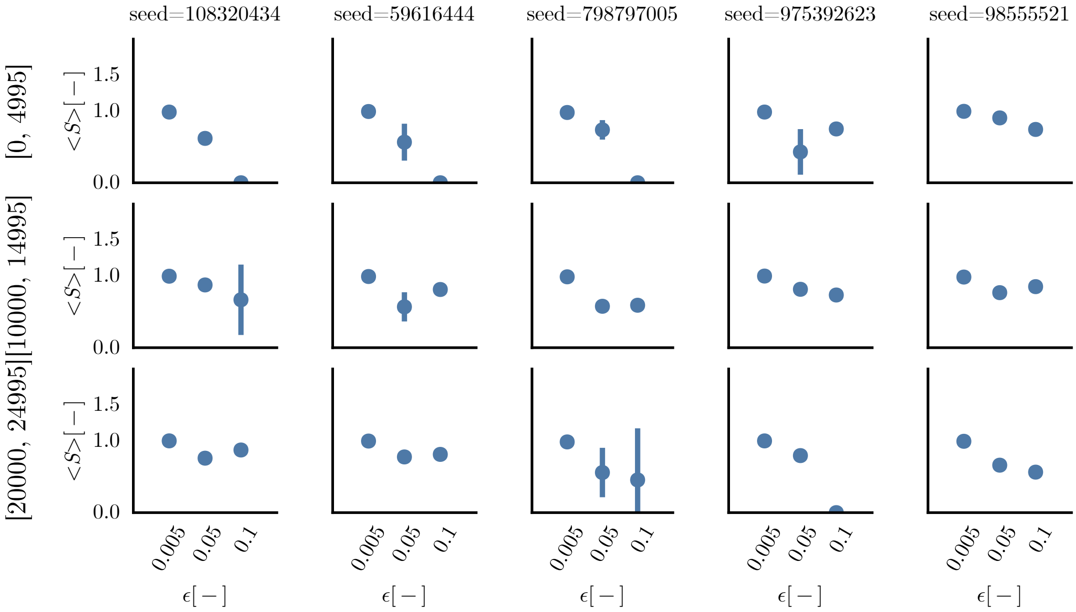


Figure 9.28: Mean and extent (errorbars) of the silhouette measure S for different simulations (columns), intervals (rows), and radii ϵ used in the DBSCAN clustering algorithms. Where no errorbars are visible, the extend is smaller than the marks. The lowest radius appears to overfit, while the largest seems to underfit substantially. $\epsilon = 0.05$ appears to offer the best compromise, although depending on the system, there are exceptions, especially for the first interval.

9.2 Baseline 2 – Multi-resource, single-cell systems without rewriting

9.2.1 Ecological niches and ecological species for the multiple resources

Ecospecies for multiple resources An easy way to increase the complexity of the niche space is to allow for multiple resources. Thus, N in Equation 9.1.1 becomes a matrix instead of a vector, allowing for a second dimension for niche differentiation. Consequently, the production of innovations can be investigated in another way, e.g., when initially organisms do not make use of all available resources such that innovations can occur that tab into a potentially large resource pool not previously accessed.

Multiple resources complicate the metabolism significantly because the firing condition of a transition must be fulfilled for all resources simultaneously. Consequently, where in the single-resource case evolutionary innovations were more or less readily attainable, we expect more complicated evolutionary trajectories in the multi-resource case, which are expected to take much longer runtimes of the simulation than can be managed on available hardware to produce the high-impact innovations found in subsection 9.1.2. The parameters for the simulations discussed in this section are given in Table 9.6, with all others being the same as before.

We indeed find substantially lower numbers of niches in Figure 9.30 than in Figure 9.10 for a single resource, and diversity stays lower for the duration of the simulation. Otherwise, the phenomenology is similar to the one discussed for the single resource case. The occurrence of an ecospecies which becomes dominant, e.g., the blue curve for seed

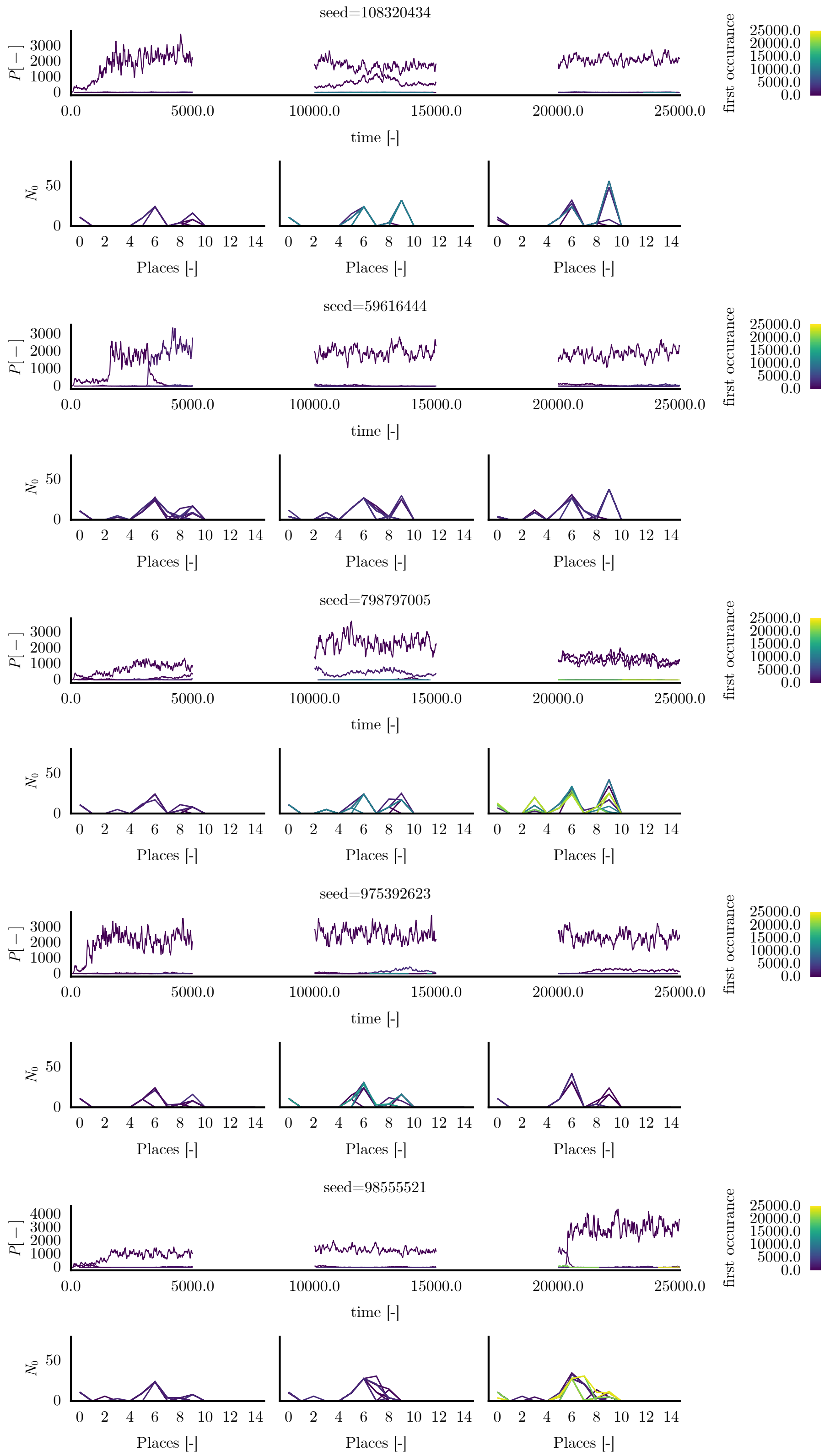


Figure 9.29: Evolution of eco-genera created by applying the DBSCAN clustering algorithm with $\epsilon = 0.05$. We find that most adaptive radiations reduce to intra-genera radiations (as is the case in nature) and that late surviving genera still can give rise to adaptive radiations in daughter genera (lower right.). Coexistence of multiple eco-genera is still frequent, but it is almost always asymmetric with one being dominant and the other suppressed. The dominant genera arise always quickly and stay for the duration of the simulation, indicating that evolution of habitat utilization evolves more slowly than can be represented with current runtimes, in agreement with earlier results.

S_{lower}	45
S_{upper}	95
r_{cell}	$[0,5]$
$r_{organism}$	$[1,4]$
γ_M	6400
M_{max}	32000

Table 9.6: Parameters for the multi-resource simulation presented here. r_X represents how many resources are initially utilized by organisms or filled by the habitat's resource influx.

798797005 in Figure 9.31 for $t \in [20000, 25000]$, precipitates a corresponding increase in the production of ecospecies. Generally, the system shows a punctuated equilibrium in which a more or less neutral succession of ecospecies is interrupted by rare adaptive innovations which then lead to differentiation of this ecospecies, i.e., there is a background species overturn that is mostly driven by neutral evolution, which every now and then is punctuated by an advantageous mutation which can disrupt the system, lead to the extinction of older populations and give rise to a new phase of ecospecies production (e.g., the rightmost panel for seed 798797005.). This succession process will likely always have random and adaptive components, with the relative proportion of both depending on the populationsize of the niches involved and the relative differences in reproductive efficiency of the corresponding ecospecies. This is corroborated in Figure 9.31, where we find an initial increase in population being followed by a differentiation phase. However, neither does the populationsize reach the same level as it does for Figure 9.15 in the first interval, nor does it produce an adaptive succession for all the simulations. Indeed, for the interval $t \in [0, 5000]$, this only happens for the seeds 98555521 and 59616444, with 975392623 exhibiting a single unambiguous adaptive niche replacement only for $t \in [20000, 25000]$. This corresponds to a situation in which the population adapts to the initially occupied part of the niche space without creating innovations that make other parts accessible. Therefore, the situation for the multi-resources case presented here is similar to seed 98555521 in the single-resources case where the system "got stuck" in much the same way and produced innovations that increase the carrying capacity only much later.

From the analysis presented for the single resource case, it can be derived that the slower evolution observed for multiple resources, despite the higher material and energy supply in the system, is caused by a less evolvable metabolism, as conjectured at the beginning of this section, which is the next focus for analysis.

Niche evolution for multiple resources Comparing Figure 9.33 and Figure 9.34 with Figure 9.15,

In Figure 9.33 and Figure 9.34, we find indeed that the system does not produce significant innovations in resource- or habitat-utilization. While for seeds 59616444 and 798797005 there are small usages of resource $r = 0$ and $r = 5$ in the middle panel (timesteps 10000 to 15000), these do not persist over longer times and we find that niche differentiation for a more or less frozen habitat usage pattern (i.e., usage pattern of places) appears to be overall similar for many resources when compared to a single resource, with the variation across resources being broadly similar, too. This is attributed to the fact that the size of the resources that are available in the system is large, such that, once a workable metabolism is established, the individual resources behave more or less similarly. A scenario with many more resources which each are much smaller than the current ones would shed more light

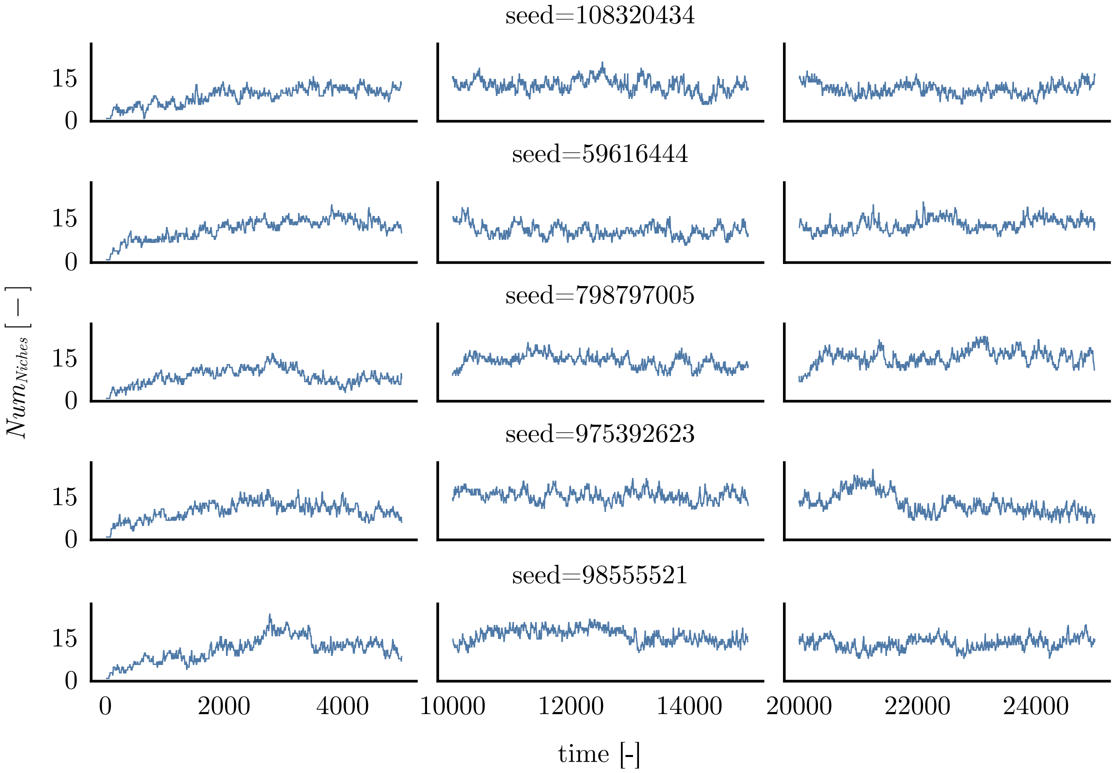


Figure 9.30: Number of different ecospecies over time for a simulation with five different resources. We find significantly lower numbers of different niches than before, and the initial increase in niche number observed in Figure 9.10 levels off earlier. This is consistent with a failure to produce innovations that yield higher metabolic efficiency, such that the simulation almost immediately reaches the differentiation regime.

on this part of the system's phenomenology, because it would produce a stronger fitness impact of the form of the niche. Finally, we find more strongly developed bifurcations in the usage of certain places, mostly place 3 for simulations with seeds 98555521 and 59616444, which in that sharpness was absent from earlier simulations and can be explained by a stronger differentiation of the fewer existing ecospecies.

Aside from a random extinction of the respective ecospecies, there are other reasons that could preclude ecological innovations. As seen before in Figure 9.1.2, metabolic evolution is complex and the simple addition of a new resource input without proper integration into the metabolic processing does not automatically translate into adaptive advantages. In more detail, the usage of new resources must not only include the addition of the resource in the input weight but also in all other weights on a metabolic path from the input to the output interface. Because these cannot occur all at once, and the sequence of mutations that establishes such links is very unlikely when compared with the random death of the individual that carries the initial additional resource mutation, the loss of such a mutation is much more likely than the further addition of a constructive mutation that would add to such a system. Consequently, resource innovation is very unlikely and probably needs much larger populationsizes and longer runtimes. Higher mutationrates would help with generating the necessary variation, but always also produce more deleterious mutations.

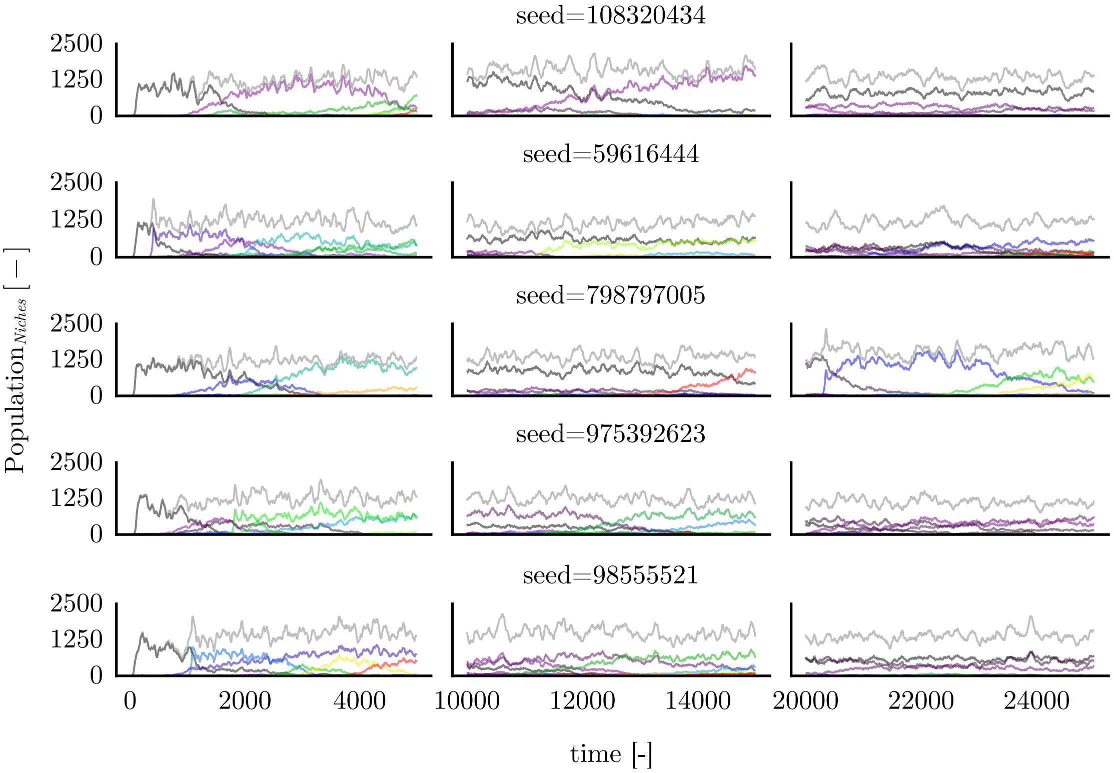


Figure 9.31: Ecospecies populations over time for the three different intervals investigated. Total population in grey, populations of individual niches in colors. Colors only separate species visually and are reset for every interval. We find largely the same phenomenology as in Figure 9.11, although the rapid attainment of high populationsizes and its subsequent stasis indicates that for the multi-resource simulations shown here, the system remains within the initial subspace of fitness and habitat-/resource utilization and does not produce carrying-capacity modifying innovations as for the single-resource case, reaching the differentiation phase rather quickly.

The main problem for the IPN-based system in the simulations presented here is that the usage of a previously unused resource must emerge *de novo* completely. This is in contrast to many natural innovations in resource usage which relink existing metabolic or developmental pathways or coopt other existing features, e.g., via gene duplication or non-coding sequences. This would correspond to a situation in which parts of the metabolism layer for resources 0 and 5 would exist, or parts from other layers could act on them, while only the access to these resources does not exist. In the simulations presented here, however, neither does the former exist nor the latter. Therefore, enough variation on these layers must accumulate in the population for evolution to act upon, and this will most likely happen neutrally at first because a mutation that immediately produces a metabolic benefit for a never-before encountered resource while not disrupting existing pathways is highly unlikely, in nature as in *Amee*. In other words, the creation of completely new (heritable) information is a very hard problem, arguably in all complex constructive systems, i.e., in learning systems and in biology. *Amee* thus opens a new path to the investigation of how novel functions and major evolutionary innovations are created in biology.

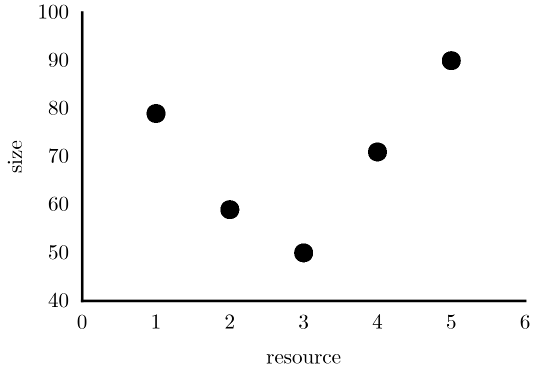


Figure 9.32: Number of symbols in each resource for all simulations presented in this section. Resource sizes are chosen at random at the start of the simulation, and can vary here in $[S_{lower}, S_{upper}] = [45, 95]$. Here, we find that the initially utilized resources (2, 3, 4) are coincidentally the shortest ones.

Evolution of the metabolism networks for multiple resources Given the networked nature of organisms' metabolism, there is a plethora of possible analysis that can be applied, with research in the structure, topology and functional organization of biological metabolic networks being a large and active research topic. Indeed, tools and concepts from this field could be transferred to this model in order to further test its phenomenology, or tailor it towards simulating the evolution of such networks in a more detailed way. That being said, network analysis is a highly specialized field that would go well beyond the scope of this thesis, and is thus forgone here. Rather, we limit ourselves to a cursory glance, deriving conclusions from the visible network structure in a small sample of individuals.

We only consider the two most interesting RNG[5] seeds here, with the others being shown in section 12.1. In Figure 9.36, 9.37, we find a generally similar phenomenology to the single resource case, where we discovered the "survival of the flattest" mechanism, i.e., a general tendency to increase short metabolic paths between input and output interfaces where these arise, or at least maintain them. However, it seems that while the "flat" metabolism does arise multiple times, e.g., via the P3 $\mapsto$ T0 $\mapsto$ P6/P0 pathway, other pathways often play a larger role or are at least as important. The simulations for seeds 59616444, 98555521 and 975392623 behave similarly to Figure 9.36 and can be found in section 12.1.

Investigating the shortest pathlengths analogously to Figure 9.18 reveals that in fact, more complex paths are found mostly for the third resource, while for the second and fourth, only seed 798797005 produces ecospecies that have a higher mean shortest pathlength when compared to the single resource case. For other simulations, there is either a trend towards shorter paths over time and thus "survival of the flattest" (primarily for the second resource), while for the fourth resource, the original level which has mostly short paths to begin with, is largely maintained or is fluctuating more strongly due to a smaller number of ecospecies. We also find that the third resource, for which the mean weighted shortest paths are the longest, is the resource with the smallest number of provided building blocks (see Figure 9.32), lending support to the case made above that more complex metabolisms should develop for smaller resources, i.e., the more complex the underlying resource system, the more complex metabolisms should evolve.

Returning to topology, a differentiation of the individual layers is apparent, where effective metabolic pathways for individual resources are optimized over time (top to bottom in the aforementioned figures). Considering the resources together, we would expect that evolution optimizes the metabolic networks in such a way that the most

---

[5] Random-Number-Generator

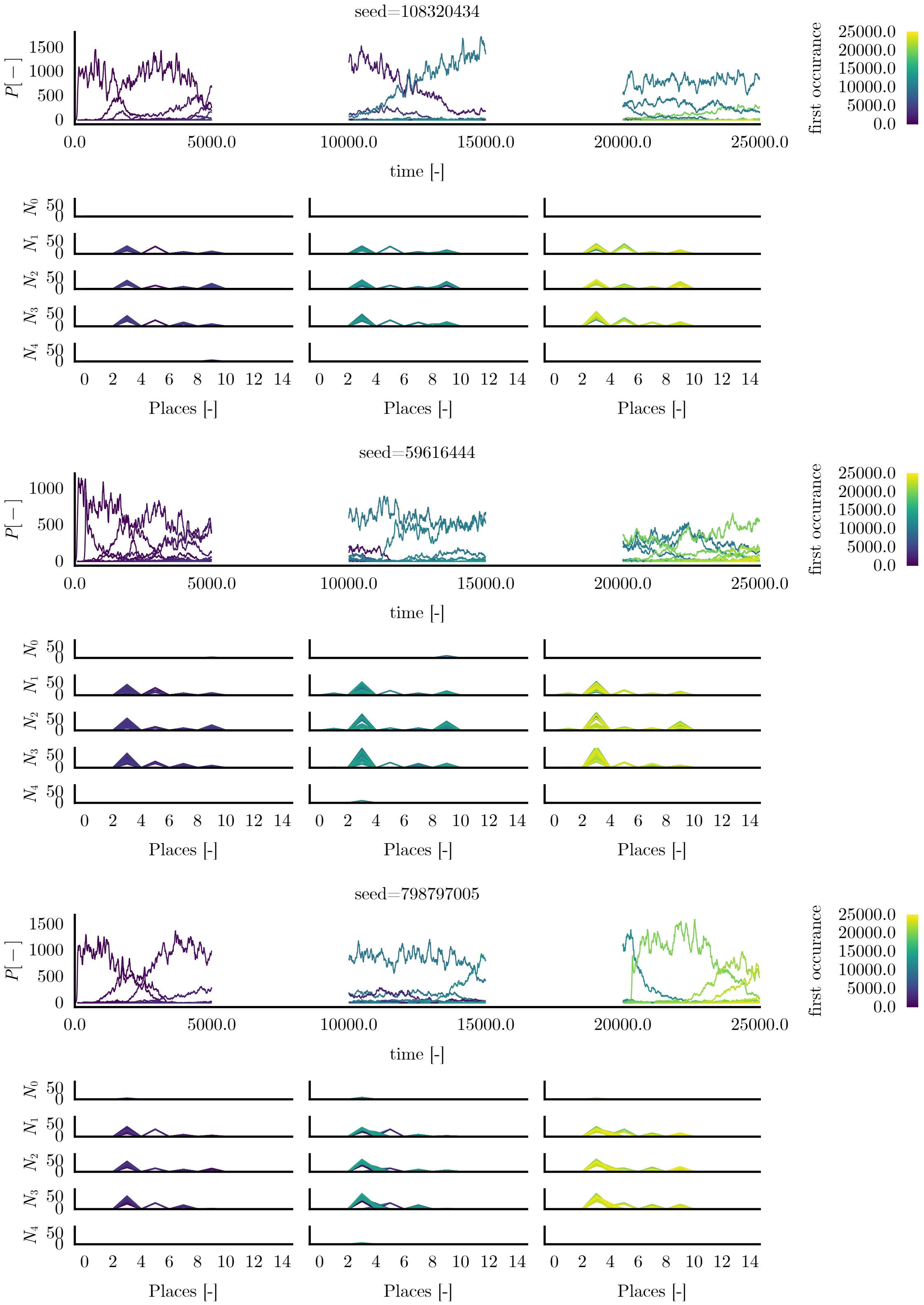


Figure 9.33: Ecospecies evolution for simulation with five different resources, shown from top to bottom in the lower panel analogous to Figure 9.15. Total populations not shown for clarity. We find a lack of innovation in resource utilization and generally limited innovation, consistent with the more intricate multi-resource system metabolism that poses more requirements for variations to not be deleterious. The two missing simulations are shown in Figure 9.34

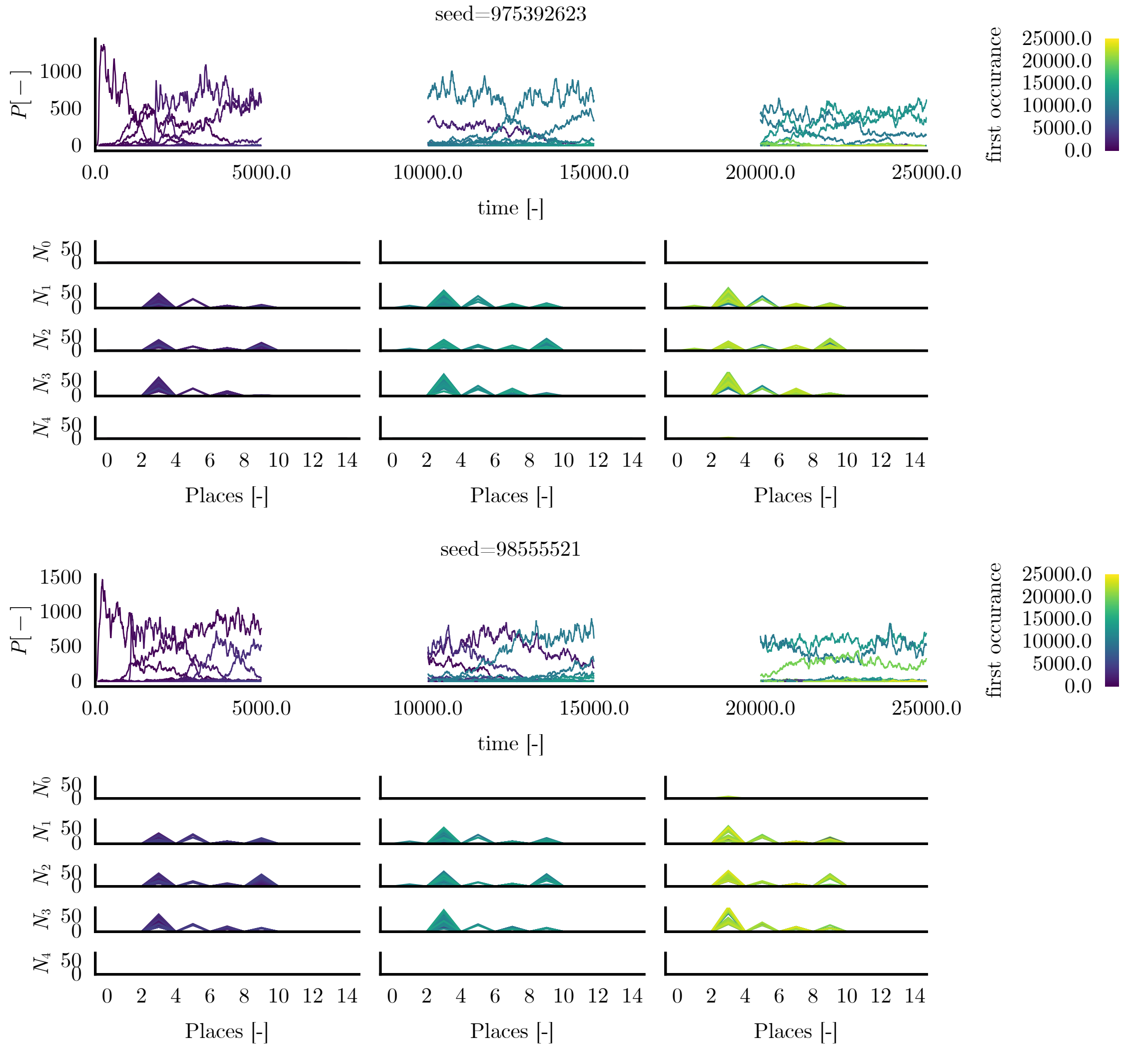


Figure 9.34: Second part of Figure 9.34. We find strong niche differentiation, and evolution of habitat utilization that is otherwise consistent with the findings for a single resource. What is lacking here, most likely due to the constraints a multi-resource metabolism system poses, is innovation in habitat- and resource utilization that gives rise to new bouts of adaptive radiation.

“useful” resources are produced on the output interface (green in Figure 9.37), i.e., the input resource configuration should be transformed in such a way that the configuration that appears on the output interface at the end yields building blocks (here, place labels) in numbers that allow for ever more efficient reproduction. Moreover, Figure 9.37 shows that innovation in the niche alone, i.e., the usage of resource 1 for t=12500, is not enough, because the connection P3 ↦ T1 is not integrated into the rest of the network at this layer and is therefore just a neutral element that, evolutionarily, hitchhikes along on adaptive elements and is lost eventually. This phenomenon has been discussed before and is merely illustrated further here.

Taken together, we find that the introduction of multiple resources results in additional complications for evolution that make it much harder for the population to evolve towards

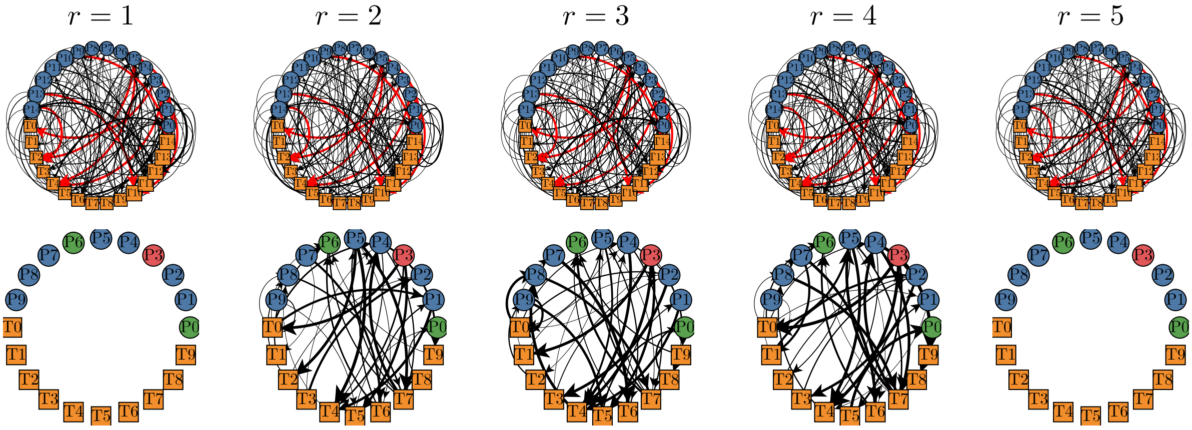


Figure 9.35: Initial metabolism network for organisms (lower row) and habitat resource networks (upper row) for all five resources r_i. We find some variation in weights, although the topology of the networks is similar for different resources. Habitat networks show some inhibitor arcs (red), which have been eliminated by the GA creating the initial conditions.

different regions of the niche space, which was observed before for a single resource. However, evolution of the metabolism networks themselves, independent of the niches, is much richer and provides a lot of opportunities for further research. A caveat in this regard is the implementation of the function that computes if a transition is enabled or not, which currently evaluates on a per-resource basis, thus forcing metabolic evolution to proceed similarly across all layers. Changing this to a condition that refers to the set of involved building blocks would allow for easier resource transformation across transitions and consequently for a more active metabolic evolution because of an increased evolvability of the metabolism networks.

In section 12.2 in the appendix, an example of the simulations shown here but with increased mutationrates is added. It is found there that an increase in mutationrates indeed can precipitate evolutionary innovations that lead to higher system carrying capacity and the invasion of other regions in the niche space of the system even for the multiple-resources case. This, however, has only been observed for a single simulation.

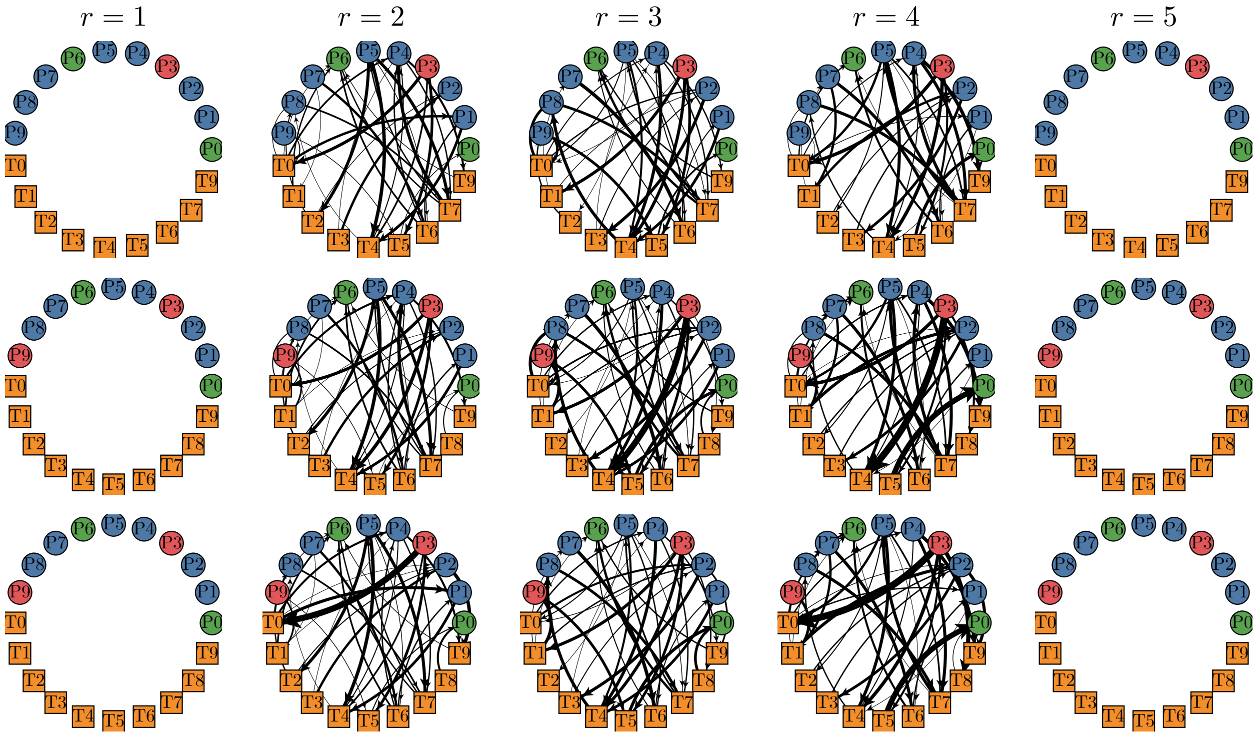


Figure 9.36: Metabolism networks for the simulation with seed 108320434 for randomly chosen individuals of the largest niche in timesteps (from top to bottom) 1250, 12500, 22500, and for resources 2, 3, 4 (from left to right) with 1,5 being all zero. Input interfaces are colored red, output interfaces green, as before. We find, analogous to the initial condition, that topological differences are minor between resources, while weight differences are significant. Their evolution is largely consistent with the findings for a single resource.

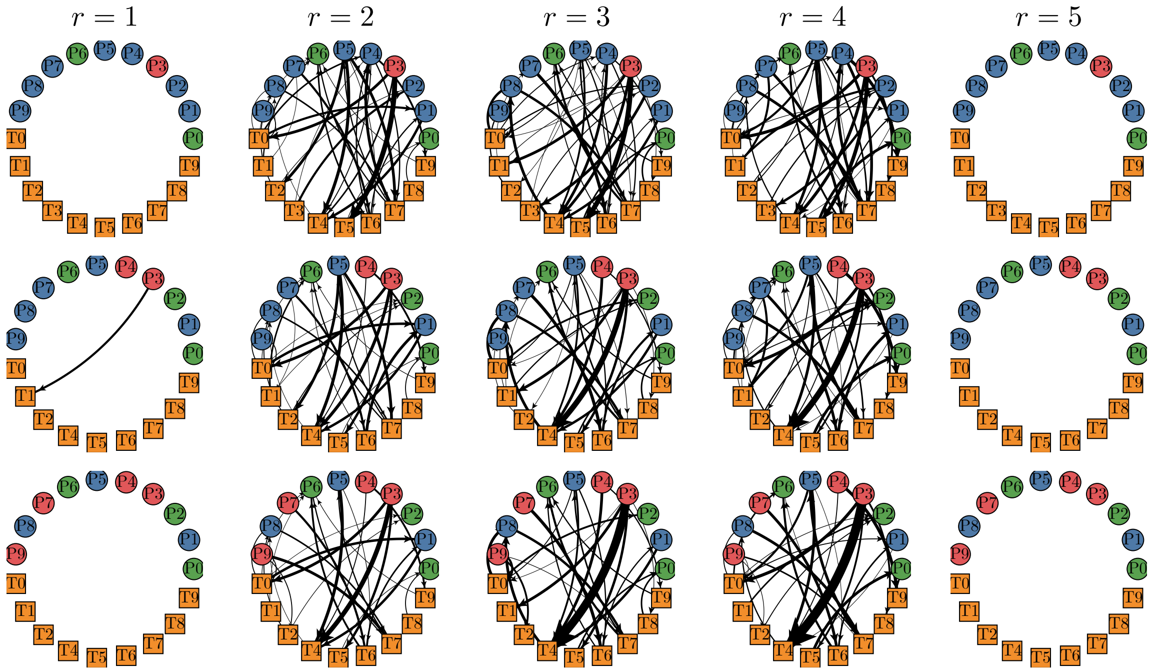


Figure 9.37: Metabolism networks for seed 798797005 for randomly chosen individuals of the largest niche in timesteps (from top to bottom) 1250, 12500, 22500. Input interfaces are colored red, output interfaces green, as before. For the middle row, we see that the niche innovation discussed before does lack connection to the rest of the network and thus is just a neutral element.

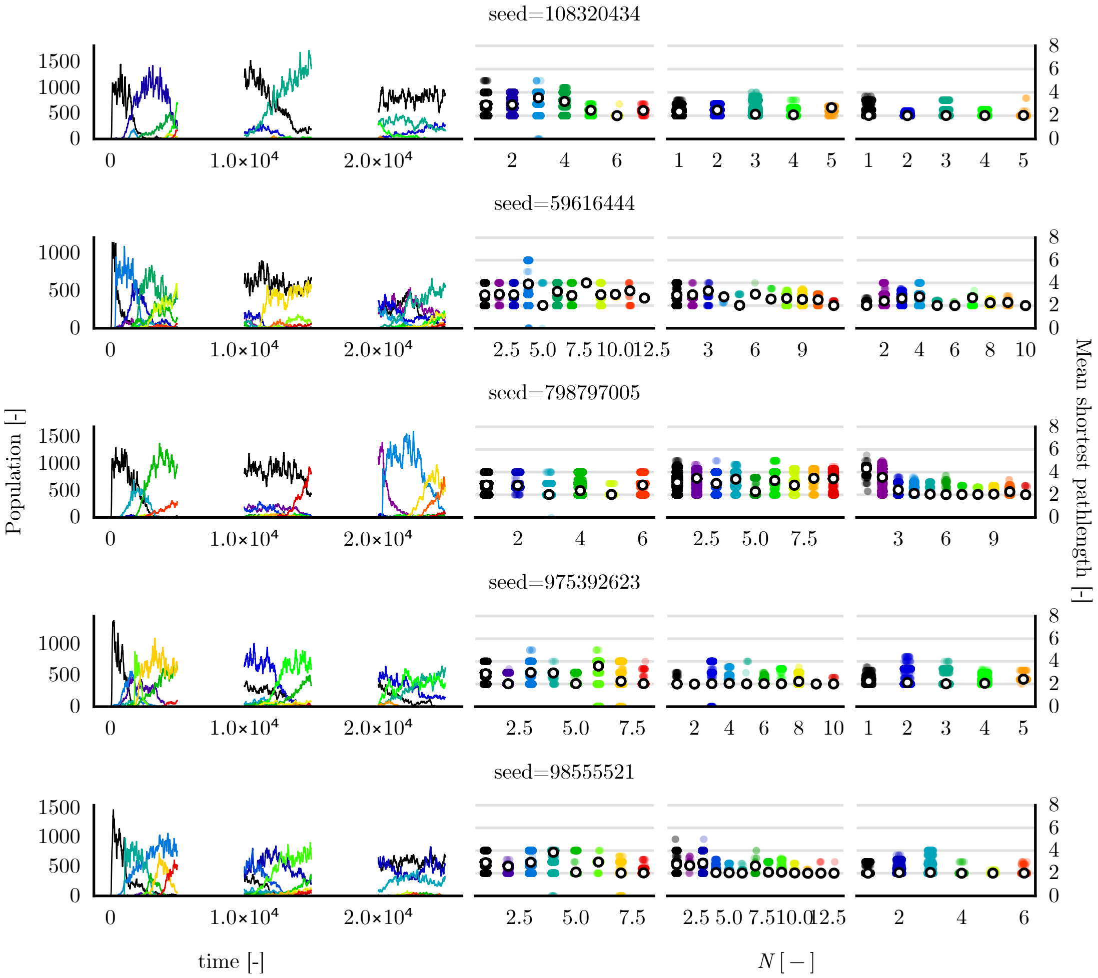


Figure 9.38: Ecospecies populations for time intervals $[0, 5000]$, $[10000, 15000]$, $[20000, 25000]$, for the second resource, analogously to Figure 9.18. We find that for this resource, the situation is similar to the single resource case, although for some simulations we find a downward trend in the mean shortest paths over time, consistent with "survival of the flattest". For seed 798797005, the shortest paths stay comparatively long for longer times but show an eventual downward trend.

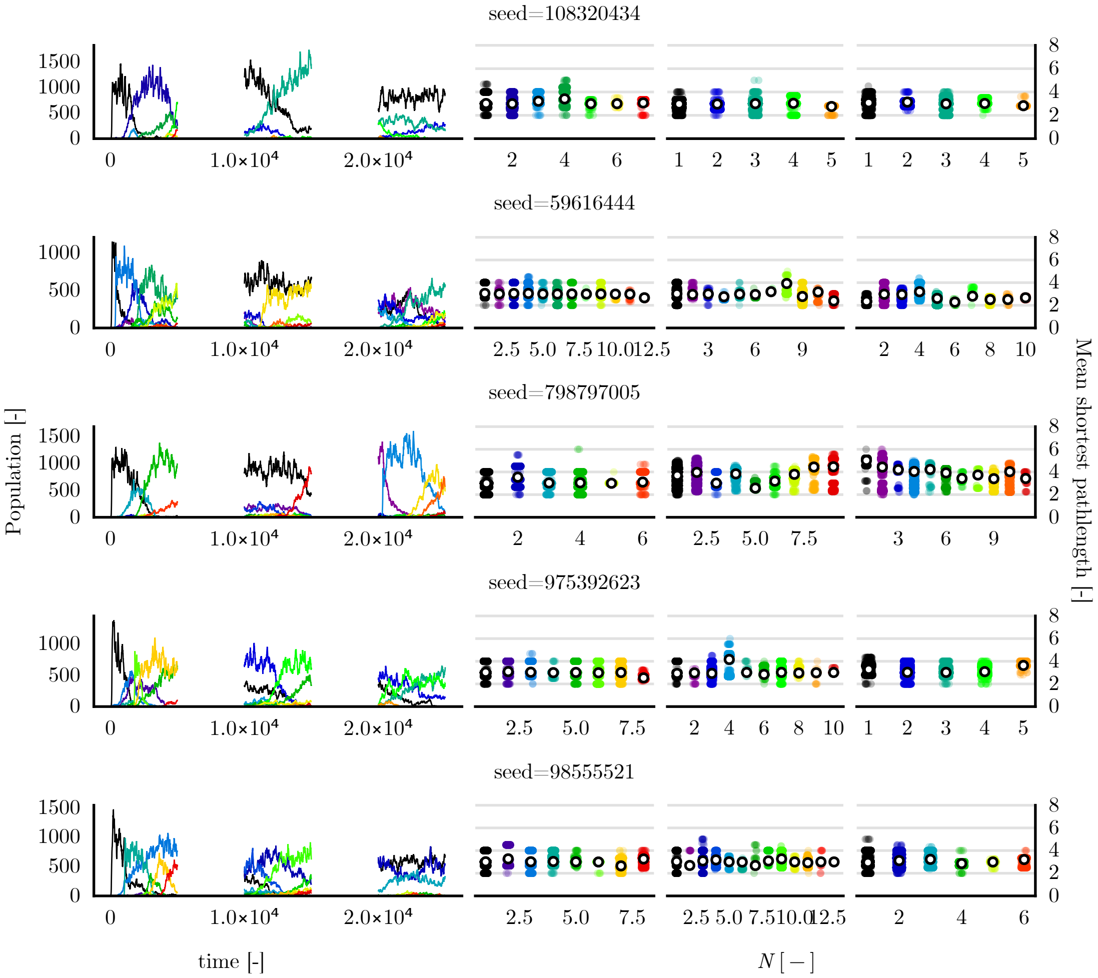


Figure 9.39: Ecospecies populations for time intervals $[0, 5000]$, $[10000, 15000]$, $[20000, 25000]$, for the third resource, analogously to Figure 9.38. Resource 3 is the shortest, and the mean shortest paths are the longest for all simulations when compared to Figure 9.38 and 9.40, consistent with the expectation that the more complex the underlying resource system, the more complex metabolisms should evolve.

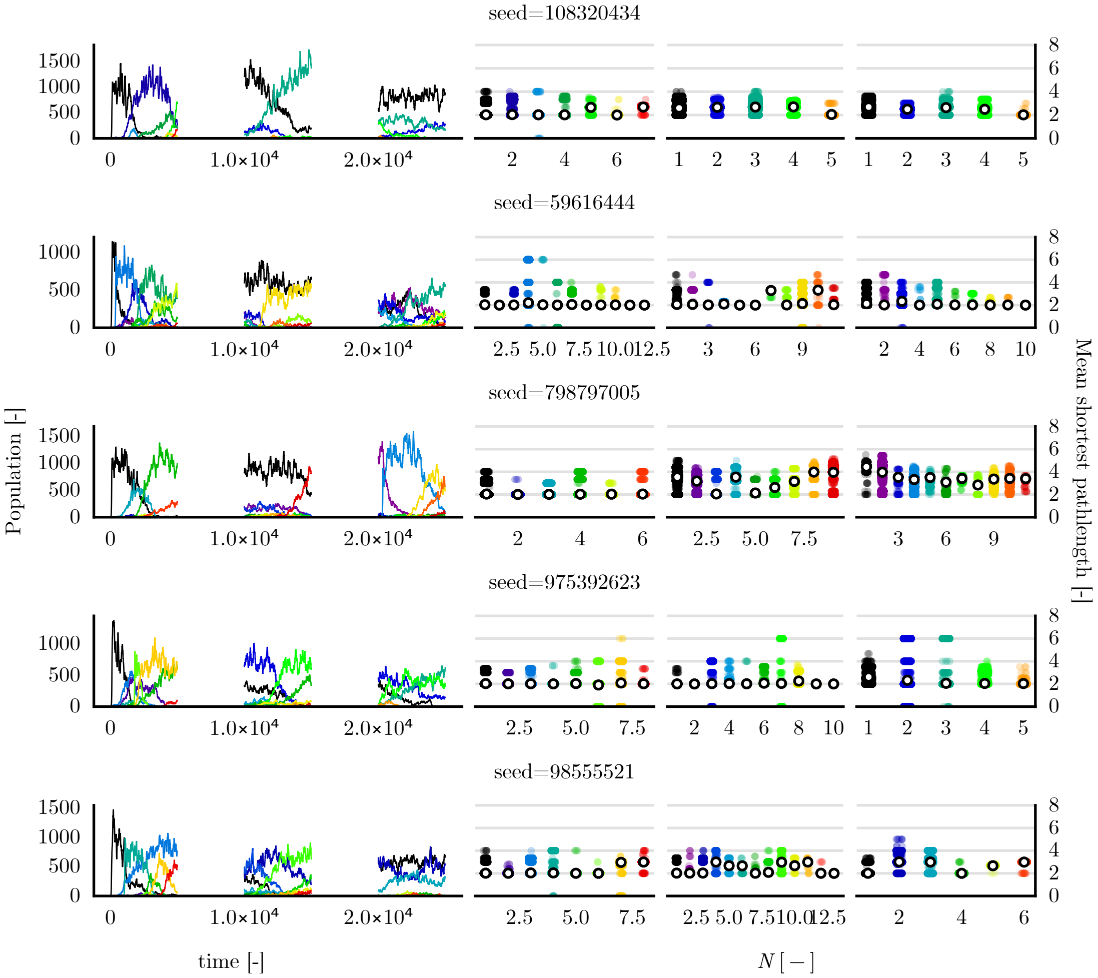


Figure 9.40: Ecospecies populations for time intervals $[0, 5000], [10000, 15000], [20000, 25000]$, for the fourth resource, analogously to Figure 9.38. The same phenomena as in Figure 9.38 and Figure 9.18 are observed.

10. Phenomenology II: Development, plasticity and niche construction

We consider rewriting processes in the single-resource case presented in section 9.1. Again, this aims at an illustration of the phenomenological breadth the model can produce instead of focusing on individual scenarios in full detail.

10.1 Single-resource, single-cell system with rewriting

We extend the single-resource system with all possible combinations of development, plasticity, and niche construction. This includes the simulations presented in section 9.1, i.e., where all these processes are off, for comparison. We compute five simulations for each combination of rewriting processes, initialized with the same RNG-seeds as before. Two different cases are presented, one with a given set of parameters for the rewriting processes, i.e., development, plasticity and niche construction, and one where the size and number of the available rewriting rules have been reduced to investigate the influence of rewriting intensity. The initial condition of each simulation contains at least one active rewriting rule for each process, such that any loss of a rewriting process during the simulation is secondary.

variable	basic	reduced
R	$[0, 30]$	$[0, 15]$
L_R	$[0, 3]$	$[0, 2]$
I_{constr}	$[10, 20]$	$[5, 10]$
I_{plast}	$[20, 30]$	$[0, 5]$
I_{dev}	$[0, 10]$	$[10, 15]$
$N_{0,constr,dev,plast}$	3	1
$e_{constr,dev,plast}$	0.001	0.001

Table 10.1: Rewriting related parameters for the basic scenarios presented in section 10.1. All other parameters are the same as for section 9.1. We investigate one case ("basic") with a given set of rewriting parameters, and a second case ("reduced") where the number and size of rewriting rules have been reduced.

Because the number of simulations in this chapter is much greater due to the investigated parameter combinations (forty instead of five) we focus on the time-development of the statistics of relevant quantities instead of high-resolution data in specific intervals. The observed phenomena are presented and possible explanations are discussed. The time-development of populationsizes for each combination of rewriting parameters are shown in Figure 10.1 and 10.2. Therein, the panels for $[c = \texttt{false}, (d = \texttt{false}, p = \texttt{false})]$ (upper left) represent the baseline case presented in section 9.1, i.e., where all rewriting processes are off.

10.1.1 Niche construction

Niche construction consists of rewriting the habitat's resource network by organisms. Across all cases investigated here, we find niche construction to be mostly deleterious to the survival or growth of the population. In Figure 10.1, niche construction leads to the extinction (or to the shrinking of the population to only a few individuals) of every simulated population irrespective of other influences, although the absence of plasticity seems to allow for longer survival for some cases. For Figure 10.2, where the possible rewriting rules are reduced in number and size, we find cases where the opposite is true, however. For seed 798797005 and 97592623 and niche construction only (lower left panel in Figure 10.2), niche construction allows populationsizes to be attained which are more than twice the size of the populationsizes attained when all rewriting is off (upper left in the aforementioned figure), at least temporarily. A similar phenomenon can be found when niche construction is combined with development *and* plasticity for seed 59616444, which however eventually dies off. These population increases are fast and thus can be assumed to be caused by some adaptive effect, either an adaptation in the metabolism occurring via natural selection that reacts to a changed habitat resource network, an adaptive niche constructive change to the latter or a combination of both. This can be likened to an evolutionary innovation, although one that can be enabled by an active change to the habitat and not via evolution by natural selection. The rerouting of resources or introduction of new possibilities for exploitation, i.e., new places in the habitat that organisms can connect to could have a similar effect. Note that if this modifies the input interface sizes of the habitat network, these newly introduced places will be the target of the resource influx and thus will be attractive targets to adapt to. We therefore conclude that niche construction (or –change), as implemented here indeed modifies the evolutionary niches of populations as was discussed in chapter 6 and by K. Laland, Matthews, et al. [2016], although the exact mechanism has not been determined. Moreover, while it apparently can be adaptive in itself or facilitate adaptive change, this effect is rare, and overall, strong changes to the habitat networks seem to be mostly deleterious.

Moreover, extinctions happen rapidly, which can be explained by the conditional nature of Equation 8.3.9 in conjunction with additive rewriting, i.e., a single rewriting in the habitat resource network can result in the firing condition of an important transition becoming disabled, which precludes resource flow to its post-set places. When the affected downstream places are part of an ecological niche of an ecospecies it will lose part of its sustenance, and the population will rapidly starve to death when the resource supply on the respective places is exhausted. If that is the case, however, it implies that diversity is low and the system is dominated by one or a small number of closely related ecospecies, which is consistent with early stages of evolution as seen in chapter 9.

The fact that the high populationsizes for rewriting are only attained when rewriting rules are small hints at an insufficient evolutionary history of the system, i.e., ecological modifications are not integrated into the ontogeny of the organisms adaptively, and this can only be accomplished during the simulation when their impact is individually small.

Considering the impact of rewriting on the habitat network further, we find in Figure 10.3, 10.4, 10.5, 10.6 that, where the number of non-zero connections in the input– and output matrices of the single habitat network in the system is shown, that niche

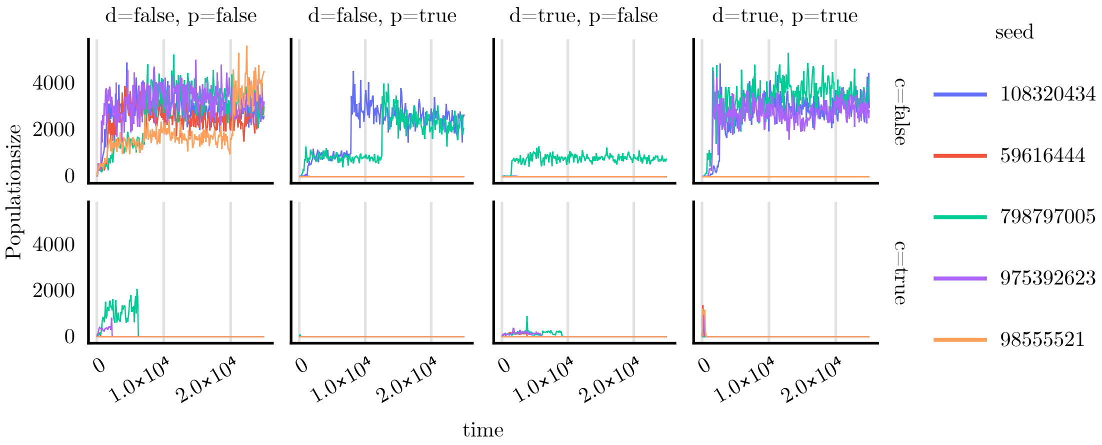


Figure 10.1: Populationsize over time for the basic rewriting parameters given in Table 10.1. c, d, p indicate whether niche construction, development and plasticity are on or off, i.e., the top left panel corresponds to the baseline case presented in section 9.1. Niche construction always leads to extinction, whereas plasticity and development only allow for populations reaching the same densities as for the non-rewriting case when they are active concurrently, and reduce populationsize otherwise or delay the growth of the population. Moreover, no case for which rewriting is active is free of extinctions or at least very low populationsizes.

construction leads to an increase in connection density in the habitat's resource network, which cannot be reduced again autonomously because modifications do not decay by themselves. This yields a mechanistic hypothesis for the cause of extinctions: The lack of evolutionary integration of the niche construction rules likely disables too many transitions in the baseline case, such that the habitat network becomes effectively inactive. In the case of reduced rewriting rules, their impact is lessened, transitions are not affected as much and organisms either adapt via rewriting or via metabolic innovation. We find in Figure 10.5 and 10.6 for the construction-only case and seed 97539263 that while the input density reaches a constant value around which it fluctuates, the output density is still increasing when the simulation ends. The significance of this finding is unclear at this stage, however.

On the other hand, it would also be possible for the system to survive if organisms would exploit the habitat's input places where the external resource influx is placed on. However, as seen in chapter 9, the evolution of new place utilization patterns by organisms is difficult and rare, i.e., metabolic evolution is unlikely to keep up with comparatively rapid changes to the habitat network. If this effect plays a role in the high-populationsize simulations for the reduced-rewriting case, which in principle could still be possible due to the path dependency of evolution, has not been investigated, though it seems unlikely.

Plasticity and development in principle could increase the evolvability of the metabolisms, though this requires, again, the evolutionary integration of the different processes. Therefore, ontogenetic integration of different rewriting processes becomes a question of timescale synchronization between effects on the habitat network and the evolvability of the metabolism networks.

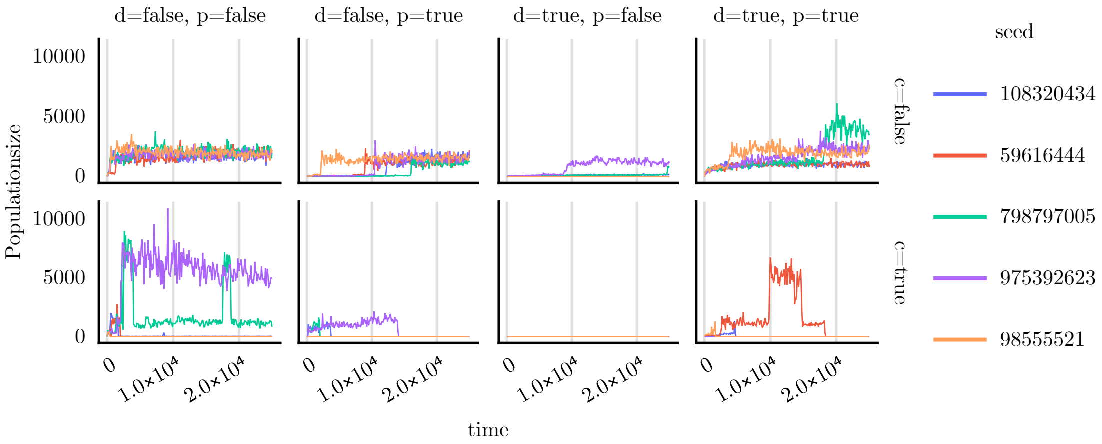


Figure 10.2: Overview of populationsizes over time analogously to Figure 10.1 but with reduced number and size of rewriting rules as described in Table 10.1. In contrast to Figure 10.1, populations survive under niche construction for longer, but only rarely for the entirety of a simulation. Niche construction alone allows for some simulations to reach populationsizes over twice the size of the baseline case (compare upper left with lower left panel). For seed 798797005 this is present only temporarily, whereas seed 975392623 exhibits a large stable population for the duration of the simulation. A comparable phenomenon can be seen for seed 59616444 in conjunction with plasticity and development (lower right), which leads to extinction eventually. Survival rates without niche construction appear to be higher than for the baseline scenario, especially for the plasticity-only case. This hints at rewriting allowing for potentially advantageous modifications which however appear to be difficult to attain and maintain. Given that niche constructions do not decay, the question remains whether construction is maintained afterward or if it dies out eventually.

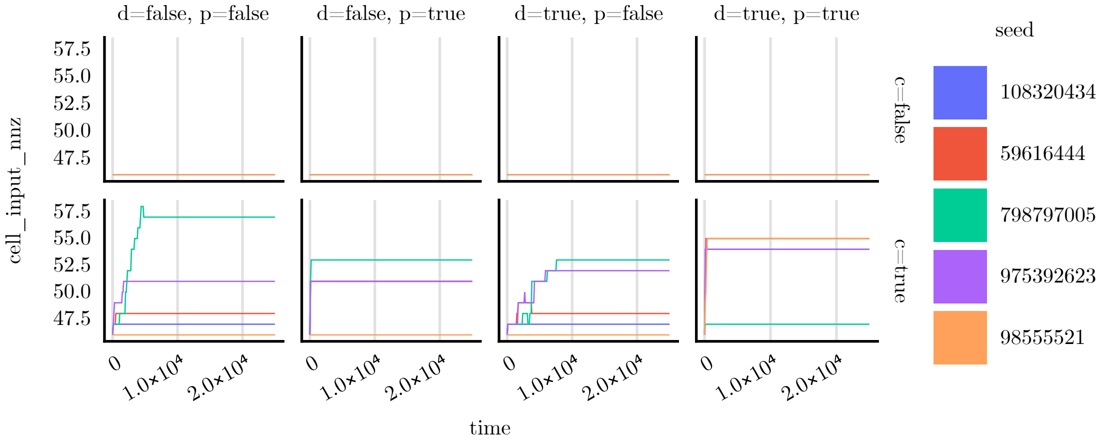


Figure 10.3: Number of nonzero elements in the input matrix of the sole habitat resource network for the basic simulations with rewriting over time. Niche construction leads to a quick increase in nonzero elements, which likely affects the net's ability to process resources and thus inhibits resource acquisition. Flat lines correspond to extinction.

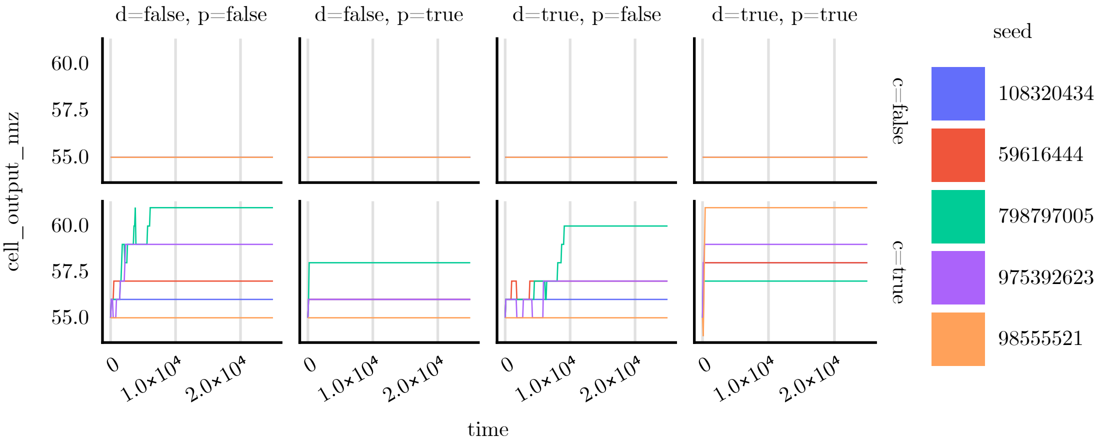


Figure 10.4: Number of nonzero elements in the output matrix of the sole habitat resource network for the basic simulations with over time rewriting. We find the same phenomenology as in Figure 10.3. Flat lines correspond to extinction.

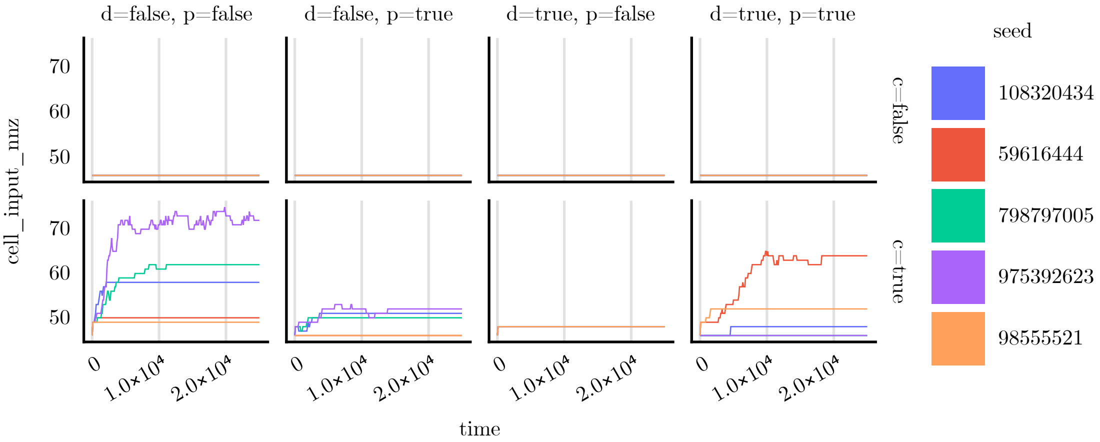


Figure 10.5: Number of nonzero elements in the input matrix of the sole habitat resource network with reduced rewriting rules over time. Connection density increases in the beginning before reaching a constant value, with flat lines indicating that no construction happens due to extinction or inactivity.

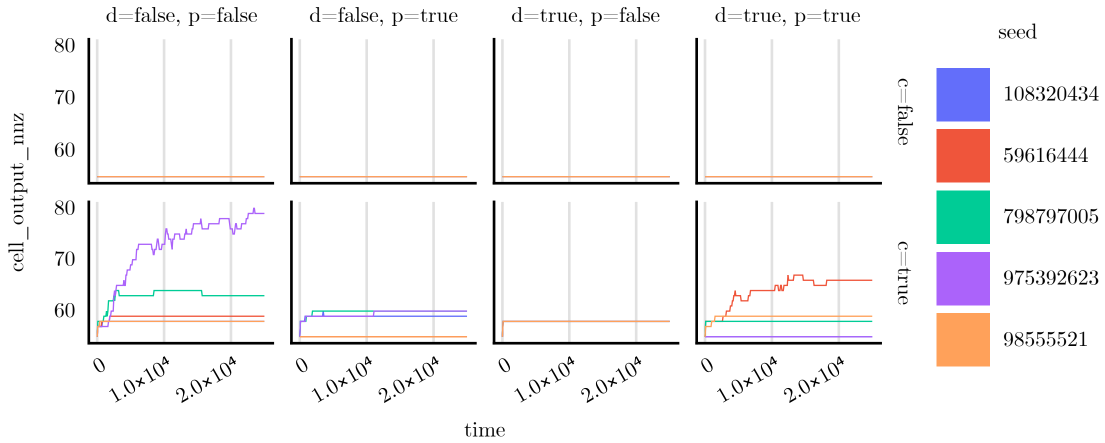


Figure 10.6: Number of nonzero elements in the output matrix of the sole habitat resource network with reduced rewriting rules over time. For seed 975392623, lower left, we find a continuing increase in connection density over time, in contrast to the input density in Figure 10.5. The significance of this finding has not been investigated further, however.

10.1.2 Development and plasticity

We now focus no the cases without niche construction, i.e., on the upper row in Figure 10.1 and Figure 10.2. Development and plasticity are qualitatively different from niche construction because they act upon individuals, i.e., their effect is fully monopolized by organisms. Consequently, we can attribute the evolutionary maintenance or extinction of these two processes directly to natural selection on individual populations in which they are expressed. Overall, we find that development and plasticity also lead to frequent extinctions or at least greatly reduced populationsize, with only some simulations being able to reach populationsizes comparable to the baseline case or exceed them. The largest populationsizes are thereby attained when development and plasticity are active concurrently, with populationsize breakthroughs, previously associated with evolutionary innovation, being delayed for many simulations in the development-only case when compared to the cases without rewriting.

Rewriting is always costly. Aside from the hypothesis that plasticity– and developmental rewriting rules interact beneficially, there is also the possibility that because of the additive cost of both, selection against rewriting occurs more strongly for the surviving populations. Consequently, rewriting will be lost more quickly, which results eventually in a scenario corresponding to the baseline case where all rewriting is off, but with altered initial conditions. The quick increase in populationsize for seed 798797005 in Figure 10.2, upper right can be likened to an evolutionary innovation as seen in chapter 9, but the question remains whether this comes about through metabolic adaptation or plasticity or development. A loss of rewriting would be more likely for the reduced-rewriting scenario in Figure 10.2 because the initial organism only has one functioning initial rule per active rewriting process which consequently can be lost more easily than the three required rules in Figure 10.1.

Development and plasticity can also be maladaptive because the initial algorithm does not force the rules to be advantageous. A maladaptive rewriting rule applied to the metabolism network incurs a twofold cost because it costs material to apply and damages the metabolism such that resource acquisition becomes less efficient or effective. Consequently, maladaptive rewriting plays a similar role to the accumulation of deleterious mutations in the mutational meltdown process, which can lead to extinction as well. Indeed, where there is variation in rewriting, we would assume that maladaptive rewriting will be lost quickly. While maladaptive plasticity does occur in nature, for the simulations presented here we attribute its existence more to the lacking integration of the rewriting into the lifecycle of the organisms, i.e., natural selection had not enough time to shape them into integrated, adaptive forms.

All these phenomena possibly can occur in the simulation results presented here, spanning a wide range of natural phenomena and allowing for deeper investigations in the future. When interpreting these results, however, it must be noted that the modifications made to the habitat network do not decay over time, which would normally be the case in nature, at least on a long enough timescale, and that the system is not spatial, which can arguably alleviate extinctions. The investigation of these effects is left for future work, however.

10.1.3 The evolution and maintenance of rewriting processes

In order to investigate how rewriting rules evolve, we focus on how much energy organisms invest into rewriting. As shown in Table 10.1, rewriting a single symbol in any rewriting rule costs 0.2 energy units, such that the statistics over the energy investment into rewriting rules over time tells us something about how rewriting processes are influenced by natural selection and how they, in turn, influence it. Moreover, trends in the overall number of rewriting rules can hint at the adaptiveness of rewriting or its neutrality, and at the interaction between different rewriting processes.

Plasticity and development for the baseline case We first focus on the baseline case (compare Table 10.1) and on the evolution of the mean overall number of rewriting rules in Figure 10.7. We do not consider the evolution of niche construction for this scenario, because the case with reduced rewriting rules shows the same phenomena, but additionally includes surviving and thriving populations for pure niche construction, such that an analysis of the reduced scenario is deemed sufficient.

For the baseline scenario, all simulations start out with nine initial rewriting rules. For the case where all rewriting processes are off, the rules still exist, i.e., the evolution of their number is neutral with respect to the effect of rewriting on the evolutionary trajectories. When development and plasticity are active concurrently, the number of rules increases in a stepwise fashion without much of a reverse trend being apparent. Since this stepwise increase is absent in the baseline case without rewriting, we interpret this as evidence that development and plasticity in conjunction are adaptive, such that metabolic rewriting becomes an important part of the system's evolutionary trajectory. This is less clear for the development-only case, where simulations 798797005 and 108320434 show somewhat contradictory behavior, indicating distinct evolutionary trajectories.

For deeper insight, we turn to the energy investment of organisms into each type of rewriting process, which illustrates how individual rewriting processes evolve and trade off against each other. In Figure 10.8 and Figure 10.9, a complex interplay between plasticity and development is apparent: When both are active at the same time, the simulation with seed 795797005 shows intricate behavior where plasticity– and developmental investment increase together around timestep 12000 after the latter was predominantly lost, only for development to become reduced again. While there appears to be a downward trend

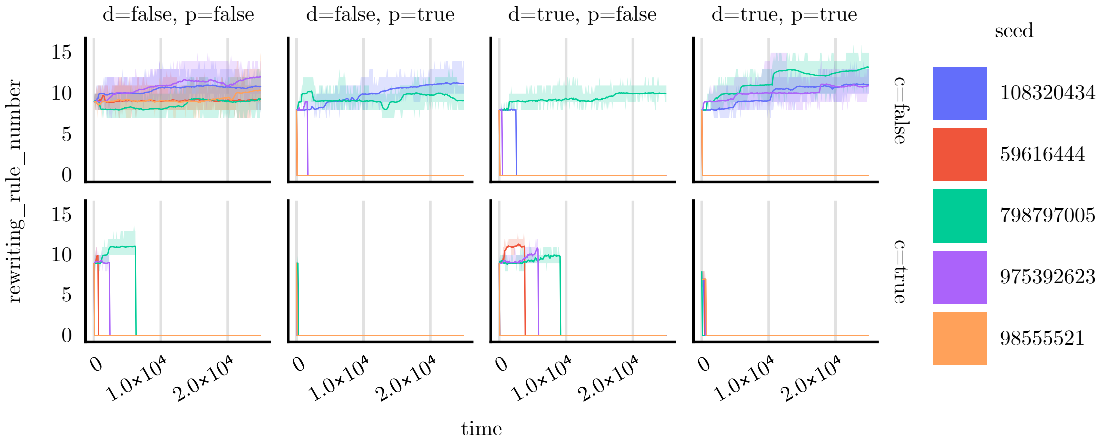


Figure 10.7: Statistics of the number of rewriting rules (accumulated over all rewriting processes). Mean as solid lines and (min, max) range as shaded area. For the top left panel, there is no rewriting but the rules are still part of the genome, i.e., this case can serve as a baseline for comparison. For $d = \texttt{true}, p = \texttt{true}, c = \texttt{false}$, we find a stepwise increase in the number of rules with very little variation and without subsequent loss of rules, which hints at evolutionary advantages being conferred by an increased number in rewriting rules. This trend does not persist for all other process combinations, which hints at evolutionary relevance of active development and plasticity. When niche construction is active, upward trends are apparent for most surviving populations, which raises the question if there is selection for compensatory rewriting either in the habitat IPN or the metabolism.

in the last 5000 timesteps for developmental investment in Figure 10.8, this seems to be weak and longer simulations would be needed to determine the further evolution of mean developmental investment more clearly. Seed 975392623 appears to tend towards losing its rewriting system completely over time, although temporarily a similar behavior to seed 798797005 can be gleamed at $t = 20000$. Mean plasticity investment, however, becomes quickly very low and appears to be mostly lost by timestep 10000. Finally, seed 108320434 shows purely development, with no plasticity investment. Moreover, it appears to tend towards an increase in developmental investment over time. We note the asymmetry between developmental and plastic rewriting rules: The former are only applied once when the organism is first created, while the latter can be applied repeatedly over the lifetime of the organism and incur material– and energy costs each time. Moreover, development is neutral with respect to the material cost, because developmental rules are always considered part of an organism's genome, irrespective of whether development is active or not. Consequently, development does not incur further material costs while plasticity (and niche construction) do. Therefore, if costs are a relevant factor that precipitates selection against rewriting, we would expect this to be more relevant for plasticity. However, no clear pattern emerges for the sample sizes shown here that would allow us to strengthen the case for selection against plasticity.

Plasticity is mostly lost where it is the sole rewriting process available. Development, on the other hand, is often retained by a fraction of the population, hinting at developmental rewriting becoming integrated into the lifecycle of organisms, with possibly multiple different strategies being involved. However, although development does not incur additional material costs when it is active, it is not maintained by the majority of the population, indicated by the mean development investment being close to zero where it alone is active in Figure 10.8.

Loss of plasticity where it alone is active in Figure 10.9 coincides with a strong increase in populationsize for seed 798797005, although for seed 108320434 loss of plasticity appears to be delayed compared to population increase. Such a phenomenon has previously been identified as an evolutionary innovation that allows access to different, and more productive, parts of the niche space, but the differing temporal correlation between population increase and loss of plasticity makes it unlikely that the same mechanism causes both events. A multitude of possible explanations present themselves at this point. Genetic assimilation of adaptive rewriting effects is possible via the crossover mutation. Of primary interest, of course, is the possibility of plasticity-led evolution. For seed 108320434 and for seed 798797005, when only plasticity is active in Figure 10.7, we find a dip in rewriting rule size coinciding with an increase in populationsize, followed or paralleled, respectively, by a loss of rewriting. A plausible mechanism here is a crossover mutation which assimilated a rewriting rule into the genome. In fact, a closer look at the time evolution of seed 108320434 reveals that the increase in populationsize coincides with an initial increase in plasticity investment when plasticity is the only active rewriting process, and only after that is plasticity eventually lost. This is consistent with a scenario in which a more plasticity-reliant variant increases in frequency in the population and gives rise to a daughter variant that either has assimilated a plastic effect or developed other adaptive mutations that enabled it to eventually replace the plastic variant. Further research is required to establish the exact mechanism, however, the possibility of plasticity-led evolution is intriguing.

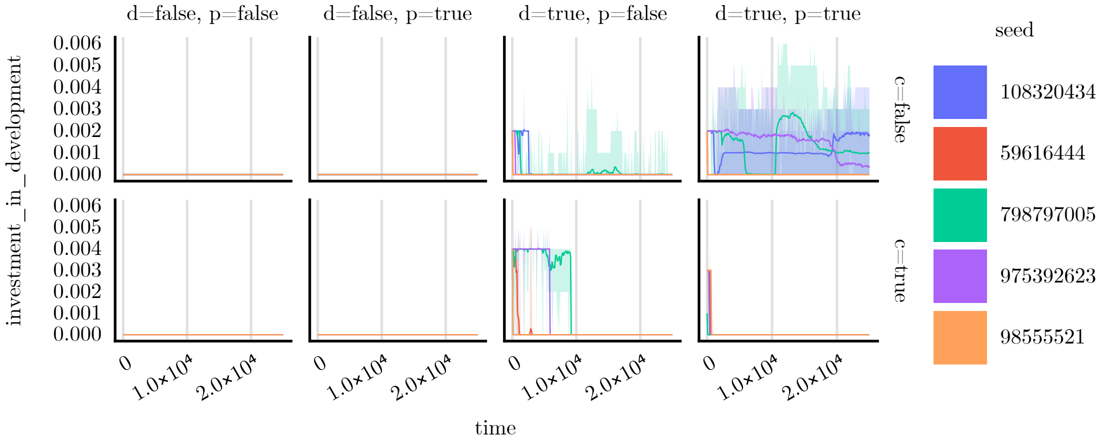


Figure 10.8: Energy investment statistics into development over time for the baseline rewriting scenario. Mean shown as a solid line, with minimum–maximum range shown as shaded areas. Where absent, minimum and maximum coincide. Where d = `false`, development doesn't occur and investment is consequently zero. We find no obvious correlation between rewriting investment and populationsize, although fluctuations in the mean investment hint at internal evolution. There is no clear trend for or against development, with both increases and decreases in developmental investment being present.

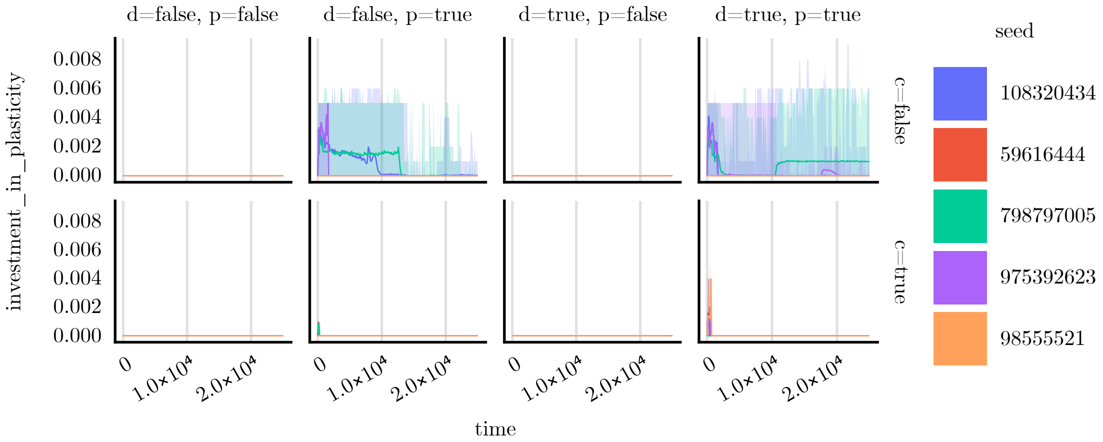


Figure 10.9: Energy investment into plasticity for basic simulations with rewriting. When plasticity alone is active, plastic behavior is predominantly lost, whereas when it occurs together with development, it is kept for seed 798797005 but lost for the others. Therefore, adaptive interactions between the metabolic rewriting systems appear possible, but are comparatively difficult to achieve.

For the seed 798797005, where the clear delay is absent, the evolutionary innovation responsible for the population increase could, however, also be a simple adaptive replacement of an existing lineage with a new one that has lost rewriting. Given the quick drop in plasticity investment, this appears to be the more plausible alternative. This illustrates that *Amee* shows interactions between evolutionary processes as we would expect from a comprehensive model of evolutionary dynamics, and confirms its applicability to modern research questions. The large variability between individual simulations however would require larger sample sizes for deeper investigations.

Because we have no dynamic environmental structure and no spatiality in the baseline scenario, no other strategies like bet-hedging or habitat choice can evolve. Longer simulations could reveal if rewriting strategies can reemerge at a later time after they have been lost, given that the number of rewriting rules shows no general downward trend even when they are mostly inactive. Finally, developmental rewriting appears to prolong survival when niche construction is active and mean developmental investment is comparatively high for this case, again confirming interaction between rewriting processes. Indeed, niche construction can enable or disable metabolic rewriting rules which are dependent on habitat structures to become active. Consequently, the interaction network among rewriting rules would be an interesting future line of investigation.

However, in contrast to natural systems, the processing systems for biological information that allows natural phenotypes to successfully construct the succession of complex phenotypes that makes up the developmental dynamics is, as far as can be seen from the results so far, not well integrated because the simulated evolutionary history is not long enough for it to develop. Indeed, an important future research direction would be to investigate under what conditions evolving rewriting systems as shown here can produce systems we could interpret as effective, fitness-enhancing biological information processors. While mechanistically there is no apparent reason why they should be fundamentally unable to do so, given that a nonlinear input-reaction-output system is in place, evolutionary trajectories producing them are likely far longer with larger populations than can be realized at the moment.

Reflections on the ecospecies concept, realized– and fundamental niches with rewriting Development only occurs once when the organism is born, and therefore it does not dynamically influence the ecological niche of an organism during its lifetime. Development rules can, however, be vital for the successful creation of a suitable ecological niche and thus can contribute to the organism-environment fit characteristic of natural evolution.

The same is true for plasticity in principle, but because plasticity-rewriting can be applied in each timestep, it can change the ecological niche over an organism's lifetime, such that a single organism belongs to multiple ecospecies. This is not unlike the examples for ontogenetic niche shift discussed in chapter 5. Similarly, niche construction by the same or other organisms can produce temporal variation in the habitat IPN, which in turn can change the realized niche of an organism. Indeed, these shifts are at the core of niche construction's effect on natural selection pressures.

For *Amee*, this means that any organism can have multiple niches, both fundamental and realized, during its lifetime when niche construction and plasticity are active, while niche construction alone can only lead to shifts in the realized niches because it does not affect metabolism networks. Indeed, plasticity-rewriting, together with the inclusion of environmental signals into its conditions, captures the principles of environmentally-dependent ontogenesis.

That being said, with regard to the ecospecies concept we employed in chapter 9, we find that given the intra-lifetime variability in ecological niches that is introduced through plasticity and niche construction, the ecospecies concept is insufficient to describe the system's dynamics. As such, they are not investigated in this chapter. Instead of relying on ecospecies alone, we would need a more encompassing concept. This could be based on the genome and the rewriting rules it defines, i.e., be roughly equivalent to a genetic species in nature. This would allow us to investigate the evolution of ontogenesis in conjunction with ecological effects, such that the modification of natural selection on genetic species can be understood as a function of the developmental effect on ecological interactions. While possible in principle, more conceptual work is necessary to map a suitable species concept to *Amee*.

Evolution of development and plasticity for reduced rewriting rules Finally, we investigate the evolution of rewriting investment when rewriting is reduced in impact. We find in Figure 10.10 that the increase in the number of rewriting rules over time relative to the initial number is stronger for the reduced-rewriting case when compared to the baseline case discussed previously, reaching similar levels for some simulations (e.g., seed 798797005 for the plasticity-only case) despite starting with three rules (one per process) instead of nine (three per process). An increase is also present when rewriting is off completely, leveling off within the first third of the simulation time. While the number of attained rewriting rules is mostly within the variability shown by the neutral case, we find the behavior of the mean number of rewriting rules to be qualitatively different, with stepwise increases and often reduced range of variation being apparent, which hints at selection being active, i.e., that the existing rules play a role in the fitness of the organisms. Without niche construction, the individual simulations in each panel also appear to qualitatively differ in their time evolution in Figure 10.10, while the neutral case naturally shows strong overlap. Rewriting consequently alters the system's evolutionary trajectory, as seen before for the baseline case.

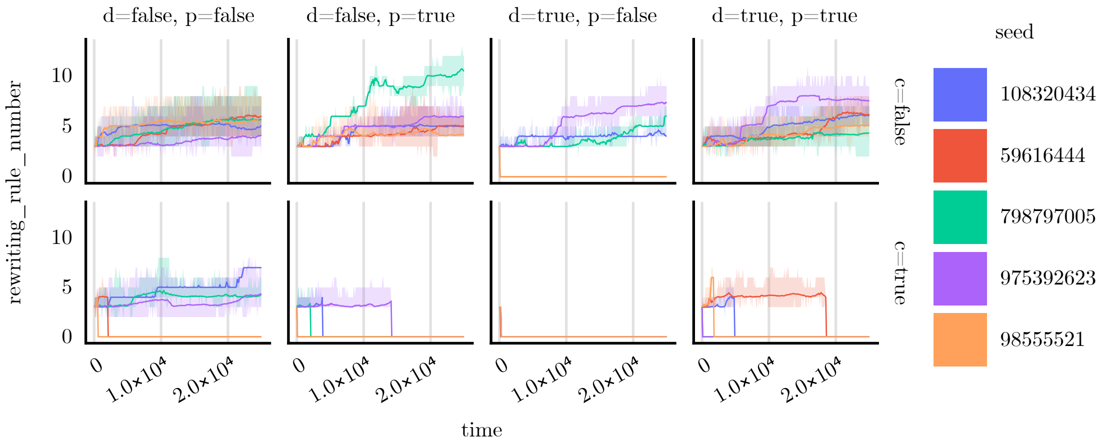


Figure 10.10: Evolution of the mean rewriting rule number (solid line) and the (min, max)-area (shaded) over the population over time for the reduced rewriting case (compare Table 10.1). While mostly within the range of values the neutral case (all rewriting processes off, upper left panel) shows, the behavior is qualitatively different when the rewriting of any kind is on, indicating that at least some rewriting rules are influenced by natural selection.

Developmental rewriting once more appears to be preferred over plastic rewriting when both are active in Figure 10.11, although no clear trend across simulations is apparent. This shows that many different alternative strategies are possible with respect to development, plasticity and genetic evolution. When the mean investment into a rewriting process falls below 0.001 energy units (the minimum rewriting cost corresponding to a single rewritten symbol), this indicates that many organisms in the population don't invest in the respective rewriting process, while a minority still does. Where this is maintained over longer times, this hints at the coexistence of different strategies. Their emergence, evolution and maintenance is an interesting topic because rewriting is costly and thus involves a trade-off against investment into reproduction, but this is left for future research.

Furthermore, we find in Figure 10.12 that plasticity is maintained to a higher degree than before and for longer times, at least by some organisms, although the mean values come very close to zero for seeds 798797005, 975392623 and 98555521. Therefore, the reduced impact of shorter rewriting rules appears to lead to less detrimental effects when compared to the more impactful rules used before. We furthermore find that the temporary increase in populationsize for the case where all three rewriting processes are active in Figure 10.2 for seed 59616444 is connected to a weak increase in developmental energy investment. This is consistent with a scenario in which the innovation that allows for the more effective resource exploitation which must underly this increase arises as a developmental modification. Interestingly, this follows the near complete loss of development after it was maintained for roughly 5000 timesteps, and the increase in mean development investment precedes the increase in population. Moreover, the subsequent extinction is associated with a similar effect, although the exact underlying mechanism remains unclear at this point. Given that the population size reduces quickly, statistical artifacts cannot be ruled out. Similar phenomena occur for seed 798797005 in the leftmost panel of Figure 10.2, lower row. These do not lead to extinction and do not rely on plasticity or

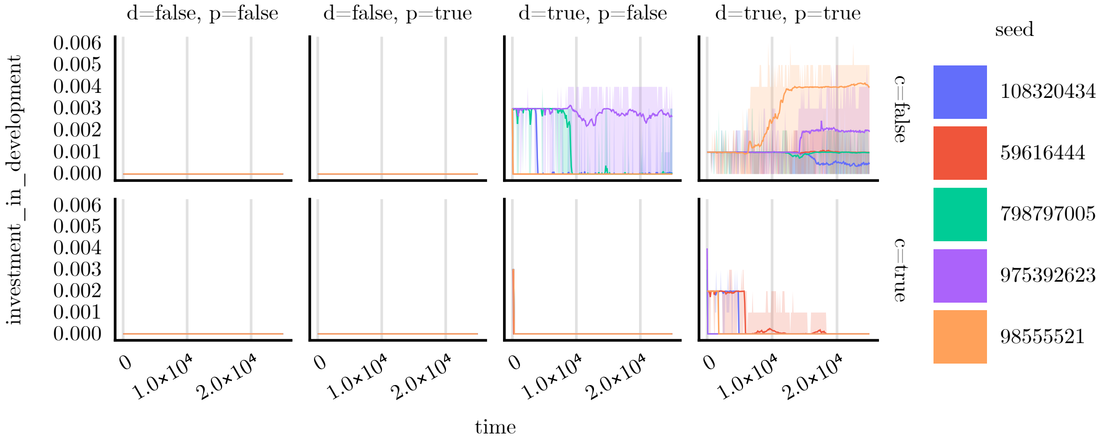


Figure 10.11: Mean investment into development (solid lines) and extend of variation ((min, max), shaded areas) with reduced rewriting rules. We find that development is maintained for longer and to a larger extent relative to the starting point than for the baseline case. When development and plasticity are active concurrently, development is always maintained, although trends in the mean investment into developmental rules are not uniform, indicating that which strategy comes to be prevalent in a system depends on its history and details. The population dies out for development + niche construction, in contrast to the baseline case.

development because they are inactive. Therefore, other effects than metabolic rewriting are more likely to result in these temporary increases, as previously discussed. Moreover, in contrast to the baseline case where development was maintained in conjunction with niche construction, for the reduced case considered now, plasticity is maintained. Therefore, smaller rewriting rules appear to favor plasticity overall, likely because they reduce the cost impact. This show that plasticity and development, acting on the same network, can act in somewhat interchangeable ways for small rules which reflects their symmetric implementation. Irrespective of the parameters, however, we find from the presented results that ontogeny in *Amee* must consequently be seen as a combination of both, plus the initial creation of the metabolism network from the genome. This is in agreement with natural systems as discussed in chapter 5.

In chapter 13, statistics over time are shown of how often organisms have applied a given rewriting process. We find that most organisms apply a rewriting rule only once at most, with many not using them at all, especially for plasticity. Therefore, where rewriting allows for the survival of the population, it seems to rely on its lessened impact, highlighting again the lack of integration of the rewriting system into the evolving population's developmental system.

Evolution of niche construction for reduced rewriting In contrast to rewriting of the metabolic networks, niche construction is frequency dependent because all organisms can rewrite sequentially the same habitat motif. This is a consequence of making rewriting additive which is a modeling choice. Therefore, a plausible hypothesis for why niche construction leads to extinction is that too many organisms with the same rewriting rules affect the same pathway in the habitat network, thereby disrupting it and degrading the environment. We would expect evolution to react to such disruption if the timescales of evolution and niche construction are similar enough, i.e., if the effects of niche construction

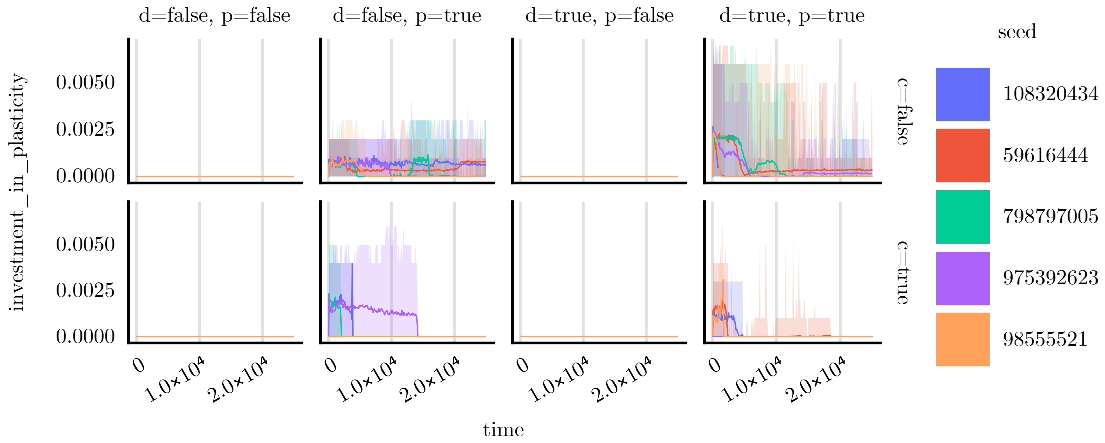


Figure 10.12: Mean investment into plasticity (solid lines) and extend of variation ((min, max), shaded areas) with reduced rewriting rules. We find plasticity to be maintained to a higher degree than before, although this often hinges on comparatively few individuals (mean close to zero, maximum greater zero). For the plasticity-only case, higher rates of plasticity investment are maintained, though mostly without significant trends. We find that plasticity is maintained under niche construction until extinction, in contrast to the baseline case.

on the environment are such that genetic change or plasticity can follow, and conclude further that for the current parameters and the current implementation, this timescale problem is not addressed.

We focus on the case where the number of rewriting rules is reduced because this scenario shows populationsizes for pure niche construction, which is absent for the baseline case.

In Figure 10.13, a comparison with Figure 10.2 shows that niche construction investment is maintained on a very low level for seed 975392623 when it is the sole rewriting process. This is also the only system where a high population size larger than 5000 organisms is maintained continuously. Seed 798797055 shows some niche construction activity that is temporally correlated with the increase in population around timestep 2000 in the lower left panel of Figure 10.2. A continuously maintained niche construction maximum greater than zero means that in every timestep there is at least one organism that engaged in niche construction. Moreover, the amount of maximum energy investment shown corresponds mostly to the application of a rewriting rule that is comprised of a single symbol that is rewritten once, with values above that being found only temporarily. Note again that rewritten structures in the habitat do not decay and that rewriting is additive. Rewriting the same symbol continuously would therefore correspond to an increase in its weight. If this affects an input symbol from a place to a transition, this eventually will reduce the firing frequency of the transition or disable it completely. A single rewrite operation, however, still can have large effects if it changes the structure of the network in an important pathway, or just crosses the threshold that deactivates a transition. That being said, in conjunction with Figure 10.5 and 10.6, niche construction on the low level found here appears to be not harmful. A behavior as found for seed 975392623 in Figure 10.13 when niche construction is the sole active rewriting process consequently allows for the existence of variation in rewriting rule activity and for the integration into the ontogenetic process of organisms. Furthermore, we conclude that intermediate or low levels of active niche construction should be the norm in *Amee*. Indeed, the relatively intensive rewriting

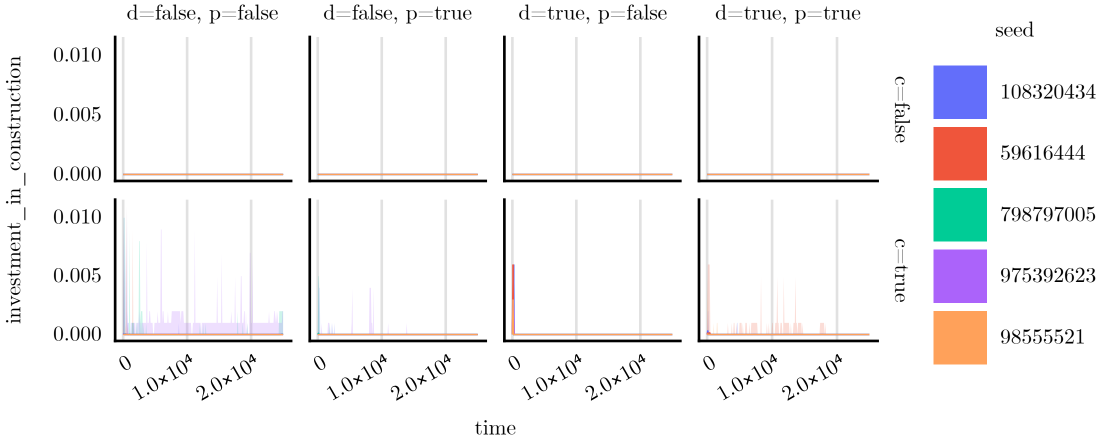


Figure 10.13: Energy investment statistics into niche construction for reduced rewriting rules. Mean being shown as solid lines, (min, max)-extend as shaded areas. Niche construction is never maintained by the majority of the population, and a continuous nonzero maximum is only found for seed 975392623 when niche construction is the sole active rewriting process. Where it is combined with development and plasticity is off, we find relatively intensive rewriting in the beginning which leads to immediate extinction, whereas niche construction is only sporadically active for the rest of the population.

at the beginning of the (d=`true`, p=`false`,c=`true`)-simulation in Figure 10.13 shows a potential case where it had strongly deleterious effects. The spike in investment, produced by the first or the first few individuals, corresponds to three symbols being rewritten, and comparison with Figure 10.14 shows that this is indeed a single rule with three or so different symbols that are affected by it. All in all, it appears that niche construction mostly produces some kind of deleterious influence that cannot be compensated for in the long run. Given that most organisms do not engage in niche construction, this is likely an effect of its nonlinearity, exacerbated by the conditional activity of IPN transitions which is quite easily disrupted.

Given that niche construction influences the environmental structure and that developmental and plastic rewriting rules can conditionally depend on the presence of such structures, and given that they in principle can be important for the formation of an effective metabolism, niche construction also could disrupt these signaling pathways and therefore, especially at the start of a simulation where the population is low and homogeneous, disrupt the necessary developmental or plastic rules where metabolic rewriting is active. Indeed, rewriting rules can be quite specific, especially for small rules, such that any change to the few target nodes such rules have can disrupt them completely. This again hints at a lack of integration of rewriting into a functioning developmental system for the current simulations. All that being said, the system shows a plethora of promising phenomena for future research, but especially for the effect of niche construction, no clear explanations have been derived for its observed behavior in isolation or in combination with other rewriting processes.

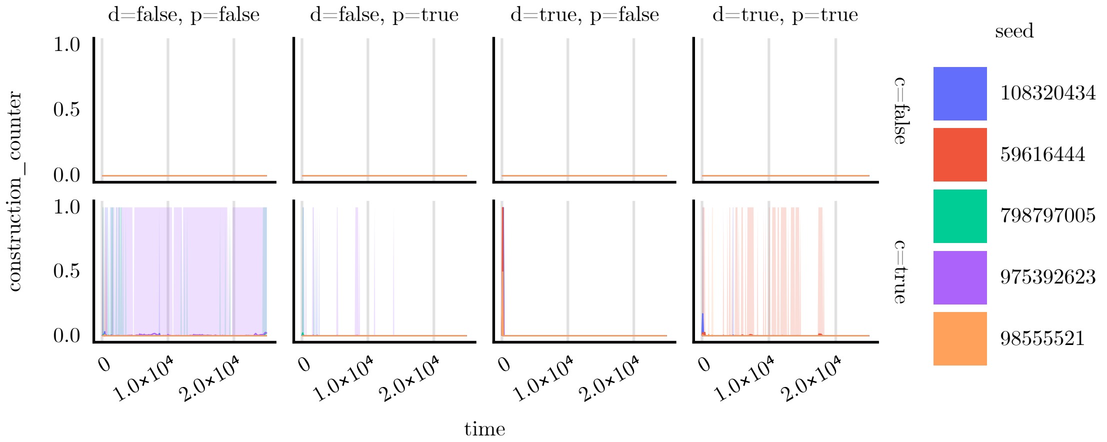


Figure 10.14: Statistics over how often organisms in the population have engaged in niche construction over time, mean as solid lines, (min, max)-range as shaded areas. We find that organisms don't rewrite the habitat more than once and that niche construction remains sporadic for the vast majority of simulations, reflecting niche construction's permanence and the conditional nature of the rewriting rules.

Discussion We forego further analysis and description at this point, because the core effects leading to the observed behavior, even if the details vary, appear to have become clear. Invoking the view of natural organisms as dynamic processing systems for biological information (see subsection 4.1.6), the lack of evolutionary history in the system appears to be the main cause of deleterious effects of rewriting, i.e., the rewriting processes are only required to be active as part of the creation of the initial condition but their adaptive effect over the organism's lifetime is not considered initially, and only unfolds once the simulation is started. Therefore, their interaction is essentially random, and given their complexity it is hardly surprising that rewriting processes generally are not interacting in a way that enhances the fitness of organisms or reflects natural systems more closely which again results in the expectation that larger, longer running systems should eventually evolve more integrated systems. Another way to alleviate this issue lies with the initial genetic algorithm, which currently takes the first solution (for given conditions) that it finds and stops immediately. Instead, it could be kept running after this point with a different fitness function geared towards rewriting integration, such that the "missing" evolutionary history of the system could be included to some degree. Another line of investigation would be larger IPN networks with more places, but without increasing the rewriting rules concurrently. In this way, their effect selective effect could be more limited and the increased possible variation would allow for natural selection to form their interaction over time. That being said, even for the cases presented here, we found possible plasticity-led evolution and advantageous niche construction effects, which allows the expectation that more substantial results are not far off from the current implementation. The interaction network of rewriting rules would be a prime target for future research because it reveals the structure of the developmental system of an organism and could produce insights into the assembly and evolution of such systems. Multiple resources have also been left out of consideration, and given the complexity of the single-resource scenario shown, the impact of multiple resources is difficult to judge. It must be said, however, that the breadth of phenomena *Amee* shows with regard to plasticity, development and niche construction is arguably too large to present in one work, and so the insight gained at this point remains limited.

10.2 Outlook: Spatiality

The results presented here are limited to a single habitat, but all natural systems are spatial in some way. Indeed, spatiality plays a vital role in speciation, ecology, and the differential survival of different lifecycle strategies and is thus essential for the origin and maintenance of biological diversity. *Amee* can represent spatial systems as a two-dimensional, rectangular grid of habitats, each with its own resource-processing IPN, connected by the 8- or 4-neighborhood. Each habitat can be settled by an arbitrary number of organisms, i.e., *Amee* simulates extended habitats as single grid cells inhabited by many organisms.

While not explicitly shown here, in a spatial system, we expect extinctions to be reduced because organisms can move to other habitats if the current one is no longer suitable. Niche construction, assuming it to be integrated into the ontogeny of the individuals in a way that is at least not harmful, can structure the grid and introduce spatial heterogeneity, which in turn increases the niche space size. We, therefore, expect more ecospecies to exist in a spatially heterogeneous system. Moreover, niche construction could enable the emergence of distinct biomes by facilitating the existence of a set of “species” in one part of the grid while excluding others, with niche-constructing organisms playing the role of ecosystem engineers. The disruption of environmentally enabled developmental– or plasticity-pathways provides a powerful mechanism for such exclusions. Additional variability in the population brought about by spatiality could also play a role in the faster evolution of the rewriting systems, thus possibly allowing for the evolution of better integrated ontogenetic systems even without modifying the creation of the initial conditions.

This opens a multitude of future research pathways. For example, possible trade-offs between relocatory– and perturbatory niche construction could be investigated more deeply, thus continuing work by Scheiner, Barfield, et al. [2021a]. The evolution of developmental systems can be studied in a more realistic setting than presented herein, shedding light on the occurrence of plasticity-led evolution. Different regions of a network contain different habitat networks, producing different environmental signals for development, which would allow for polyphenism in a way not possible for non-spatial simulations. The evolution of organism mobility, either in a purely ecological setting or together with different combinations of rewriting processes, is also of interest because it provides a niche space dimension that can be exploited to facilitate coexistence. Finally, the structuring of an initially homogeneous habitat by niche-constructing species and the emerging spatial patterns could be investigated in a system that allows for the easy addition or removal of ontogenetic and ecological processes.

11. Summary and Outlook

11.1 Summary

This thesis is divided into two major parts. In the first part, modern evolutionary theory was reviewed with a focus on self-organization, inheritance, and development, which resulted in viewing evolving entities as processing systems for biological information that unfold via guided self-organization. Development has thereby been identified as a central element and as the focus of evolution.

This view is grounded on the concept of complex constructive systems derived from complex adaptive systems to describe complex adaptive systems that produce, over time, complex and cumulative adaptations to external conditions and *de-novo* innovations in an open-ended fashion. This includes especially the autonomous emergence of higher-level structures, which come about as motifs in lower-level interaction networks.

The idea of CCS rests on the concept of complex adaptive systems, interpreted as networks of interacting agents which themselves are complex and interact nonlinearly. We postulate that a subclass of these systems has a process that evaluates the performance of traits of a CAS or of its elements and a second process that produces new traits on the basis of this performance. As a corollary, these systems produce a reservoir of traits, accessed and modified by the trait discovery process, which we know as memory in the context of learning systems or as the gene- (or, more general, trait-) pool of evolving systems. Indeed, through these three properties, complex constructive systems conceptually unify learning- and evolving systems. "Evolving" does not just refer to evolution by natural selection in a population of biological organisms but includes all possible systems in which Lewontin's three principles – variation, reproduction, and inheritance – are active. This includes culture, which combines learning and evolutionary processes but on a more Lamarckian basis. Lamarckian and Darwinian evolution have been examined in chapter 3 in terms of complex constructive systems, and their respective differences have been established concerning trait discovery, performance evaluation, and trait memory, which was also used to establish some basic properties of Darwinian natural selection and biological evolution in general, like evolutionary relativity and continuity, advantages of sexual selection and the role of neutral evolution.

A large section of this work has been devoted to inheritance processes (chapter 4) and the way they operate in nature, given their central role in the process of Darwinian evolution and the formation of biological CCSs. A dynamical systems interpretation showed how the inheritance process decays into different elements based on the concept of a von Neumann universal constructor, which is made up of an encoding function for heritable information, a decoder system (syntax), and a translator system (semantics). This syntax-semantics split has been identified as a central element and a watershed that divides heritable information systems into those with unlimited encoding capability, which enables the construction of arbitrary phenotypes, and those that lack this split and rely on template-based reproduction, thus being limited by the properties of the system to be reproduced. This property, consequently, introduces a two-level separation in the concept of complex constructive systems, which divides them into an essentially unlimited class (those with syntax-semantics split) and a limited class (those that lack this). Chapter 4 introduced the concept of biological and, building on that, heritable information and showed that the process of copying heritable information from one individual to the next is only one step of the reproduction process, with the translation of this information into a system that can interact with its environment on the one hand, and growth of the produced phenotype into a reproductively viable state on the other being necessary complements to it that are a direct consequence of the assumed nature of biological organisms as CAS. The presentation then proceeded through a zoo of facts from the genetic encoding mechanism over translation and transcription mechanisms to gene-regulatory networks and genetic expression. This introduced the notion of genomes as processing systems for biological information, with gene-regulatory networks implementing the processing logic. The same discussion was then extended to other inheritance forms- epigenetic, ecological, and cultural inheritance. For all of these, it was found that a symbol system in analogy to genetics is largely lacking or at least not fully developed, with cultural inheritance in humans being arguably the most developed. This walk through the garden of inheritance also served to show how one inheritance scheme can create a scaffold for the emergence of another by coopting some of its structural elements, a concept introduced before in the context of Darwinian evolution.

The depth of discussion in this chapter allowed for a more shallow treatment of development and plasticity in chapter 5, which focuses on the processes of ontogenesis and growth and their evolutionary significance, in itself a direct corollary of the dynamical system's view of inheritance. Building on chapter 4, it was examined how these processes create and modify phenotypic variation on which natural selection can act and that, as a consequence, an evolving population of organisms must be regarded as an evolving population of lifecycles, i.e., organisms are not just static vehicles for genes or interactors with constant structures, but are dynamic entities that inherit, grow and reproduce in response to environmental conditions and thereby form the substrate on which natural selection ultimately acts. While development is strictly only a property of multicellular organisms, the fact that unicellular organisms develop in organized cycles allows us to see lifecycles, i.e., developmental processes, as the fundamental element on which evolution acts.

This idea was then illuminated from an environmental point of view in chapter 6, where it was shown how the concept of organisms as dynamic processes extends into their environment through their actions and the artifacts they produce. These change how they themselves, but also many others, interact with their environment and thus modify the selection regime they are subject to. Thus, natural organisms shape their own evolution in two ways - through developmental and phenotypic plasticity on the one hand and through niche construction and extended phenotypes on the other. In other words, developmental processes are not passive consumers of biological information but have evolved elements that actively modify and create the signals they use to unfold. Moreover, environmental modifications can produce spillover effects that change the dimensionality of the niche space and the system's carrying capacity both in terms of pure populationsizes but also in terms of niche- and species diversity.

Moving over to the second part of the thesis, the utopia project has been presented in chapter 7, which is a software framework for the modeling of complex adaptive systems using C++ and Python, the data-management part of which has been developed as a part of the present work. The basis of this module is formed by the HDF5 package, with a modern wrapper library and management system for the scheduling and execution of complex data extraction tasks.

The Utopia package was then used to implement a model for biological CCS (*Amee*) (presented in chapter 8), which is based on the process-modeling language of Petri-Nets and represents the first attempt at employing such an architecture to model evolution. The guiding principles have thereby been derived from the theory part, which results in a comprehensive model that includes genetic evolution, plasticity, development, and niche construction in addition to a basic resource-consumer ecology.

The IPN-language provides us with a set of basic building blocks that can be seen as self-organizing, i.e., building blocks form Petri-nets by virtue of their own underlying dynamics. This highly simplistic “artificial chemistry” forms the lowest level of organization and allows for the creation of complex, adaptive systems through the combination of its elements based on simple rules.

Evolution then acts on the second layer of organization in the model: Organisms are conceived as being comprised of a self-reproducing, growing metabolism, which in turn is modeled with an open IPNs. The habitat these organisms live in is imagined as an IPN as well, which represents a section of a natural ecosystem that we see as a resource-consumer network. The organisms can attach to the nodes of this network, extract resources from them, process them through their metabolism, and use them to reproduce their own structure, which is encoded in a linear genome that encodes their metabolism IPN. Thus, we assume a syntactic-semantic split to be present and do not examine its evolution. The highest layer of the system is comprised of net rewriting systems, i.e., a set of rules attached to each organism that defines how a given motif in the network can be replaced with another IPN. While based on NRSs, this system has been modified significantly to include environmental and organismal signals into whether a rewriting rule can be applied. This scheme implements development, phenotypic plasticity, and niche construction with a single algorithm. In total, *Amee* was aimed at implementing the principle structure of biological CCSs in a simplified manner, which is still complex enough to create a wide variety of natural phenomena.

Finally, experiments have been performed for the non-rewriting part of the system with single and multiple resources in a non-spatial setting that elucidated the emerging ecology of the system. It has been shown that the system unfolds via two processes: Adaptive radiation of an ecospecies into a group of successively dominant but related ecospecies that leads to the exploration of a subset of the available niche space and the eventual attainment of a local fitness maximum. Thereafter, the system can stay confined to this subspace for long times, with further ecological turnover most likely being driven by neutral processes or, at most, by weak selection. This regime leads to a differentiation of ecospecies such that a group of coexisting but related ecospecies evolves. While most are very shortlived because each new ecospecies starts out with a single organism, the evolution of the system nevertheless contains remarkably long-lived ecospecies. For a single resource, this process gives rise to evolutionary innovation that makes it possible to leave the local niche space volume and go on to produce adaptations that allow for higher carrying capacity and more efficient usage of the available resources. These innovations could happen at any time in the simulation and give rise to a new bout of adaptive radiation that eventually leads to the attainment of a new local fitness maximum and the re-establishment of the niche-differentiation regime. All these findings are largely consistent with expectations based on observation. Thus, *Amee* produces a form punctuated ecological equilibrium, i.e., ecological dynamics with a neutrally dominated species overturning rate, interrupted by evolutionary innovations that can lead to strong regime shifts. The influence of the network structure of the habitat as a structuring pattern for the evolution of the population has been left for future work. However, it can be expected to be very significant for the adaptation options of the population and, thus, for the observed evolutionary dynamics.

For multiple resources, it has been found that, due to the more complex metabolism, evolution is more restricted and consequently much slower, with almost no substantial evolutionary innovations being observed that would yield a change in the carrying capacity of the system on the level of those seen for a single resource. Rather, the system transitioned quickly into the punctuated equilibrium regime, largely restricted to the initially occupied niche space volume. In all cases, it was revealed that the system mostly stays confined to the initial habitat utilization pattern, hinting at the difficulty of fundamental niche change, which would likely require much larger or longer simulations.

A closer investigation of the evolution of the metabolism revealed a highly complex phenomenology that ranges from a “survival of the flattest” effect in which organisms evolve short, high-efficiency metabolic pathways to metabolic switches in which “parasitic transitions” that have no post-set are turned off by increasing their input weights to the point where they are no longer enabled at the same time as important metabolic pathways that otherwise would be harmed by their presence.

On the other hand, it has been shown that the metabolism is evolutionarily “brittle”, i.e., its evolvability is restricted, most likely due to the way the firing condition for transitions is implemented and because the network architecture is, due to the complicated interdependencies of its pathways, challenging to change. The former issue can be changed relatively simply while the latter, however, is intrinsic to such an architecture, which begs the question of how the biosphere solved this problem. The most likely explanation is time and sample size, but also probably modularity (compare chapter 4).

Finally, it has been shown that the system produces higher-order groupings of ecospecies, which should be expected given previous results and the nature of evolutionary change. While the methodology has potential for improvement, the current version was sufficient to qualitatively show how the individual ecospecies evolution discussed above merges, for higher-order clusters (“eco-genera”), into a longer-term punctuated equilibrium with far lower overturning rates, i.e., much increased lifetimes of eco-genera when compared to ecospecies, which represents a natural outcome for such systems. Therefore, we can see the radiation of an initial niche as an adaptive radiation of a single genus. Whether this resulted in the coexistence of multiple such eco-genera on a similar population level largely depended on system details, but as a rule of thumb, a single one was dominant while others existed on low levels. However, it has been shown that long-existing eco-genera that always had low populations can still give rise to ecological innovations that change the carrying capacity of the system, an observation consistent with the evolutionary behavior of many natural families of animals.

Overall, the analysis was somewhat hampered by computational limitations, which made the restriction to intervals of 5000 timesteps necessary. Increasing these to the full simulation length would produce deeper insight into the various evolutionary phenomena, as would longer simulation runs with larger sizes.

Finally, it has been shown how sensitive the system reacts to variation in the initial condition and small changes in the sequence of random events in a simulation, which is consistent with the behavior of natural evolving and generally nonlinear systems. This behavior is primarily owed to the complexity of the organism structure and must be considered when planning experiments with the model.

Rewriting has only been investigated superficially because this work aims to showcase the potential *Amee* has for future research instead of presenting individual case studies in depth. Evolving interactions have been found between development and plasticity. We found hints at plasticity-led evolution, which provides a promising route for further investigation. Given that both of these processes have been implemented in the same way and are irreversible, they both contribute to ontogeny and thus, these interactions are not surprising. How and if this interaction plays out, however, depends on the system details and the evolutionary history of the population in question. Indeed, extinctions are frequent as soon as rewriting processes are included, especially for niche construction. Surprisingly, while all niche constructing populations die out for the baseline rewriting case, niche construction allows for very high populationsizes when the size and number of rewriting rules are reduced, indicating that it can have facilitating effects, too. However, the details of the underlying mechanisms have not been elucidated more deeply and remain largely open for further investigation.

Given that facilitating effects of rewriting, no matter the exact process, are comparatively rare, we concluded that the rewriting rules are not sufficiently integrated into the unfolding of the organisms, i.e., the evolutionary history of the system is too short for them to form a processing system for biological information that would stably contribute to individual fitness in the way it happens in natural organisms. A few possibilities to alleviate this have been discussed, with the most promising variant probably being a run-up phase attached to the genetic algorithm that produces the initial condition such that a longer evolutionary history before the actual data-producing simulation can be emulated. Mechanistic issues have been investigated, too, with transitions either being fully enabled or not at all being a primary candidate for reducing evolvability. While various alternatives do exist (continuous Petri-nets, [Alla and David, 1998]), none have been considered in depth within *Amee* in the present work. Spatiality has only been theoretically discussed. However, this aspect has received much attention in the literature in a much deeper way than can be accomplished within the confines of this work.

11.2 Outlook

Naturally, a model as complex as *Amee* cannot be comprehensively presented in one thesis, and so, only the surface has been scratched here, the aim being a proof of concept in conjunction with the underlying conceptual theory. This provides many opportunities for future research, mostly with respect to the effect of rewriting rules.

The rewriting processes represent an aspect of the theory of evolution that is controversial (plasticity-led evolution, niche construction theory, and the extended evolutionary theory), hard to investigate in nature, and has received much theoretical attention over the last few years and even decades. Therefore, their further exploration should probably be the primary route for further exploration. In particular, *Amee* allows for the investigation of the interactions of plasticity, development, and niche construction on the basis of complex, self-organizing individuals where organism fitness is an emergent quantity and not an ad-hoc parameter that is baked into the model structure. Therefore, we can expect any result to be more fundamental, although also more dependent on the implementation details. In particular, the interaction network between rewriting rules which belong to different rewriting processes could be investigated, which could open the door to comparisons with the architecture of natural regulatory systems.

This aspect can be combined with the analysis of evolving resource-consumer networks, a property of *Amee* that has received little attention in the present work. With this, *Amee* can be developed further into an individual-based ecological model that arguably includes all fundamental evolutionary processes. The rewriting mechanism and the mechanism of open Petri-Nets furthermore allow for the implementation of other ecological interactions than resource competition and constructive facilitation or harming, e.g., parasitism or predation via attaching organisms not to the habitat but to other organisms, which opens the door for deeper ecological considerations.

On the more technical side, longer runtimes with larger populations should reveal more of the evolutionary phenomenology of the systems, which, however, is always the case for sufficiently complex evolutionary systems. Nevertheless, because it has been found that rewriting systems needs to be integrated better via natural selection, this is a promising avenue. Finally, other parameter combinations for the number of resources, habitat- and organism node sets should be investigated, for it is likely that the system behaves differently for a large number of small, variable resources and many metabolism nodes than for the large, relatively homogeneous resources and low node numbers that where used here. This could yield more complex resource– and metabolism systems, which should result in more interesting and complex metabolic evolution. Spatiality is also an unexplored avenue that, as Scheiner, Barfield, et al. has shown, can interact with niche construction or plasticity in complicated and interesting ways. Finally, the implementation of the IPNs can be changed, in particular with respect to the firing condition for the transitions. If this was made to allow for more flexible resource conversion, we could expect evolution for multiple resources to be more variable, i.e., for evolvability to increase.

The concepts represented here can be extended further to allow for information processing instead of, or in conjunction with, resource processing. Indeed, there is no reason why the Petri-Net formalism shown herein, which is a Turing-complete modeling language, should not be able to represent learning systems, at least in a very simplistic case. While this would arguably require deeper additions to the mechanics of the system, such changes are rather easily made, opening the model towards studying complex evolving systems with social learning or towards investigations of sociological systems.

Part III

Appendix

12. Appendix 1: Additional Material about single-habitat simulations without rewriting

12.1 Additional visualizations for “Baseline 2 – Multi-resource, single-cell systems without rewriting”

Shown here are additional visualizations for the simulations in Figure 9.2.1.

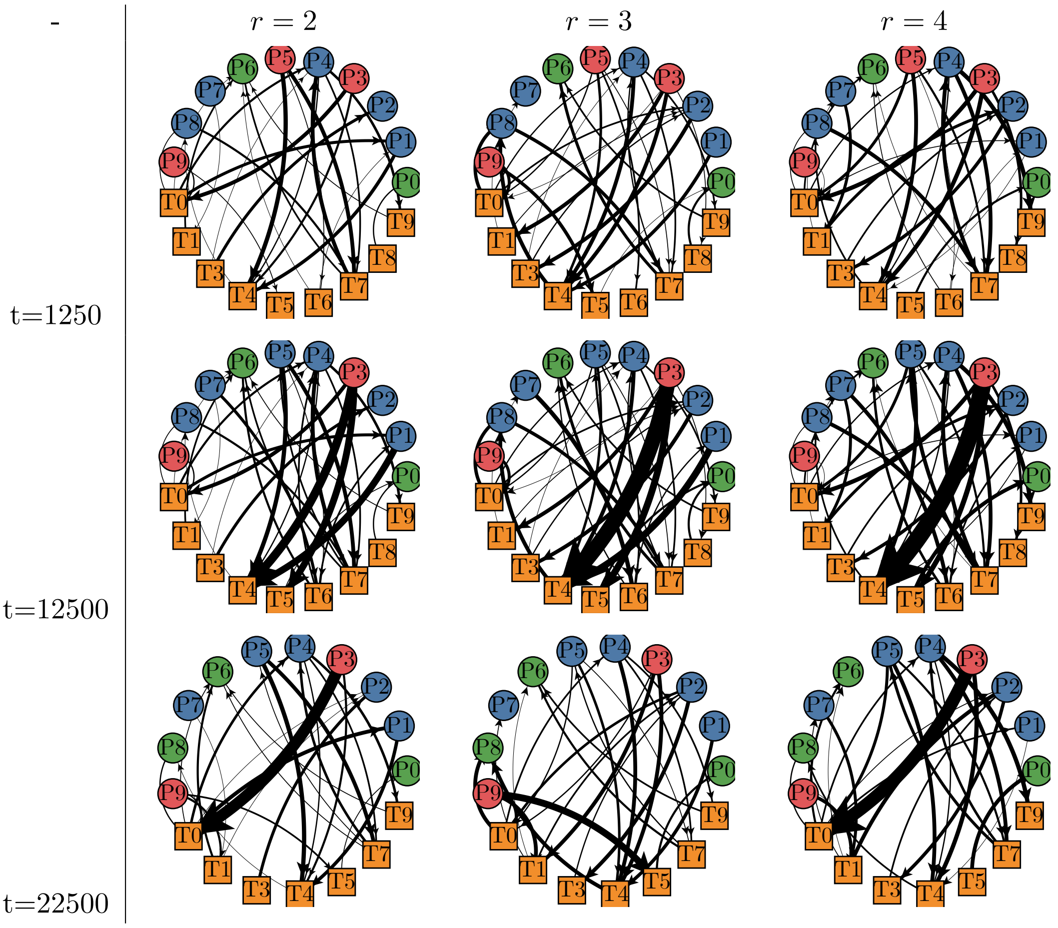


Figure 12.1: Metabolism networks for the simulation with seed 59616444 for randomly chosen individuals of the largest niche in timesteps (from top to bottom) 1250, 12500, 22500, and for resources 2, 3, 4 (from left to right) with 1,5 being all zero. Input interfaces are colored red, output interfaces green, as before.

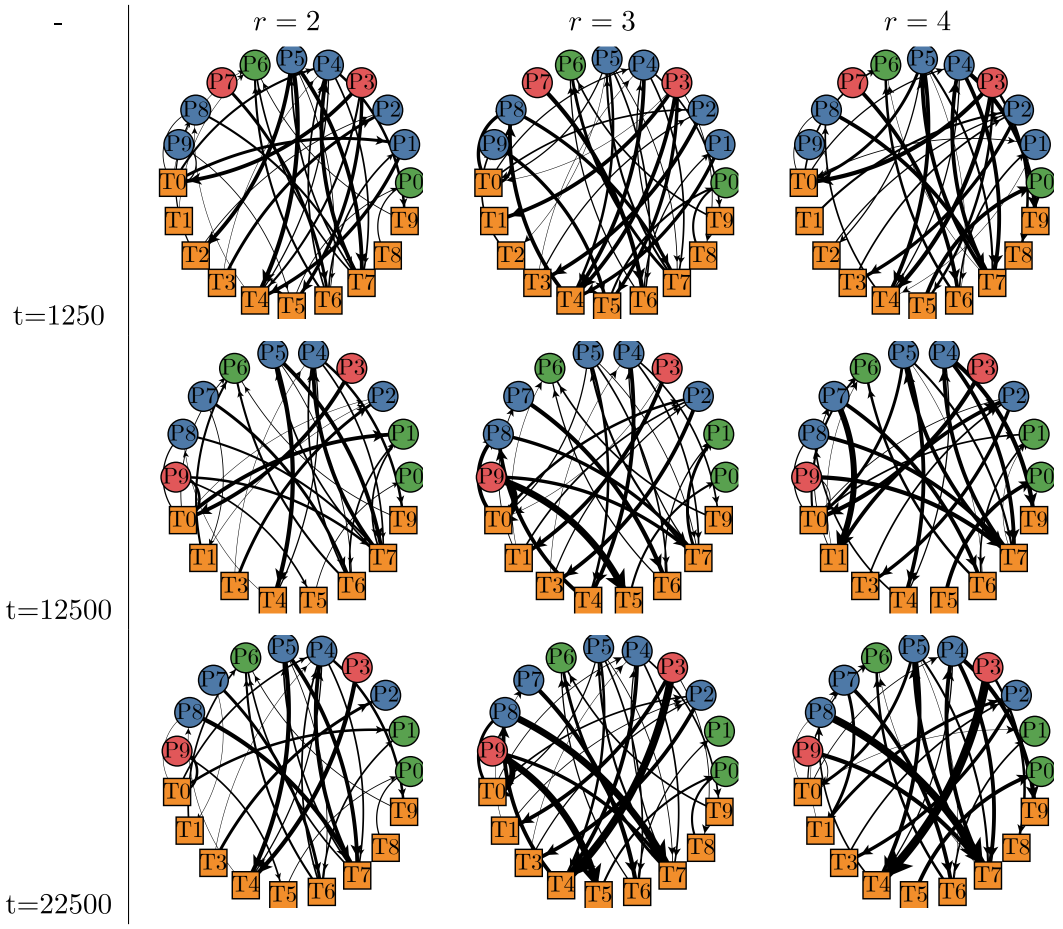


Figure 12.2: Metabolism networks for the simulation with seed 98555521 for randomly chosen individuals of the largest niche in timesteps (from top to bottom) 2500, 12500, 22500, and for resources 2, 3, 4 (from left to right) with 1,5 being all zero. Input interfaces are colored red, output interfaces green, as before.

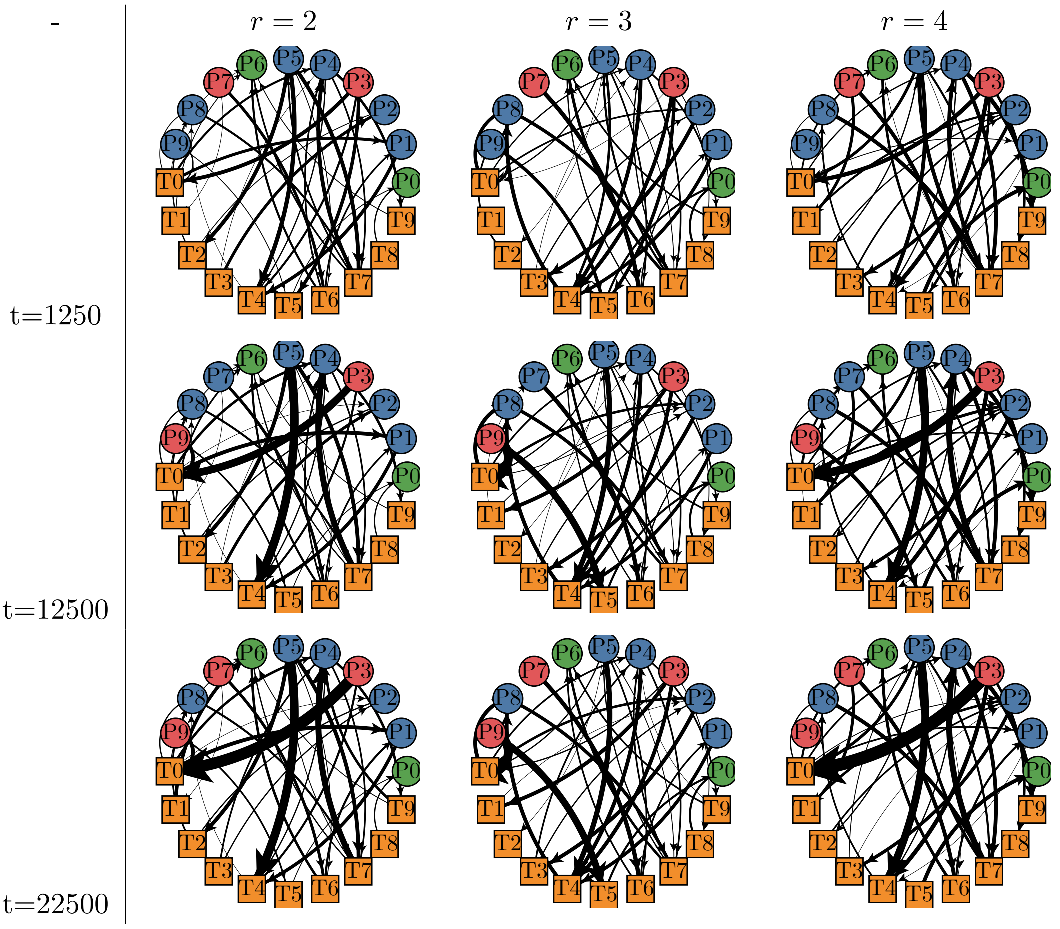


Figure 12.3: Metabolism networks for the simulation with seed 975392623 for randomly chosen individuals of the largest niche in timesteps (from top to bottom) 2500, 12500, 22500, and for resources 2, 3, 4 (from left to right) with 1,5 being all zero. Input interfaces are colored red, output interfaces green, as before.

12.2 Ecological niches and ecological species for multiple resources with higher mutation rate

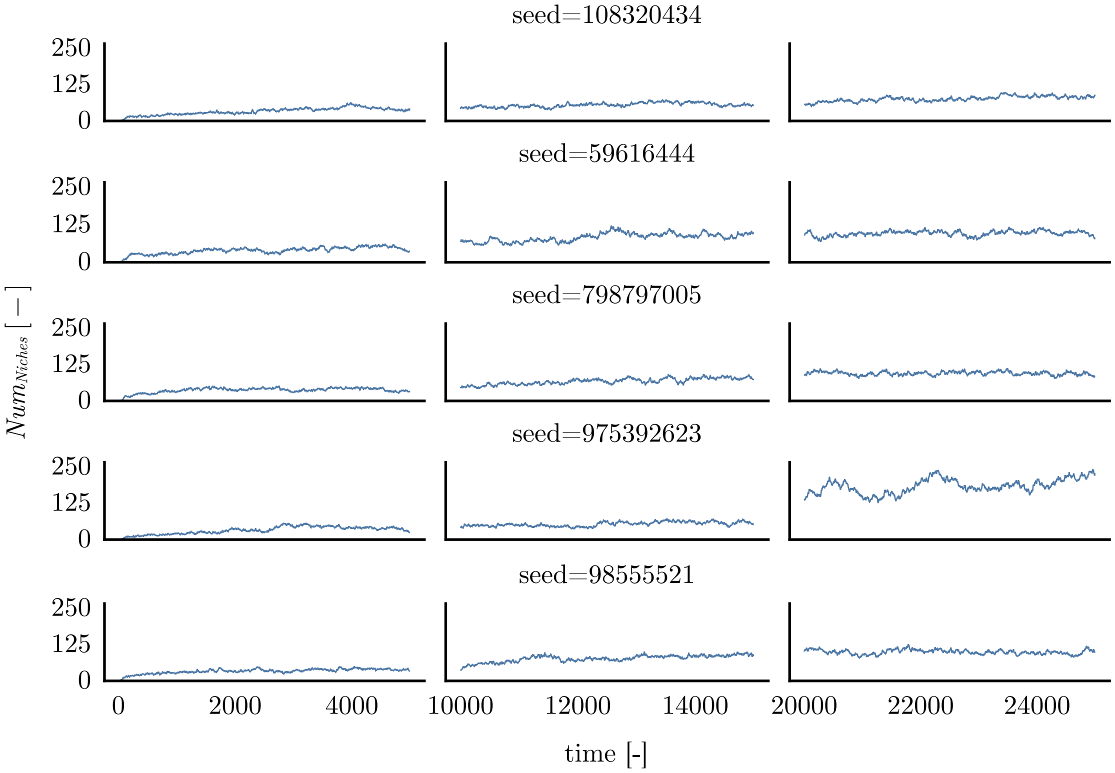


Figure 12.4: Evolution of the number of ecospecies for higher mutationrates. A generally increasing trend can be seen for the first two intervals, with the simulation with seed 975392623 being of particular interest because it produces more than double the diversity of the next highest for the latest interval.

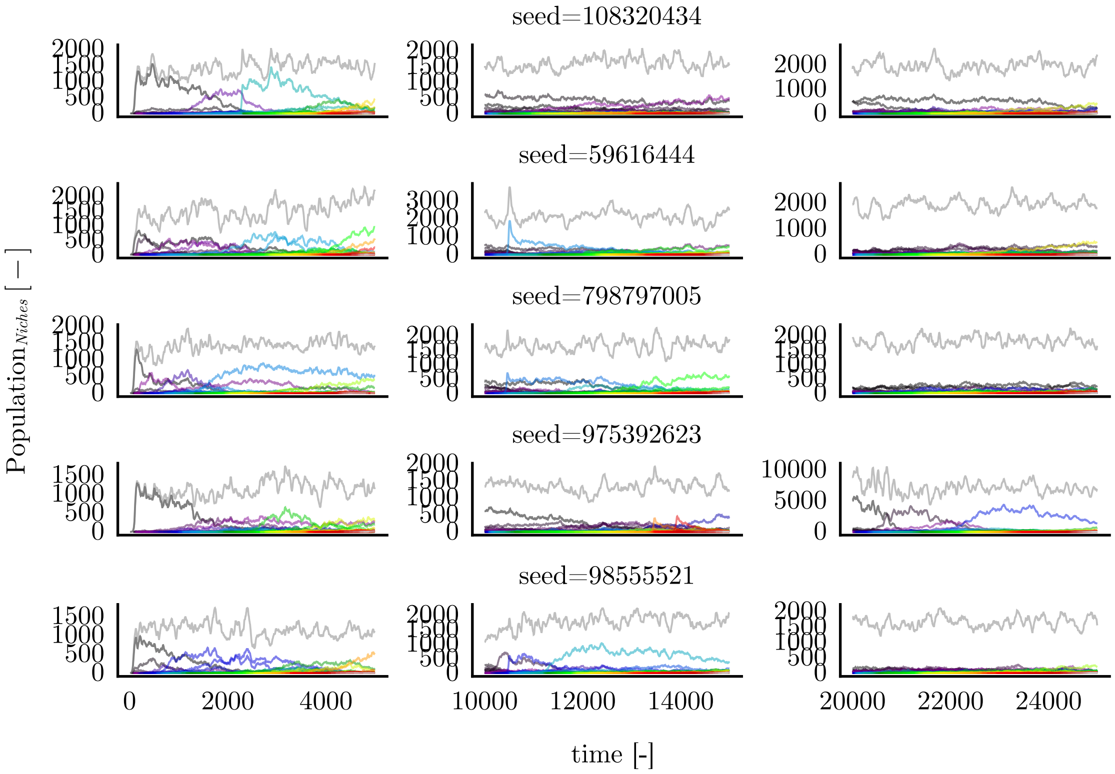


Figure 12.5: Evolution of the population of ecospecies over time, analogous to Figure 9.11 and Figure 9.31. We find that while the populationsize is mostly comparable to the simulations for lower mutationrate and to the single resource case, seed 975392623 produces a roughly five times higher population than the next higher, with dominant ecospecies succeeding each other. Note that given the higher mutation rates, the background diversity is much higher, too (compare Figure 12.4). While the exact transition from the low to the high population state has not been recorded, we identify this simulation as an instance where the higher mutationrate allowed for an evolutionary innovation that allow the system to access a different part of the niche space where higher population densities are possible.

13. Appendix 2: Additional Material about single-habitat simulations with rewriting

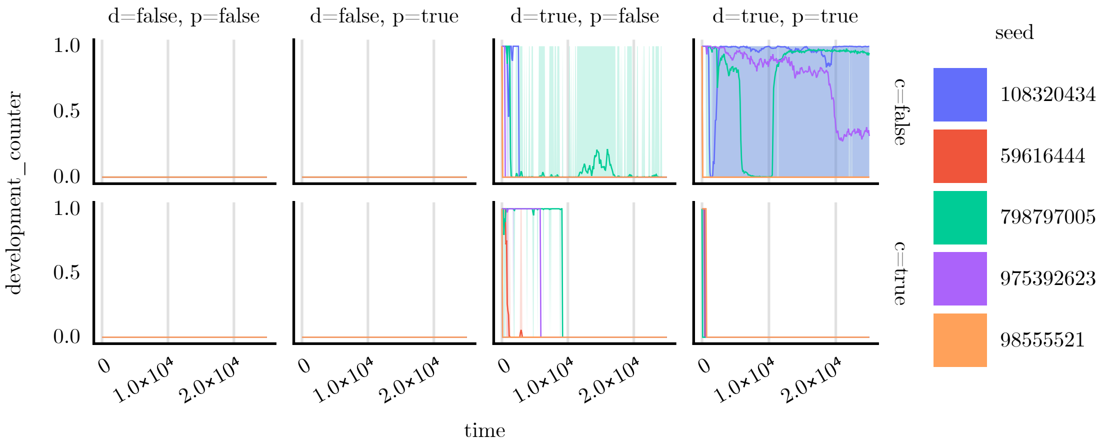


Figure 13.1: Statistics of development counter over time for simulations for the baseline rewriting case. Development is only active once at most, such that only one rule is active.

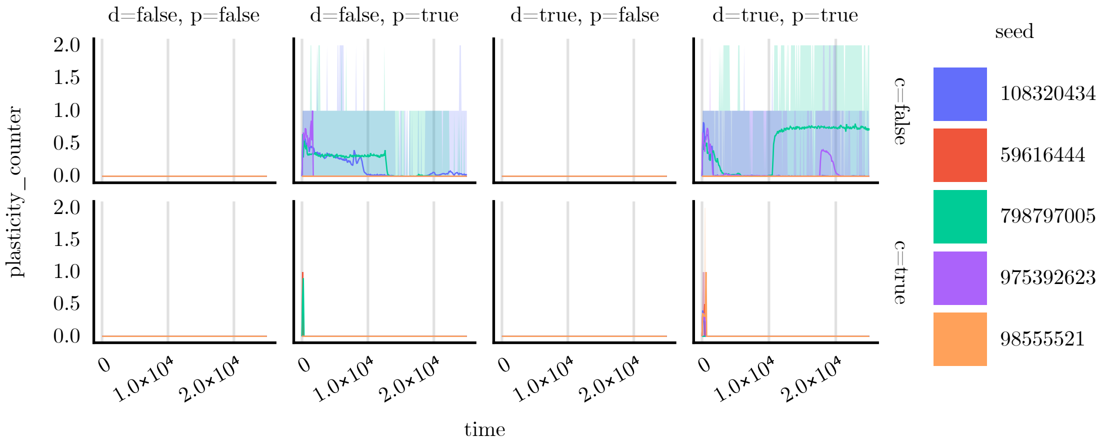


Figure 13.2: Statistics of plasticity counter over time for simulations for the baseline rewriting case. Plasticity only happens once at most for most timesteps, with higher numbers of active rewriting rules being infrequent and not sustained over long times.

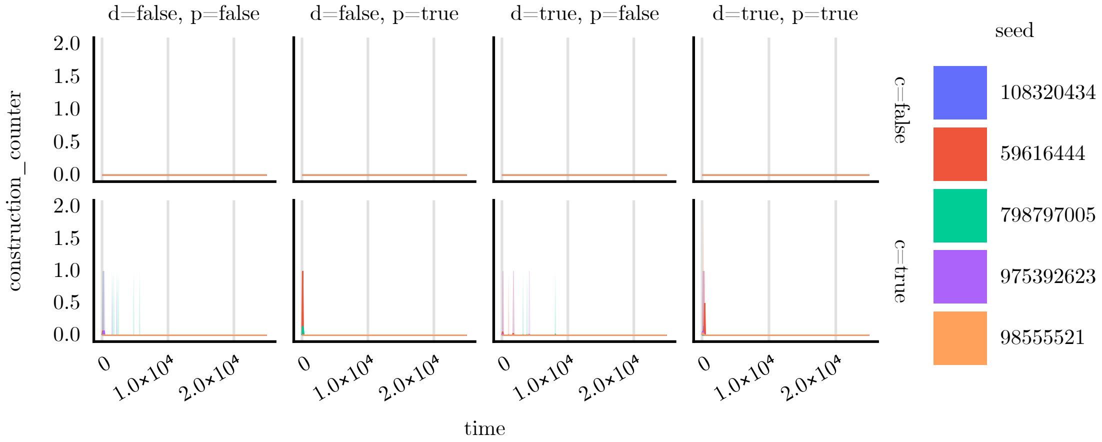


Figure 13.3: Statistics of niche construction counter over time for simulations for the baseline rewriting case. Niche construction leads to extinction for the baseline case, but where it occurs in the beginning, it only occurs once.

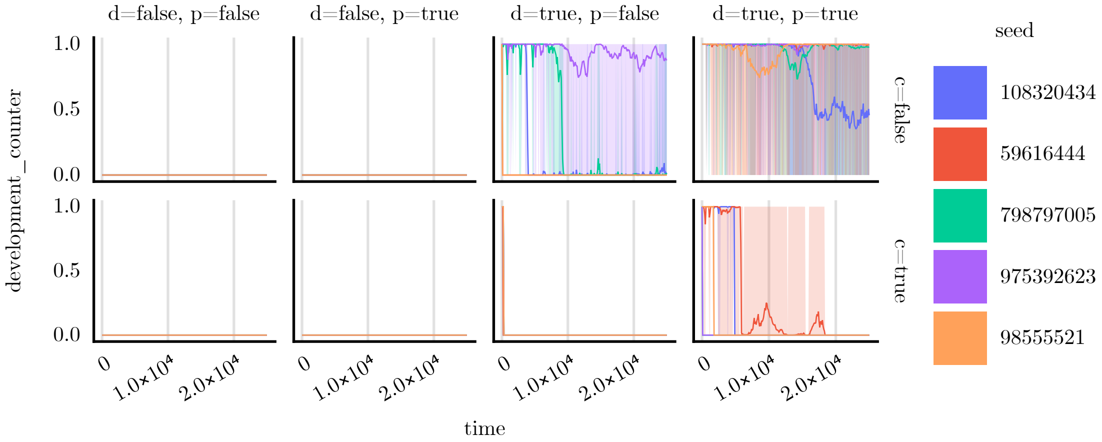


Figure 13.4: Statistics of development counter over time for simulations for the reduced rewriting case. We find a largely analogous behavior to Figure 13.1, but with more active populations and long-term survival even when all rewriting processes are active.

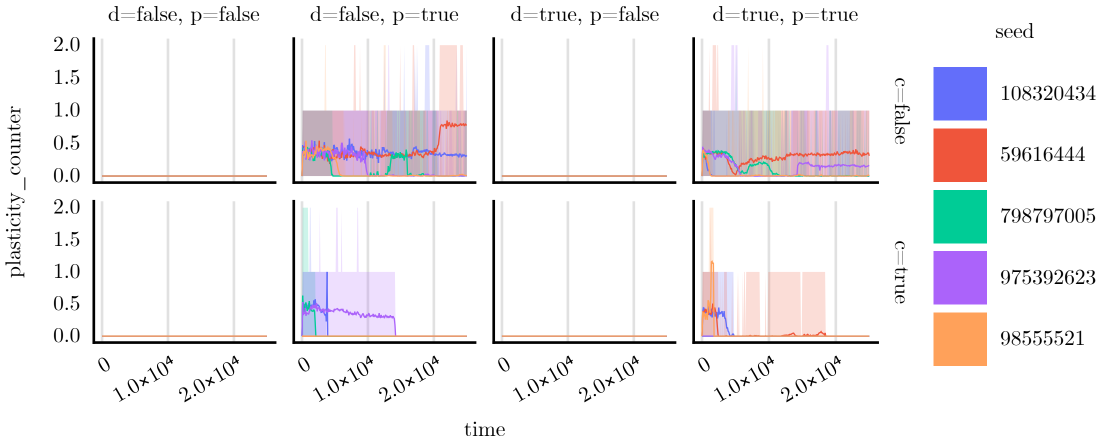


Figure 13.5: Statistics of plasticity counter over time for simulations for the reduced rewriting case. We find a largely analogous behavior to Figure 13.2, but with more active populations and long-term survival even when all rewriting processes are active.

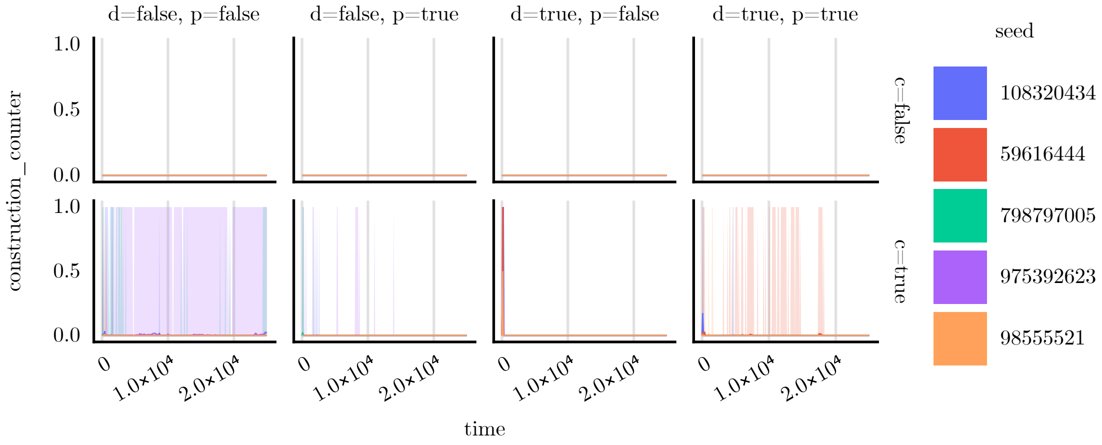


Figure 13.6: Statistics of niche construction counter over time for simulations for the reduced rewriting case. Niche construction is only sustained for seed 975392623 and only in a very small number of organisms. Additionally, no trends are visible, probably indicating selective neutrality at this stage.

14. Acknowledgement

First of all, I would like to thank Prof. Dr. Kurt Roth for many helpful discussions and inspiration for all his support, patience and refusal to give up on me. Furthermore, I would like to thank Prof. Dr. Michael Hausmann for agreeing to be the second referee on this thesis and for his patience and understanding. A big thanks to my fellow Ph.D. students Lukas Riedel, Yunus Sevinchan and Benjamin Herdeanu for many fruitful discussions and many a laugh. To Yunus Sevinchan and Benjamin Herdeanu my special thanks for proofreading. I'd like to thank Angelika Gassama for helping with many an organizational problem and for always being there for a little chat. My gratitude goes also to my family, without whose support I could not have finished this work. Finally, I'm grateful to Hannes Bauser, Daniel Berg and Stefan Jaumann, who completed their theses a while before me and to whom I could look up for guidance.

Acronyms

IPN	Petri-net with inhibitor arcs.
PN	Petri-net.
GA	Genetic algorithm
NRS	Net-rewriting system
RNG	Random-Number-Generator
Amee	**A** **m**odel for **e**co-**e**vo-devo. Name for the agent-based computer model presented in this thesis.
CAS	Complex Adaptive System
CCS	Complex Constructive System
seed	The initial value for a random number generator algorithm.
RNN	Recurrent neural network, a kind of artificial neural network that allows connections between nodes to form cycles.
ANN	Artificial neural network.
LUCA	Last Universal Common Ancestor, i.e., the youngest organism to which all living things today are related.
GRN	Gene regulatory network
RNA	Ribo-nucleic acid. Analogous to DNA but uses ribose as a sugar backbone and uracil instead of thymine.
mRNA	Messenger RNA, used in the transcription process from DNA that is the first phase of gene translation to protein.

tRNA	transfer-RNA. Small RNA molecules that bind to an amino acid on one side and to a codon on the other. Important during the translation phase of gene expression. Implement the genetic code.
DNA	Desoxyribo-nucleic acid. Macromolecule made up of a sugar-phosphate backbone with one of four bases (adenine, thymine, cytosine, guanine) attached to it, used in long chains to store genetic information in cells.

Own Publications

I contributed to the following publications as coauthor. They present the *Utopia* modeling framework which has been discussed in chapter 7.

- L. Riedel et al. [2020]. "Utopia: A Comprehensive and Collaborative Modeling Framework for Complex and Evolving Systems." In: *Journal of Open Source Software* 5.53, p. 2165. DOI: `10.21105/joss.02165`
- Y. Sevinchan, B. Herdeanu, H. Mack, et al. [2020]. "Boosting Group-Level Synergies by Using a Shared Modeling Framework." In: *Lecture Notes in Computer Science.* Ed. by V. V. Krzhizhanovskaya et al. Springer International Publishing, pp. 442–456. ISBN: 978-3-030-50436-6. DOI: `10.1007/978-3-030-50436-6_32`

Bibliography

Adami, C., C. T. Brown, and W. Kellogg (1994). "Evolutionary learning in the 2D artificial life system "Avida"." In: *Artificial life IV*. Vol. 1194. Cambridge, MA: MIT Press, pp. 377–381. DOI: `https://doi.org/10.7551/mitpress/1428.001.0001`.

Agrawal, A. A., C. Laforsch, and R. Tollrian (1999). "Transgenerational induction of defences in animals and plants." In: *Nature* 401.6748, pp. 60–63. DOI: `10.1038/43425`.

Alattas, R. J., S. Patel, and T. M. Sobh (2018). "Evolutionary Modular Robotics: Survey and Analysis." In: *Journal of Intelligent & Robotic Systems* 95.3-4, pp. 815–828. DOI: `10.1007/s10846-018-0902-9`.

Alla, H. and R. David (1998). "Continuous and hybrid Petri nets." In: *Journal of Circuits, Systems, and Computers* 8.01, pp. 159–188.

Allen, G. E. (2014). "The history of evolutionary thought." In: *The Princeton Guide of Evolution*, pp. 10–27.

Anderson, P. W. (1972). "More Is Different." In: *Science* 177.4047, pp. 393–396. DOI: `10.1126/science.177.4047.393`.

Anderson, P. W. (2018). *The Economy As An Evolving Complex System*. en. CRC Press. ISBN: 978-0-429-96139-7.

Ariew, A. and R. C. Lewontin (2004). "The Confusions of Fitness." In: *The British Journal for the Philosophy of Science* 55.2, pp. 347–363. DOI: `10.1093/bjps/55.2.347`.

Badyaev, A. V. (2009). "Evolutionary significance of phenotypic accommodation in novel environments: an empirical test of the Baldwin effect." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1520, pp. 1125–1141. DOI: `10.1098/rstb.2008.0285`.

Baedke, J. (2020). "Mechanisms in Evo-Devo." In: *Evolutionary Developmental Biology*. Ed. by L. Nuno de la Rosa and G. Müller. Springer International Publishing, pp. 1–14. ISBN: 978-3-319-33038-9. DOI: `10.1007/978-3-319-33038-9_94-1`.

Bailey, J. K. et al. (2009). "From genes to ecosystems: an emerging synthesis of eco-evolutionary dynamics." In: *New Phytologist* 184.4, pp. 746–749. DOI: `10.1111/j.1469-8137.2009.03081.x`.

Bak, P., C. Tang, and K. Wiesenfeld (1987). "Self-organized criticality: An explanation of the 1/f noise." In: *Physical Review Letters* 59.4, pp. 381–384. DOI: `10.1103/physrevlett.59.381`.

Baldwin, J. M. (1896). "A New Factor in Evolution." In: *The American Naturalist* 30.354, pp. 441–451. DOI: `10.1086/276408`.

Barbieri, M. (2008). "Biosemiotics: a new understanding of life." In: *Naturwissenschaften* 95.7, pp. 577–599. DOI: `10.1007/s00114-008-0368-x`.

Barker, J. S. ((2009). "Defining Fitness in Natural and Domesticated Populations." In: *Adaptation and Fitness in Animal Populations*. Springer Netherlands, pp. 3–14. DOI: `10.1007/978-1-4020-9005-9_1`.

Bateson, B. (2009). “Problems of Heredity as a subject for Horticultural Investigation.” In: *William Bateson, Naturalist: His Essays and Addresses Together with a Short Account of His Life.* Cambridge Library Collection - Darwin, Evolution and Genetics. Cambridge University Press, pp. 171–180. DOI: `10.1017/CBO9780511693946.003`.

Battiston, F. et al. (2020). “Networks beyond pairwise interactions: Structure and dynamics.” In: *Physics Reports* 874, pp. 1–92. DOI: `10.1016/j.physrep.2020.05.004`.

Begon, M., C. R. Townsend, and J. L. Harper (2005). *Ecology. From Individuals to Ecosystems.* Blackwell Publishing Limited, p. 738. ISBN: 9781405111171.

Belousov, B. P. (1959). “A Periodic Reaction and Its Mechanism.” In: *Collection of Short Papers on Radiation Medicine for 1958* 147.145, pp. 145–147.

Bezanson, J. et al. (2017). “Julia: A Fresh Approach to Numerical Computing.” In: *SIAM Review* 59.1, pp. 65–98. DOI: `10.1137/141000671`.

Black, A. J., P. Bourrat, and P. B. Rainey (2020). “Ecological scaffolding and the evolution of individuality.” In: *Nature Ecology & Evolution* 4.3, pp. 426–436. DOI: `10.1038/s41559-019-1086-9`.

Boccaletti, S. et al. (2006). “Complex networks: Structure and dynamics.” en. In: *Physics Reports* 424.4, pp. 175–308. DOI: `10.1016/j.physrep.2005.10.009`.

Bonabeau, E., G. Theraulaz, and J.-L. Deneubourg (1998). “Group and Mass Recruitment in Ant Colonies: the Influence of Contact Rates.” In: *Journal of Theoretical Biology* 195.2, pp. 157–166. DOI: `https://doi.org/10.1006/jtbi.1998.0789`.

Bonduriansky, R. and T. Day (2014). *Extended Heredity A New Understanding of Inheritance and Evolution. A New Understanding of Inheritance and Evolution.* Princeton University Press, p. 280. ISBN: 9780691157672.

Bottjer, D. J. (2010). “The cambrian substrate revolution and early evolution of the phyla.” In: *Journal of Earth Science* 21.S1, pp. 21–24. DOI: `10.1007/s12583-010-0160-7`.

Boulding, K. E. (1991). “What is evolutionary economics?” In: *Journal of Evolutionary economics* 1.1, pp. 9–17. DOI: `https://doi.org/10.1007/978-0-585-35155-1_2`.

Bourke, A. F. (2011). *Principles of Social Evolution.* Oxford University Press. ISBN: 9780199231157.

Bowler, P. J. (1983). *The eclipse of Darwinism. anti-Darwinian evolution theories in the decades around 1900*, p. 291. ISBN: 0801829321.

Brakefield, P. M. (2006). “Evo-devo and constraints on selection.” In: *Trends in Ecology & Evolution* 21.7, pp. 362–368. DOI: `10.1016/j.tree.2006.05.001`.

Brännström, Å., J. Johansson, and N. von Festenberg (2013). “The Hitchhiker’s Guide to Adaptive Dynamics.” In: *Games* 4.3, pp. 304–328. DOI: `10.3390/g4030304`.

Breznak, J. A. (1982). “Intestinal Microbiota of Termites and other Xylophagous Insects.” In: *Annual Review of Microbiology* 36.1, pp. 323–323. DOI: `10.1146/annurev.mi.36.100182.001543`.

Brucker, R. M. and S. R. Bordenstein (2013). “The Hologenomic Basis of Speciation: Gut Bacteria Cause Hybrid Lethality in the Genus Nasonia.” In: *Science* 341.6146, pp. 667–669. DOI: `10.1126/science.1240659`.

Burkhardt, R. W. (2013). “Lamarck, Evolution, and the Inheritance of Acquired Characters.” In: *Genetics* 194.4, pp. 793–805. DOI: `10.1534/genetics.113.151852`.

Burroughs, A. M. and L. Aravind (2020). “Identification of Uncharacterized Components of Prokaryotic Immune Systems and Their Diverse Eukaryotic Reformulations.” In: *Journal of Bacteriology* 202.24. Ed. by A. M. Stock. DOI: `10.1128/jb.00365-20`.

Camazine, S. et al. (2020). *Self-organization in biological systems.* Princeton university press.

Carmel, Y. (2023). “Human societal development: is it an evolutionary transition in individuality?” In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 378.1872. DOI: `10.1098/rstb.2021.0409`.

Carrier, T. J. and T. C. G. Bosch (2022). “Symbiosis: the other cells in development.” In: *Development* 149.13. DOI: `10.1242/dev.200797`.

Casagrande, D. (1999). “Information as Verb: Re-conceptualizing Information for Cognitive and Ecological Models.” In: *Journal of Ecological Anthropology* 3.1, pp. 4–13. DOI: `10.5038/2162-4593.3.1.1`.

Castillo-Davis, C. I. (2005). “The evolution of noncoding DNA: how much junk, how much func?” In: *Trends in genetics: TIG* 21.10, pp. 533–536. DOI: `10.1016/j.tig.2005.08.001`.

Chebii, V. J. et al. (2021). “Genomics of Adaptations in Ungulates.” In: *Animals* 11.6, p. 1617. DOI: `10.3390/ani11061617`.

Chen, H. et al. (2015). “The reverse evolution from multicellularity to unicellularity during carcinogenesis.” In: *Nature Communications* 6.1. DOI: `10.1038/ncomms7367`.

Childe, V. G. (1950). “The Urban Revolution.” In: *The Town Planning Review* 21.1, pp. 3–17.

Chisholm, R. H. et al. (2018). “The Role of Pleiotropy in the Evolutionary Maintenance of Positive Niche Construction.” In: *The American Naturalist* 192.1, pp. 35–48. DOI: `10.1086/697471`.

Chiu, L. and S. F. Gilbert (2015). “The Birth of the Holobiont: Multi-species Birthing Through Mutual Scaffolding and Niche Construction.” en. In: *Biosemiotics* 8.2, pp. 191–210. DOI: `10.1007/s12304-015-9232-5`.

Chunikhin, A. Y. (2019). “On Concept of Creative Petri Nets.” In: *CoRR* abs/1910.09326. DOI: `https://doi.org/10.48550/arXiv.1910.09326`.

Cisternino, A. (2021). *Tierra Artificial Life Simulator.* URL: `https://github.com/acisternino/tierra` (visited on 09/30/2021).

Clune, J., J.-B. Mouret, and H. Lipson (2013). “The evolutionary origins of modularity.” In: *Proceedings of the Royal Society B: Biological Sciences* 280.1755, p. 20122863. DOI: `10.1098/rspb.2012.2863`.

Conine, C. C. and O. J. Rando (2021). “Soma-to-germline RNA communication.” In: *Nature Reviews Genetics* 23.2, pp. 73–88. DOI: `10.1038/s41576-021-00412-1`.

Connelly, B. D. et al. (2015). “Negative niche construction favors the evolution of cooperation.” In: *Evolutionary Ecology* 30.2, pp. 267–283. DOI: `10.1007/s10682-015-9803-6`.

Corenblit, D., A. C. W. Baas, et al. (2011). “Feedbacks between geomorphology and biota controlling Earth surface processes and landforms: A review of foundation concepts and current understandings.” In: *Earth-Science Reviews* 106.3-4, pp. 307–331. DOI: `10.1016/j.earscirev.2011.03.002`.

Corenblit, D., J. Steiger, et al. (2007). “Darwinian origin of landforms.” In: *Earth Surface Processes and Landforms* 32.13, pp. 2070–2073. DOI: `10.1002/esp.1536`.

Cottee-Jones, H. E. W. and R. J. Whittaker (2012). “The keystone species concept: a critical appraisal.” In: *Frontiers of Biogeography* 4.3. DOI: `10.21425/f5fbg12533`.

Crick, F. (1970). “Central Dogma of Molecular Biology.” In: *Nature* 227.5258, pp. 561–563. DOI: `10.1038/227561a0`.

Crick, F. H. (1958). “On protein synthesis.” In: *Symp Soc Exp Biol.* Vol. 12. 138-63, p. 8.

Damer, B. and D. Deamer (2020). “The Hot Spring Hypothesis for an Origin of Life.” In: *Astrobiology* 20.4, pp. 429–452. DOI: `10.1089/ast.2019.2045`.

Danchin, E., L.-A. Giraldeau, et al. (2004). "Public Information: From Nosy Neighbors to Cultural Evolution." In: *Science* 305.5683, pp. 487–491. DOI: `10.1126/science.1098254`.

Danchin, E., A. Pocheville, et al. (2018). "Epigenetically facilitated mutational assimilation: epigenetics as a hub within the inclusive evolutionary synthesis." In: *Biological Reviews* 94.1, pp. 259–282. DOI: `10.1111/brv.12453`.

Danchin, É. et al. (2011). "Beyond DNA: integrating inclusive inheritance into an extended theory of evolution." In: *Nature Reviews Genetics* 12.7, pp. 475–486. DOI: `10.1038/nrg3028`.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life.* Murray.

— (1868). "Variation of plants and animals under domestication." In.

Datseris, G., A. R. Vahdati, and T. C. DuBois (2022). "Agents.jl: a performant and feature-full agent-based modeling software of minimal code complexity." In: *SIMULATION*, p. 003754972110688. DOI: `10.1177/00375497211068820`.

Davidson, E. H. (2010). "Emerging properties of animal gene regulatory networks." In: *Nature* 468.7326, pp. 911–920. DOI: `10.1038/nature09645`.

Davidson, E. H. and M. S. Levine (2008). "Properties of developmental gene regulatory networks." In: *Proceedings of the National Academy of Sciences* 105.51, pp. 20063–20066. DOI: `10.1073/pnas.0806007105`.

Dawkins, R. (1976). *The selfish gene.* Oxford University Press. ISBN: 019857519X.

— (2004). "Extended Phenotype – But Not Too Extended. A Reply to Laland, Turner and Jablonka." In: *Biology and Philosophy* 19.3, pp. 377–396. DOI: `10.1023/B:BIPH.0000036180.14904.96`.

— (2016). *Extended Phenotype The Long Reach of the Gene. The Long Reach of the Gene.* Oxford University Press, p. 440. ISBN: 9780198788911.

Day, R. L., K. N. Laland, and F. J. Odling-Smee (2003). "Rethinking Adaptation: The Niche-Construction Perspective." In: *Perspectives in Biology and Medicine* 46.1, pp. 80–95. DOI: `10.1353/pbm.2003.0003`.

Dayrell, R. L. C. et al. (2018). "Ontogenetic shifts in plant ecological strategies." In: *Functional Ecology* 32.12. Ed. by O. Godoy, pp. 2730–2741. DOI: `10.1111/1365-2435.13221`.

de Farias, S. T., M. V. Jose, and F. Prosdocimi (2021). "Is it possible that cells have had more than one origin?" In: *Biosystems* 202, p. 104371. DOI: `https://doi.org/10.1016/j.biosystems.2021.104371`.

Dean, L. G. et al. (2013). "Human cumulative culture: a comparative perspective." In: *Biological Reviews* 89.2, pp. 284–301. DOI: `10.1111/brv.12053`.

Denamur, E. and I. Matic (2006). "Evolution of mutation rates in bacteria." In: *Molecular Microbiology* 60.4, pp. 820–827. DOI: `10.1111/j.1365-2958.2006.05150.x`.

Devresse, A. (2016). URL: `https://github.com/BlueBrain/HighFive`.

Dickins, T. E. and Q. Rahman (2012). "The extended evolutionary synthesis and the role of soft inheritance in evolution." In: *Proceedings of the Royal Society B: Biological Sciences* 279.1740, pp. 2913–2921. DOI: `10.1098/rspb.2012.0273`.

DiFrisco, J. and J. Jaeger (2019). "Beyond networks: mechanism and process in evo-devo." In: *Biology & Philosophy* 34.6. DOI: `10.1007/s10539-019-9716-9`.

Dimond, C. C., R. J. Cabin, and J. S. Brooks (2011). "Feathers, Dinosaurs, and Behavioral Cues: Defining the Visual Display Hypothesis for the Adaptive Function of Feathers in Non-Avian Theropods." In: *BIOS* 82.3, pp. 58–63. DOI: `10.1893/011.082.0302`.

Dobzhansky, T. (1937). *Genetics and the Origin of Species.* Columbia University biological series. Columbia University Press.

Dobzhansky, T. (1973). "Nothing in Biology Makes Sense except in the Light of Evolution." en. In: *The American Biology Teacher* 35.3, pp. 125–129. DOI: `10.2307/4444260`.

Doebeli, M., Y. Ispolatov, and B. Simon (2017). "Towards a mechanistic foundation of evolutionary theory." In: *eLife* 6. DOI: `10.7554/elife.23804`.

Domenico, M. D. et al. (2013). "Mathematical Formulation of Multilayer Networks." In: *Physical Review X* 3.4, p. 041022. DOI: `10.1103/physrevx.3.041022`.

Doncieux, S. et al. (2015). "Evolutionary Robotics: What, Why, and Where to." In: *Frontiers in Robotics and AI* 2. DOI: `10.3389/frobt.2015.00004`.

Doran, N. A. et al. (2006). "Is extinction age dependent?" In: *Palaios* 21.6, pp. 571–579. DOI: `https://doi.org/10.2110/palo.2006.p06-055r`.

Dosi, G. and R. R. Nelson (1994). "An introduction to evolutionary theories in economics." en. In: *Journal of Evolutionary Economics* 4.3, pp. 153–172. DOI: `10.1007/BF01236366`.

Doughty, C. E. et al. (2015). "Global nutrient transport in a world of giants." In: *Proceedings of the National Academy of Sciences* 113.4, pp. 868–873. DOI: `10.1073/pnas.1502549112`.

Douglas, A. E. and J. H. Werren (2016). "Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts." In: *mBio* 7.2. Ed. by M. J. McFall-Ngai and R. J. Collier. DOI: `10.1128/mbio.02099-15`.

Du, Q. et al. (2015). "The Evolution of Aggregative Multicellularity and Cell–Cell Communication in the Dictyostelia." In: *Journal of Molecular Biology* 427.23, pp. 3722–3733. DOI: `10.1016/j.jmb.2015.08.008`.

Ebrahimkhani, M. R. and M. Levin (2021). "Synthetic living machines: A new window on life." In: *iScience* 24.5, p. 102505. DOI: `10.1016/j.isci.2021.102505`.

Eidelson, R. J. (1997). "Complex adaptive systems in the behavioral and social sciences." In: *Review of General Psychology* 1.1, pp. 42–71. DOI: `https://doi.org/10.1037/1089-2680.1.1.4`.

Eigen, M. (1971). "Selforganization of matter and the evolution of biological macromolecules." In: *Die Naturwissenschaften* 58.10, pp. 465–523. DOI: `10.1007/bf00623322`.

Eigen, M. and P. Schuster (1977). "A principle of natural self-organization." In: *Naturwissenschaften* 64.11, pp. 541–565. DOI: `10.1007/bf00450633`.

Elton, C. S. (1927). *Animal ecology, by Charles Elton, with an introduction by Julian S. Huxley.* Macmillan Co.,

Eme, L. et al. (2017). "Archaea and the origin of eukaryotes." In: *Nature Reviews Microbiology* 15.12, pp. 711–723. DOI: `10.1038/nrmicro.2017.133`.

Erwin, D. H. (2008). "Macroevolution of ecosystem engineering, niche construction and diversity." en. In: *Trends in Ecology & Evolution* 23.6, pp. 304–310. DOI: `10.1016/j.tree.2008.01.013`.

Erwin, D. H. and E. H. Davidson (2009). "The evolution of hierarchical gene regulatory networks." In: *Nature Reviews Genetics* 10.2, pp. 141–148. DOI: `10.1038/nrg2499`.

Ester, M. et al. (1996). "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining.* KDD'96. Portland, Oregon: AAAI Press, pp. 226–231. DOI: `10.5555/3001460`.

Evolution, the Merriam-Webster dictionary (2023). Merriam-Webster. URL: `https://www.merriam-webster.com/dictionary/evolution` (visited on 04/28/2023).

Fagundes, N. J. R. et al. (2022). "What We Talk About When We Talk About "Junk DNA"." In: *Genome Biology and Evolution* 14.5. Ed. by L. Katz. DOI: `10.1093/gbe/evac055`.

Farmer, J. D. et al. (2012). "A complex systems approach to constructing better models for managing financial markets and the economy." en. In: *The European Physical Journal Special Topics* 214.1, pp. 295–324. DOI: `10.1140/epjst/e2012-01696-9`.

Feschotte, C. (2008). "Transposable elements and the evolution of regulatory networks." In: *Nature Reviews Genetics* 9.5, pp. 397–405. DOI: `10.1038/nrg2337`.

Fields, C. and M. Levin (2020a). "Does Evolution Have a Target Morphology?" en. In: *Organisms. Journal of Biological Sciences* Vol 4, No 1 (2020): Special Issue: The COVID–19 Epidemic. DOI: `10.13133/2532-5876/16961`.

Fields, C. and M. Levin (2020b). "Scale-Free Biology: Integrating Evolutionary and Developmental Thinking." In: *BioEssays* 42.8, p. 1900228. DOI: `10.1002/bies.201900228`.

Fierro-Constaín, L. et al. (2017). "The Conservation of the Germline Multipotency Program, from Sponges to Vertebrates: A Stepping Stone to Understanding the Somatic and Germline Origins." In: *Genome Biology and Evolution*, evw289. DOI: `10.1093/gbe/evw289`.

Fisher, R. M. et al. (2017). "The evolution of host-symbiont dependence." In: *Nature Communications* 8.1, p. 15973. DOI: `10.1038/ncomms15973`.

Fisher, R. A. (1930). *The genetical theory of natural selection.* Clarendon Press.

Fitzhugh, B. et al. (2019). "Human ecodynamics: A perspective for the study of long-term change in socioecological systems." In: *Journal of Archaeological Science: Reports* 23, pp. 1077–1094. DOI: `10.1016/j.jasrep.2018.03.016`.

Forterre, P. (2006). "The origin of viruses and their possible roles in major evolutionary transitions." In: *Virus Research* 117.1, pp. 5–16. DOI: `10.1016/j.virusres.2006.01.010`.

Forterre, P., J. Filée, and H. Myllykallio (2004). "Origin and Evolution of DNA and DNA Replication Machineries." In: *The Genetic Code and the Origin of Life.* Springer US, pp. 145–168. DOI: `10.1007/0-387-26887-1_10`.

Frenkel-Pinter, M. et al. (2020). "Prebiotic Peptides: Molecular Hubs in the Origin of Life." In: *Chemical Reviews* 120.11, pp. 4707–4765. DOI: `10.1021/acs.chemrev.9b00664`.

Futuyma, D. J. and M. Kirkpatrick (2007). *Evolution.* Oxford University Press, p. 594. ISBN: 9781605356051.

Gabriel, W., M. Lynch, and R. Burger (1993). "Muller's Ratchet and Mutational Meltdowns." In: *Evolution* 47.6, p. 1744. DOI: `10.2307/2410218`.

Galton, F. (1871). "I. Experiments in Pangenesis, by breeding from rabbits of a pure variety, into whose circulation blood taken from other varieties had previously been largely transfused." In: *Proceedings of the Royal Society of London* 19.123-129, pp. 393–410. DOI: `10.1098/rspl.1870.0061`.

Gariépy, J.-F. et al. (2014). "Social learning in humans and other animals." In: *Frontiers in Neuroscience* 8. DOI: `10.3389/fnins.2014.00058`.

Gaskin, T. (2021). *The Utopia Project.* URL: `https://utopia-project.org/index.html` (visited on 07/30/2021).

Gerardo, N. M., K. L. Hoang, and K. S. Stoy (2020). "Evolution of animal immunity in the light of beneficial symbioses." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1808, p. 20190601. DOI: `10.1098/rstb.2019.0601`.

Getling, A. V. (1998). *Rayleigh-Bénard Convection.* WORLD SCIENTIFIC. ISBN: 978-9810226572.

Gibson, A. K., L. F. Delph, and C. M. Lively (2017). “The two-fold cost of sex: Experimental evidence from a natural system.” In: *Evolution Letters* 1.1, pp. 6–15. DOI: `10.1002/evl3.1`.

Gilad, E. et al. (2004). “Ecosystem Engineers: From Pattern Formation to Habitat Creation.” In: *Physical Review Letters* 93.9, p. 098105. DOI: `10.1103/physrevlett.93.098105`.

Gilbert, S. F. (2019). “Evolutionary transitions revisited: Holobiont evo-devo.” en. In: *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 332.8, pp. 307–314. DOI: `10.1002/jez.b.22903`.

Gilbert, S. F., T. C. G. Bosch, and C. Ledón-Rettig (2015). “Eco-Evo-Devo: developmental symbiosis and developmental plasticity as evolutionary agents.” en. In: *Nature Reviews Genetics* 16.10, pp. 611–622. DOI: `10.1038/nrg3982`.

Gillespie, D. T. (1976). “A general method for numerically simulating the stochastic time evolution of coupled chemical reactions.” In: *Journal of Computational Physics* 22.4, pp. 403–434. DOI: `10.1016/0021-9991(76)90041-3`.

Gissis, S. B. et al. (2015). *Transformations of Lamarckism From Subtle Fluids to Molecular Biology. From Subtle Fluids to Molecular Biology.* MIT Press, p. 474. ISBN: 9780262527507.

Godfrey-Smith, P. (2007). “Conditions for Evolution by Natural Selection.” In: *Journal of Philosophy* 104.10, pp. 489–516. DOI: `10.5840/jphil2007104103`.

Goldenfeld, N. and C. Woese (2011). “Life is Physics: Evolution as a Collective Phenomenon Far From Equilibrium.” en. In: *Annu. Rev. Condens. Matter Phys.* 2.1, pp. 375–399. DOI: `10.1146/annurev-conmatphys-062910-140509`.

Goodman, C. S. and B. C. Coughlin (2000). “The evolution of evo-devo biology.” In: *Proceedings of the National Academy of Sciences* 97.9, pp. 4424–4425. DOI: `10.1073/pnas.97.9.4424`.

Gould, S. J. (1992). *The Panda’s Thumb. More Reflections in Natural History.* W. W. Norton & Company, p. 344. ISBN: 9780393308198.

Gould, S. J. and E. S. Vrba (1982). “Exaptation – a Missing Term in the Science of Form.” In: *Paleobiology* 8.1, pp. 4–15. DOI: `10.1017/s0094837300004310`.

Gravel, D., F. Guichard, and M. E. Hochberg (2011). “Species coexistence in a variable world.” In: *Ecology Letters* 14.8, pp. 828–839. DOI: `10.1111/j.1461-0248.2011.01643.x`.

Grosberg, R. K. and R. R. Strathmann (1998). “One cell, two cell, red cell, blue cell: The persistence of a unicellular stage in multicellular life histories.” In: *Trends in Ecology & Evolution* 13.3, pp. 112–116. DOI: `10.1016/s0169-5347(97)01313-x`.

Grunert, K. et al. (2021). “Evolutionarily stable strategies in stable and periodically fluctuating populations: The Rosenzweig–MacArthur predator–prey model.” In: *Proceedings of the National Academy of Sciences* 118.4. DOI: `10.1073/pnas.2017463118`.

Hagen, O. et al. (2017). “Estimating Age-Dependent Extinction: Contrasting Evidence from Fossils and Phylogenies.” In: *Systematic Biology* 67.3. Ed. by M. Alfaro, pp. 458–474. DOI: `10.1093/sysbio/syx082`.

Haken, H. (2008). “Self-organization.” In: *Scholarpedia* 3.8, p. 1401. DOI: `10.4249/scholarpedia.1401`.

Haldane, J. B. S. (1932). *The causes of evolution.* London: Longmans, Green and co. ISBN: 9789333359009.

Hall, B. K. and B. Hallgrímsson (2013). *Strickberger’s Evolution.* Jones & Bartlett Learning, LLC, p. 646. ISBN: 9781449691929.

Hamilton, W. D. (1964). “The genetical evolution of social behaviour. I.” In: *Journal of Theoretical Biology* 7.1, pp. 1–16. DOI: `10.1016/0022-5193(64)90038-4`.

Hammarlund, S. P. et al. (2016). “The evolution of cooperation by the Hankshaw effect.” In: *Evolution* 70.6, pp. 1376–1385. DOI: `10.1111/evo.12928`.

Hammerstein, P. and E. Hagen (2005). “The second wave of evolutionary economics in biology.” In: *Trends in Ecology & Evolution* 20.11, pp. 604–609. DOI: `10.1016/j.tree.2005.07.012`.

Handel, A. E. and S. V. Ramagopalan (2010). “Is Lamarckian evolution relevant to medicine?” In: *BMC Medical Genetics* 11.1, p. 73. DOI: `10.1186/1471-2350-11-73`.

Hardin, G. (1960). “The Competitive Exclusion Principle.” In: *Science* 131.3409, pp. 1292–1297. DOI: `10.1126/science.131.3409.1292`.

Hardy, G. H. (1908). “Mendelian Proportions in a Mixed Population.” In: *Science* 28.706, pp. 49–50. DOI: `10.1126/science.28.706.49`.

Hedrick, P. W. and S. T. Kalinowski (2000). “Inbreeding Depression in Conservation Biology.” In: *Annual Review of Ecology and Systematics* 31.1, pp. 139–162. DOI: `10.1146/annurev.ecolsys.31.1.139`.

Helbing, D. (2012). “Agent-Based Modeling.” In: *Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior*. Ed. by D. Helbing. Springer Berlin Heidelberg, pp. 25–70. ISBN: 978-3-642-24004-1. DOI: `10.1007/978-3-642-24004-1_2`.

Herdeanu, B. (2021). “Emergence of Cooperation in Evolutionary Social Interaction Networks.” PhD thesis. Heidelberg University.

Herrmann, H.-J. (2006). “Pattern Formation of Dunes.” In: *Nonlinear Dynamics* 44.1-4, pp. 315–317. DOI: `10.1007/s11071-006-2016-3`.

Hodgkin, A. L. and A. F. Huxley (1952). “A quantitative description of membrane current and its application to conduction and excitation in nerve.” In: *The Journal of Physiology* 117.4, pp. 500–544. DOI: `10.1113/jphysiol.1952.sp004764`.

Holland, J. H. (1992). “Complex Adaptive Systems.” In: *Daedalus* 121.1, pp. 17–30.

— (2006). “Studying Complex Adaptive Systems.” In: *Journal of Systems Science and Complexity* 19.1, pp. 1–8. DOI: `10.1007/s11424-006-0001-z`.

Holovatch, Y., R. Kenna, and S. Thurner (2017). “Complex systems: physics beyond physics.” In: *European Journal of Physics* 38.2, p. 023002. DOI: `10.1088/1361-6404/aa5a87`.

Holterhoff, K. (2014). “The History and Reception of Charles Darwin’s Hypothesis of Pangenesis.” In: *Journal of the History of Biology* 47.4, pp. 661–695. DOI: `10.1007/s10739-014-9377-0`.

Hordijk, W. and M. Steel (2018). “Autocatalytic Networks at the Basis of Life’s Origin and Organization.” In: *Life* 8.4, p. 62. DOI: `10.3390/life8040062`.

Howe, J. et al. (2022). “Multicellularity in animals: The potential for within-organism conflict.” In: *Proceedings of the National Academy of Sciences* 119.32. DOI: `10.1073/pnas.2120457119`.

Hoyer, S. and J. Hamman (2017). “xarray: N-D labeled arrays and datasets in Python.” In: *Journal of Open Research Software* 5.1. DOI: `10.5334/jors.148`.

Hunter, J. D. (2007). “Matplotlib: A 2D graphics environment.” In: *Computing in Science & Engineering* 9.3, pp. 90–95. DOI: `10.1109/MCSE.2007.55`.

Hutchinson, G. E. (1957). “Concluding Remarks.” In: *Cold Spring Harbor Symposia on Quantitative Biology* 22.0, pp. 415–427. DOI: `10.1101/sqb.1957.022.01.039`.

Hutchinson, G. E. (1965). *The ecological theater and the evolutionary play.* Yale UNIVERSITY PRESS. ISBN: 978-0300005868.

Huxley, J. (1942). *Evolution: The Modern Synthesis.* London: George Alien & Unwin Ltd.

Ille, A. M., H. Lamont, and M. B. Mathews (2022). "The Central Dogma revisited: Insights from protein synthesis, CRISPR, and beyond." In: *WIREs RNA* 13.5. DOI: `10.1002/wrna.1718`.

Isaeva, V. V. (2012). "Self-organization in biological systems." en. In: *Biology Bulletin* 39.2, pp. 110–118. DOI: `10.1134/S1062359012020069`.

Jablonka, E. (2002). "Information: Its Interpretation, Its Inheritance, and Its Sharing." In: *Philosophy of Science* 69.4, pp. 578–605. DOI: `10.1086/344621`.

— (2017). "The evolutionary implications of epigenetic inheritance." In: *Interface Focus* 7.5, p. 20160135. DOI: `10.1098/rsfs.2016.0135`.

Jablonka, E. and M. J. Lamb (2008). "Soft inheritance: challenging the modern synthesis." In: *Genetics and Molecular Biology* 31.2, pp. 389–395. DOI: `10.1590/s1415-47572008000300001`.

Jones, C. G., J. H. Lawton, and M. Shachak (1994). "Organisms as ecosystem engineers." In: *Ecosystem management.* Springer, pp. 130–147.

— (1997). "POSITIVE AND NEGATIVE EFFECTS OF ORGANISMS AS PHYSICAL ECOSYSTEM ENGINEERS." In: *Ecology* 78.7, pp. 1946–1957. DOI: `https://doi.org/10.1890/0012-9658(1997)078[1946:PANEOO]2.0.CO;2`.

Kandori, M. (1992). "Social Norms and Community Enforcement." In: *The Review of Economic Studies* 59.1, p. 63. DOI: `10.2307/2297925`.

Karreth, F. A., Y. Tay, and P. Pandolfi (2014). "Target competition: transcription factors enter the limelight." In: *Genome Biology* 15.4, p. 114. DOI: `10.1186/gb4174`.

Kauffman, S. A. (1992). "Origins of Order in Evolution: Self-Organization and Selection." en. In: *Understanding Origins: Contemporary Views on the Origin of Life, Mind and Society.* Ed. by F. J. Varela and J.-P. Dupuy. Boston Studies in the Philosophy and History of Science. Dordrecht: Springer Netherlands, pp. 153–181. ISBN: 978-94-015-8054-0. DOI: `10.1007/978-94-015-8054-0_8`.

Kays, R. and A. Allison (2001). "Arboreal tropical forest vertebrates: current knowledge and research trends." In: *Tropical Forest Canopies: Ecology and Management: Proceedings of ESF Conference, Oxford University, 12–16 December 1998.* Ed. by K. E. Linsenmair et al. Dordrecht: Springer Netherlands, pp. 109–120. ISBN: 978-94-017-3606-0. DOI: `10.1007/978-94-017-3606-0_9`.

Kazil, J., D. Masad, and A. Crooks (2020). "Utilizing Python for Agent-Based Modeling: The Mesa Framework." In: *Social, Cultural, and Behavioral Modeling.* Ed. by R. Thomson et al. Cham: Springer International Publishing, pp. 308–317. ISBN: 978-3-030-61254-2.

Kelly, B. P., A. Whiteley, and D. Tallmon (2010). "The Arctic melting pot." In: *Nature* 468.7326, pp. 891–891. DOI: `10.1038/468891a`.

Kermack, W. O., A. G. McKendrick, and G. T. Walker (1927). "A contribution to the mathematical theory of epidemics." In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772, pp. 700–721.

Keynes, M. and T. M. Cox (2008). "William Bateson, the rediscoverer of Mendel." In: *Journal of the Royal Society of Medicine* 101.3, pp. 104–104. DOI: `10.1258/jrsm.2008.081011`.

Kimura, M. (1968). "Evolutionary Rate at the Molecular Level." en. In: *Nature* 217.5129, pp. 624–626. DOI: `10.1038/217624a0`.

— (1983). *The neutral theory of molecular evolution*. Cambridge University Press. ISBN: 978-0-521-23109-1.

Klug, W. S. (2012). *Concepts of genetics*. Pearson Education. ISBN: 9780321724120.

Kondepudi, D. K., B. D. Bari, and J. A. Dixon (2020). "Dissipative Structures, Organisms and Evolution." In: *Entropy* 22.11, p. 1305. DOI: `10.3390/e22111305`.

Koonin, E. V. (2012). "Does the central dogma still stand?" In: *Biology Direct* 7.1, p. 27. DOI: `10.1186/1745-6150-7-27`.

— (2015). "Why the Central Dogma: on the nature of the great biological exclusion principle." In: *Biology Direct* 10.1. DOI: `10.1186/s13062-015-0084-3`.

Koonin, E. V., K. S. Makarova, and L. Aravind (2001). "Horizontal Gene Transfer in Prokaryotes: Quantification and Classification." In: *Annual Review of Microbiology* 55.1, pp. 709–742. DOI: `10.1146/annurev.micro.55.1.709`.

Koonin, E. V. and A. S. Novozhilov (2017). "Origin and Evolution of the Universal Genetic Code." In: *Annual Review of Genetics* 51.1, pp. 45–62. DOI: `10.1146/annurev-genet-120116-024713`.

Koranne, S. (2011). "Hierarchical data format 5: HDF5." In: *Handbook of Open Source Tools*. Springer, pp. 191–200.

Krakauer, D. C., K. M. Page, and D. H. Erwin (2009). "Diversity, Dilemmas, and Monopolies of Niche Construction." In: *The American Naturalist* 173.1, pp. 26–40. DOI: `10.1086/593707`.

Kronfeldner, M. E. (2006). "Is cultural evolution Lamarckian?" In: *Biology & Philosophy* 22.4, pp. 493–512. DOI: `10.1007/s10539-006-9037-7`.

Kruger, K. et al. (1982). "Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena." In: *Cell* 31.1, pp. 147–157. DOI: `10.1016/0092-8674(82)90414-7`.

Kühnlein, A., S. A. Lanzmich, and D. Braun (2021). "tRNA sequences can assemble into a replicator." In: *eLife* 10, e63431. DOI: `10.7554/elife.63431`.

Kun, Á., M. Santos, and E. Szathmáry (2005). "Real ribozymes suggest a relaxed error threshold." In: *Nature Genetics* 37.9, pp. 1008–1011. DOI: `10.1038/ng1621`.

Laland, K. N., F. J. Odling-Smee, and M. W. Feldman (1996). "The evolutionary consequences of niche construction: a theoretical investigation using two-locus theory." In: *Journal of Evolutionary Biology* 9.3, pp. 293–316. DOI: `https://doi.org/10.1046/j.1420-9101.1996.9030293.x`.

— (1999). "Evolutionary consequences of niche construction and their implications for ecology." In: *Proceedings of the National Academy of Sciences* 96.18, pp. 10242–10247. DOI: `10.1073/pnas.96.18.10242`.

Laland, K., B. Matthews, and M. W. Feldman (2016). "An introduction to niche construction theory." en. In: *Evolutionary Ecology* 30.2, pp. 191–202. DOI: `10.1007/s10682-016-9821-z`.

Laland, K., T. Uller, et al. (2014). "Does evolutionary theory need a rethink?" In: *Nature* 514.7521, pp. 161–164. DOI: `10.1038/514161a`.

Laland, K. N. and M. J. O'Brien (2011). "Cultural Niche Construction: An Introduction." In: *Biological Theory* 6.3, pp. 191–202. DOI: `10.1007/s13752-012-0026-6`.

Laland, K. N., J. Odling-Smee, and S. F. Gilbert (2008). "EvoDevo and niche construction: building bridges." In: *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 310B.7, pp. 549–566. DOI: `10.1002/jez.b.21232`.

Laland, K. N., J. Odling-Smee, and S. Myles (2010). "How culture shaped the human genome: bringing genetics and the human sciences together." In: *Nature Reviews Genetics* 11.2, pp. 137–148. DOI: `10.1038/nrg2734`.

Laland, K. N., T. Uller, et al. (2015). "The extended evolutionary synthesis: its structure, assumptions and predictions." In: *Proceedings of the Royal Society B: Biological Sciences* 282.1813, p. 20151019. DOI: `10.1098/rspb.2015.1019`.

Lamarck, J.-B. (1830). *Philosophie zoologique.* Paris: Duminil-Lesueur.

Landmann, F. et al. (2014). "Co-evolution between an Endosymbiont and Its Nematode Host: Wolbachia Asymmetric Posterior Localization and AP Polarity Establishment." In: *PLoS Neglected Tropical Diseases* 8.8, e3096. DOI: `10.1371/journal.pntd.0003096`.

Langerhans, R. B. and T. J. DeWitt (2002). "Plasticity constrained: over-generalized induction cues cause maladaptive phenotypes." In: *Evolutionary Ecology Research* 4.6, pp. 857–870.

Laubichler, M. D. and J. Renn (2015). "Extended evolution: A conceptual framework for integrating regulatory networks and niche construction." en. In: *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 324.7, pp. 565–577. DOI: `10.1002/jez.b.22631`.

Lehmann, L. (2008). "The adaptive dynamics of niche constructing traits in spatially subdivided populations: evolving posthumous extended phenotypes." In: *Evolution* 62.3, pp. 549–566. DOI: `10.1111/j.1558-5646.2007.00291.x`.

Leibold, M. A. (1995). "The Niche Concept Revisited: Mechanistic Models and Community Context." In: *Ecology* 76.5, pp. 1371–1382. DOI: `10.2307/1938141`.

Lenski, R. E., C. Ofria, et al. (2003). "The evolutionary origin of complex features." In: *Nature* 423.6936, pp. 139–144. DOI: `10.1038/nature01568`.

Lenski, R. E., M. J. Wiser, et al. (2015). "Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with Escherichia coli." In: *Proceedings of the Royal Society B: Biological Sciences* 282.1821, p. 20152292. DOI: `10.1098/rspb.2015.2292`.

Levin, S. A. (2002). "Complex adaptive systems: Exploring the known, the unknown and the unknowable." en. In: *Bulletin of the American Mathematical Society* 40.01, pp. 3–20. DOI: `10.1090/S0273-0979-02-00965-5`.

Levine, J. M. et al. (2017). "Beyond pairwise mechanisms of species coexistence in complex communities." In: *Nature* 546.7656, pp. 56–64. DOI: `10.1038/nature22898`.

Lewontin, R. C. (1970). "The units of selection." In: *Annual review of ecology and systematics* 1.1, pp. 1–18.

Lind, M. I. and F. Spagopoulou (2018). "Evolutionary consequences of epigenetic inheritance." In: *Heredity* 121.3, pp. 205–209. DOI: `10.1038/s41437-018-0113-y`.

Liu, Y. and Q. Chen (2018). "150 years of Darwin's theory of intercellular flow of hereditary information." In: *Nature Reviews Molecular Cell Biology* 19.12, pp. 749–750. DOI: `10.1038/s41580-018-0072-4`.

Llorens, M., J. Oliver, et al. (2004). "Structural and dynamic changes in concurrent systems: reconfigurable Petri nets." In: *IEEE Transactions on Computers* 53.9, pp. 1147–1158. DOI: `10.1109/TC.2004.66`.

Llorens, M. and J. Oliver (2004). "Introducing Structural Dynamic Changes in Petri Nets: Marked-Controlled Reconfigurable Nets." In: *Automated Technology for Verification and Analysis.* Ed. by F. Wang. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 310–

323. ISBN: 978-3-540-30476-0. DOI: `https://doi.org/10.1007/978-3-540-30476-0_26`.

López-García, P., L. Eme, and D. Moreira (2017). "Symbiosis in eukaryotic evolution." In: *Journal of Theoretical Biology* 434, pp. 20–33. DOI: `10.1016/j.jtbi.2017.02.031`.

Lorberbaum, D. S. and S. Barolo (2013). "Gene Regulation: When Analog Beats Digital." In: *Current Biology* 23.23, R1054–R1056. DOI: `https://doi.org/10.1016/j.cub.2013.10.004`.

Lotka, A. J. (1910). "Contribution to the Theory of Periodic Reactions." In: *The Journal of Physical Chemistry* 14.3, pp. 271–274. DOI: `10.1021/j150111a004`.

— (1920). "Analytical Note on Certain Rhythmic Relations in Organic Systems." In: *Proceedings of the National Academy of Sciences* 6.7, pp. 410–415. DOI: `10.1073/pnas.6.7.410`.

Louis, P.-Y. and F. R. Nardi, eds. (2018). *Probabilistic Cellular Automata*. Springer International Publishing. DOI: `10.1007/978-3-319-65558-1`.

Luisi, P. L. (2019). *Emergence of Life From Chemical Origins to Synthetic Biology. From Chemical Origins to Synthetic Biology*. Cambridge University Press, p. 478. ISBN: 9781108735506.

Lynch, M., R. Bürger, et al. (1993). "The Mutational Meltdown in Asexual Populations." In: *Journal of Heredity* 84.5, pp. 339–344. DOI: `10.1093/oxfordjournals.jhered.a111354`.

Lynch, M., J. Conery, and R. Bürger (1995). "MUTATIONAL MELTDOWNS IN SEXUAL POPULATIONS." In: *Evolution* 49.6, pp. 1067–1080. DOI: `10.1111/j.1558-5646.1995.tb04434.x`.

Ma, T. and Y. Nakamori (2005). "Agent-based modeling on technological innovation as an evolutionary process." en. In: *European Journal of Operational Research* 166.3, pp. 741–755. DOI: `10.1016/j.ejor.2004.01.055`.

MacKinnon, J. R. and K. S. MacKinnon (1980). "Niche Differentiation in a Primate Community." In: *Malayan Forest Primates: Ten Years' Study in Tropical Rain Forest*. Ed. by D. J. Chivers. Boston, MA: Springer US, pp. 167–190. ISBN: 978-1-4757-0878-3.

Majhi, S., M. Perc, and D. Ghosh (2022). "Dynamics on higher-order networks: a review." In: *Journal of The Royal Society Interface* 19.188. DOI: `10.1098/rsif.2022.0043`.

Maraun, M. et al. (2019). "Parthenogenetic *vs.* sexual reproduction in oribatid mite communities." In: *Ecology and Evolution* 9.12, pp. 7324–7332. DOI: `10.1002/ece3.5303`.

Maynard Smith, J. (1978). *The evolution of sex*. Vol. 4. Cambridge University Press Cambridge. ISBN: 978-0-521-29302-0.

Mayr, E. (1999). *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. Emersion: Emergent Village Resources for Communities of Faith Series. Harvard University Press. ISBN: 9780674862500.

Mayr, E. (1972). "Lamarck revisited." In: *Journal of the History of Biology*, pp. 55–94.

— (1976). "Species Concepts and Definitions." In: *Topics in the Philosophy of Biology*. Springer Netherlands, pp. 353–371. DOI: `10.1007/978-94-010-1829-6_16`.

— (2000a). "Darwin's Influence on Modern Thought." In: *Scientific American* 283.1, pp. 78–83.

— (2000b). "The Biological Species Concept." In: *Species Concepts and Phylogenetic Theory: A Debate*. Ed. by Q. D. Wheeler and R. Meier. Columbia University Press, pp. 17–29.

McFall-Ngai, M. J. and E. G. Ruby (1991). “Symbiont Recognition and Subsequent Morphogenesis as Early Events in an Animal-Bacterial Mutualism.” In: *Science* 254.5037, pp. 1491–1494. DOI: `10.1126/science.1962208`.

Meaning of evolution in English, Cambridge Dictionary (2023). Cambridge Dictionary. URL: `https://dictionary.cambridge.org/dictionary/english/evolution` (visited on 04/28/2023).

Mendel, G. (1866). “Versuche über Pflanzenhybriden.” In: *Verhandlungen des naturforschenden Vereines in Brunn 4: 3* 44.

Mengistu, H. et al. (2016). “The Evolutionary Origins of Hierarchy.” In: *PLOS Computational Biology* 12.6. Ed. by O. Sporns, e1004829. DOI: `10.1371/journal.pcbi.1004829`.

Mesoudi, A. (2011). *Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences.* Chicago ; London: University of Chicago Press.

— (2015). “Cultural Evolution: A Review of Theory, Findings and Controversies.” In: *Evolutionary Biology* 43.4, pp. 481–497. DOI: `10.1007/s11692-015-9320-0`.

Michod, R. E. (2003). “On the Reorganization of Fitness During Evolutionary Transitions in Individuality.” In: *Integrative and Comparative Biology* 43.1, pp. 64–73. DOI: `10.1093/icb/43.1.64`.

Michod, R. E. and D. Roze (2001). “Cooperation and conflict in the evolution of multicellularity.” In: *Heredity* 86.1, pp. 1–7. DOI: `10.1046/j.1365-2540.2001.00808.x`.

Millstein, R. L. (2022). “Evolution.” In: *The Stanford Encyclopedia of Philosophy.* Ed. by E. N. Zalta. Spring 2022. Metaphysics Research Lab, Stanford University.

Minkina, O. and C. P. Hunter (2018). “Intergenerational Transmission of Gene Regulatory Information in Caenorhabditis elegans.” In: *Trends in Genetics* 34.1, pp. 54–64. DOI: `10.1016/j.tig.2017.09.012`.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms (Complex Adaptive Systems).* The MIT Press, p. 221. ISBN: 9780262631853.

Moran, P. A. P. (1958). “Random processes in genetics.” In: *Mathematical Proceedings of the Cambridge Philosophical Society* 54.1, pp. 60–71. DOI: `10.1017/S0305004100033193`.

Muller, H. (1964). “The relation of recombination to mutational advance.” In: *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 1.1, pp. 2–9. DOI: `https://doi.org/10.1016/0027-5107(64)90047-8`.

Müller, G. B. (2007). “Evo–devo: extending the evolutionary synthesis.” en. In: *Nature Reviews Genetics* 8.12, pp. 943–949. DOI: `10.1038/nrg2219`.

— (2017a). “Why an extended evolutionary synthesis is necessary.” en. In: *Interface Focus* 7.5, p. 20170015. DOI: `10.1098/rsfs.2017.0015`.

— (2017b). “Why an extended evolutionary synthesis is necessary.” en. In: *Interface Focus* 7.5, p. 20170015. DOI: `10.1098/rsfs.2017.0015`.

Mullon, C. and L. Lehmann (2018). “Eco-Evolutionary Dynamics in Metacommunities: Ecological Inheritance, Helping within Species, and Harming between Species.” In: *The American Naturalist* 192.6, pp. 664–686. DOI: `10.1086/700094`.

Nakazawa, T. (2014). “Ontogenetic niche shifts matter in community ecology: a review and future perspectives.” In: *Population Ecology* 57.2, pp. 347–354. DOI: `10.1007/s10144-014-0448-z`.

Neveu, M., H.-J. Kim, and S. A. Benner (2013). “The Strong RNA World Hypothesis: Fifty Years Old.” In: *Astrobiology* 13.4, pp. 391–403. DOI: `10.1089/ast.2012.0868`.

Ng, E. T. H. and A. R. Kinjo (2022). “Computational modelling of plasticity-led evolution.” In: *Biophysical Reviews* 14.6, pp. 1359–1367. DOI: `10.1007/s12551-022-01018-5`.

Niklas, K. J. (2013). “The evolutionary-developmental origins of multicellularity.” In: *American Journal of Botany* 101.1, pp. 6–25. DOI: `10.3732/ajb.1300314`.

NikNaks, W. C. U. (2014). *Complex adaptive system.svg.* URL: `https://en.wikipedia.org/wiki/File:Complex_adaptive_system.svg` (visited on 04/28/2023).

Nilsson, E. E., M. B. Maamar, and M. K. Skinner (2020). “Environmentally Induced Epigenetic Transgenerational Inheritance and the Weismann Barrier: The Dawn of Neo-Lamarckian Theory.” In: *Journal of Developmental Biology* 8.4, p. 28. DOI: `10.3390/jdb8040028`.

Norell, M. A. and X. Xu (2005). “FEATHERED DINOSAURS.” In: *Annual Review of Earth and Planetary Sciences* 33.1, pp. 277–299. DOI: `10.1146/annurev.earth.33.092203.122511`.

Nowak, M. A. (2006). *Evolutionary Dynamics: Exploring the Equations of Life.* The Belknap press of Harvard University press. ISBN: 978-0674023383.

Odlin-Smee, F. J., K. N. Laland, and M. W. Feldman (2003). *Niche Construction: The Neglected Process in Evolution (MPB-37).* Princeton University Press. ISBN: 0691044376.

Odling-Smee, F. J., K. N. Laland, and M. W. Feldman (2003). *Niche construction: the neglected process in evolution.* 37. Princeton University Press. ISBN: 9780691044378.

Ohno, S. (2013). *Evolution by gene duplication.* Springer Science & Business Media. ISBN: 978-3-642-86661-6.

Okasha, S. (2006). *Evolution and the Levels of Selection.* Oxford University Press. DOI: `10.1093/acprof:oso/9780199267972.001.0001`.

Okazaki, R. et al. (1968). “Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains.” In: *Proceedings of the National Academy of Sciences* 59.2, pp. 598–605. DOI: `10.1073/pnas.59.2.598`.

Oughton, E. J. et al. (2018). “Infrastructure as a Complex Adaptive System.” In: *Complexity* 2018, pp. 1–11. DOI: `10.1155/2018/3427826`.

Paaby, A. B. and M. V. Rockman (2014). “Cryptic genetic variation: evolution’s hidden substrate.” In: *Nature Reviews Genetics* 15.4, pp. 247–258. DOI: `10.1038/nrg3688`.

Paine, R. T. (1969). “A Note on Trophic Complexity and Community Stability.” In: *The American Naturalist* 103.929, pp. 91–93. DOI: `10.1086/282586`.

Papale, F. (2020). “Evolution by means of natural selection without reproduction: revamping Lewontin’s account.” In: *Synthese* 198.11, pp. 10429–10455. DOI: `10.1007/s11229-020-02729-6`.

Parrish, J. K. and L. Edelstein-Keshet (1999). “Complexity, Pattern, and Evolutionary Trade-Offs in Animal Aggregation.” In: *Science* 284.5411, pp. 99–101. DOI: `10.1126/science.284.5411.99`.

Pattee, H. H. (2007). “The Necessity Of Biosemiotics: Matter-Symbol Complementarity.” In: *Introduction to Biosemiotics.* Springer Netherlands, pp. 115–132. DOI: `10.1007/1-4020-4814-9_4`.

— (2012a). “Cell Psychology: An Evolutionary Approach to the Symbol-Matter Problem.” In: *Biosemiotics.* Springer Netherlands, pp. 165–179. DOI: `10.1007/978-94-007-5161-3_11`.

— (2012b). “Evolving Self-reference: Matter, Symbols, and Semantic Closure.” In: *Biosemiotics.* Springer Netherlands, pp. 211–226. DOI: `10.1007/978-94-007-5161-3_14`.

Pelletier, F., D. Garant, and A. P. Hendry (2009). “Eco-evolutionary dynamics.” In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1523, pp. 1483–1489. DOI: `10.1098/rstb.2009.0027`.

Peralta, A. F., J. Kertész, and G. Iñiguez (2022). "Opinion dynamics in social networks: From models to data." In: DOI: `10.48550/ARXIV.2201.01322`.

Peter, I. S. and E. H. Davidson (2017). "Assessing regulatory information in developmental gene regulatory networks." In: *Proceedings of the National Academy of Sciences* 114.23, pp. 5862–5869. DOI: `10.1073/pnas.1610616114`.

Petri, C. A. (1962). "Kommunikation mit Automaten." PhD thesis. Technical University of Darmstadt.

Pfennig, D. W. (2021). *Phenotypic Plasticity & Evolution.* CRC Press. DOI: `10.1201/9780429343001`.

Pigliucci, M. (2009). "An Extended Synthesis for Evolutionary Biology." In: *Annals of the New York Academy of Sciences* 1168.1, pp. 218–228. DOI: `10.1111/j.1749-6632.2009.04578.x`.

Plaisance, L. et al. (2011). "The Diversity of Coral Reefs: What Are We Missing?" In: *PLOS ONE* 6.10, pp. 1–7. DOI: `10.1371/journal.pone.0025026`.

Pocheville, A. (2014). "The Ecological Niche: History and Recent Controversies." In: *Handbook of Evolutionary Thinking in the Sciences.* Springer Netherlands, pp. 547–586. DOI: `10.1007/978-94-017-9014-7_26`.

Prusiner, S. B. (1998). "Prions." In: *Proceedings of the National Academy of Sciences* 95.23, pp. 13363–13383. DOI: `10.1073/pnas.95.23.13363`.

Raff, R. A. (2000). "Evo-devo: the evolution of a new discipline." In: *Nature Reviews Genetics* 1.1, pp. 74–79. DOI: `10.1038/35049594`.

Rafiqi, A. M., A. Rajakumar, and E. Abouheif (2020). "Origin and elaboration of a major evolutionary transition in individuality." In: *Nature* 585.7824, pp. 239–244. DOI: `10.1038/s41586-020-2653-6`.

Railsback, S. F. and V. Grimm (2019). *Agent-Based and Individual-Based Modeling A Practical Introduction, Second Edition. A Practical Introduction, Second Edition.* Princeton University Press, p. 360. ISBN: 9780691190839.

Rainey, P. B. (2023). "Major evolutionary transitions in individuality between humans and AI." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 378.1872. DOI: `10.1098/rstb.2021.0408`.

Reisig, W. (2013). *Understanding petri nets: modeling techniques, analysis methods, case studies.* Springer. ISBN: 978-3-642-33278-4.

Rich, E. (2008). *Automata, computability and complexity. theory and applications.* Pearson Prentice Hall, p. 1099. ISBN: 9780132288064.

Ridley, M. (2004). *Evolution.* Blackwell Pub., p. 751. ISBN: 1405103450.

Riedel, L. et al. (2020). "Utopia: A Comprehensive and Collaborative Modeling Framework for Complex and Evolving Systems." In: *Journal of Open Source Software* 5.53, p. 2165. DOI: `10.21105/joss.02165`.

Rocha, L. M. (2001). "Evolution with material symbol systems." In: *Biosystems* 60.1-3, pp. 95–121. DOI: `10.1016/s0303-2647(01)00110-1`.

Rocklin, M. (2015). "Dask: Parallel computation with blocked algorithms and task scheduling." In: *Proceedings of the 14th python in science conference.* 130-136. DOI: `10.25080/Majora-7b98e3ed-013`.

Ross, A. A., A. Rodrigues Hoffmann, and J. D. Neufeld (2019). "The skin microbiome of vertebrates." In: *Microbiome* 7.1, p. 79. DOI: `10.1186/s40168-019-0694-6`.

Rosselló-Mora, R. (2001). "The species concept for prokaryotes." In: *FEMS Microbiology Reviews* 25.1, pp. 39–67. DOI: `10.1016/s0168-6445(00)00040-1`.

Rousseeuw, P. J. (1987). “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis.” In: *Journal of Computational and Applied Mathematics* 20, pp. 53–65. DOI: `10.1016/0377-0427(87)90125-7`.

Ruiz, M. (2022). *DNA replication.* URL: `https://commons.wikimedia.org/wiki/File:DNA_replication_de.svg` (visited on 07/17/2023).

Sardanyés, J., S. F. Elena, and R. V. Solé (2008). “Simple quasispecies models for the survival-of-the-flattest effect: The role of space.” In: *Journal of Theoretical Biology* 250.3, pp. 560–568. DOI: `10.1016/j.jtbi.2007.10.027`.

Scheiner, S. M. (2014). “Bet-hedging as a complex interaction among developmental instability, environmental heterogeneity, dispersal, and life-history strategy.” In: *Ecology and Evolution* 4.4, pp. 505–515. DOI: `10.1002/ece3.951`.

Scheiner, S. M., M. Barfield, and R. D. Holt (2021a). “Do I build or do I move? Adaptation by habitat construction versus habitat choice.” In: *Evolution* 76.3, pp. 414–428. DOI: `10.1111/evo.14355`.

— (2021b). “The evolution of habitat construction with and without phenotypic plasticity.” In: *Evolution* 75.7, pp. 1650–1664. DOI: `10.1111/evo.14226`.

— (2022). “The factors that favor adaptive habitat construction versus non-adaptive environmental conditioning.” In: *Ecology and Evolution* 12.3. DOI: `10.1002/ece3.8763`.

Schoener, T. W. (2011). “The Newest Synthesis: Understanding the Interplay of Evolutionary and Ecological Dynamics.” In: *Science* 331.6016, pp. 426–429. DOI: `10.1126/science.1193954`.

Schönfisch, B. and A. de Roos (1999). “Synchronous and asynchronous updating in cellular automata.” en. In: *Biosystems* 51.3, pp. 123–143. DOI: `10.1016/S0303-2647(99)00025-8`.

Schroeder, K., S. K. Lyons, and F. A. Smith (2021). “The influence of juvenile dinosaurs on community structure and diversity.” In: *Science* 371.6532, pp. 941–944. DOI: `10.1126/science.abd9220`.

Scott, W. G. (2007). “Ribozymes.” In: *Current Opinion in Structural Biology* 17.3, pp. 280–286. DOI: `10.1016/j.sbi.2007.05.003`.

Scott-Phillips, T. C. et al. (2014). “The Niche Construction Perspective: A Critical Appraisal.” en. In: *Evolution* 68.5, pp. 1231–1243. DOI: `10.1111/evo.12332`.

Ségurel, L. and C. Bon (2017). “On the Evolution of Lactase Persistence in Humans.” In: *Annual Review of Genomics and Human Genetics* 18.1, pp. 297–319. DOI: `10.1146/annurev-genom-091416-035340`.

Seong, H. J., S.-W. Han, and W. J. Sul (2021). “Prokaryotic DNA methylation and its functional roles.” In: *Journal of Microbiology* 59.3, pp. 242–248. DOI: `10.1007/s12275-021-0674-y`.

Sevinchan, Y. (2018). personal communication.

— (2019). personal communication.

— (2021). “Evolution Mechanics and Perspectives on Food Web Ecology.” PhD thesis.

Sevinchan, Y., B. Herdeanu, H. Mack, et al. (2020). “Boosting Group-Level Synergies by Using a Shared Modeling Framework.” In: *Lecture Notes in Computer Science*. Ed. by V. V. Krzhizhanovskaya et al. Springer International Publishing, pp. 442–456. ISBN: 978-3-030-50436-6. DOI: `10.1007/978-3-030-50436-6_32`.

Sevinchan, Y., B. Herdeanu, and J. Traub (2020). “dantro: a Python package for handling, transforming, and visualizing hierarchically structured data.” en. In: *Journal of Open Source Software* 5.52, p. 2316. DOI: `10.21105/joss.02316`.

Shannon, C. E. (1948). "A Mathematical Theory of Communication." In: *Bell System Technical Journal* 27.3, pp. 379–423. DOI: `10.1002/j.1538-7305.1948.tb01338.x`.

Shimamura, M. et al. (1997). "Molecular evidence from retroposons that whales form a clade within even-toed ungulates." In: *Nature* 388.6643, pp. 666–670. DOI: `10.1038/41759`.

Siek, J., L. Q. Lee, and A. Lumsdaine (2002). *The Boost Graph Library: User Guide and Reference Manual*. Addison-Wesley. ISBN: 978-0201729146.

Siemers, B. M. and H.-U. Schnitzler (2004). "Echolocation signals reflect niche differentiation in five sympatric congeneric bat species." In: *Nature* 429.6992, pp. 657–661. DOI: `10.1038/nature02547`.

Silver, M. and E. D. Paolo (2006). "Spatial effects favour the evolution of niche construction." In: *Theoretical Population Biology* 70.4, pp. 387–400. DOI: `10.1016/j.tpb.2006.08.003`.

Skinner, M. K. and E. E. Nilsson (2021). "Role of environmentally induced epigenetic transgenerational inheritance in evolutionary biology: Unified Evolution Theory." In: *Environmental Epigenetics* 7.1. DOI: `10.1093/eep/dvab012`.

Smith, E. (2023). "Beyond fitness: the nature of selection acting through the constructive steps of lifecycles." In: *Evolution*, qpad068. DOI: `10.1093/evolut/qpad068`.

Smolin, L. (1992). "Did the Universe evolve?" In: *Classical and Quantum Gravity* 9.1, pp. 173–191. DOI: `10.1088/0264-9381/9/1/016`.

Snustad, D. P. and M. J. Simmons (2016). *Principles of Genetics*. Wiley, p. 648. ISBN: 9781119142287.

Sommer, R. J. (2020). "Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges." In: *Genetics* 215.1, pp. 1–13. DOI: `10.1534/genetics.120.303163`.

Staps, M., J. van Gestel, and C. E. Tarnita (2019). "Emergence of diverse life cycles and life histories at the origin of multicellularity." In: *Nature Ecology & Evolution* 3.8, pp. 1197–1205. DOI: `10.1038/s41559-019-0940-0`.

Stearns, S. C. and R. F. Hoekstra (2005). *Evolution, an introduction*. Oxford University Press. ISBN: 978-0199255634.

Steffen, W., P. J. Crutzen, and J. R. McNeill (2007). "The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature." en. In: *AMBIO: A Journal of the Human Environment* 36.8, pp. 614–621. DOI: `10.1579/0044-7447(2007)36[614:TAAHNO]2.0.CO;2`.

Steiner, U. K. (2021). "Senescence in Bacteria and Its Underlying Mechanisms." In: *Frontiers in Cell and Developmental Biology* 9. DOI: `10.3389/fcell.2021.668915`.

Stotz, K. (2017). "Why developmental niche construction is not selective niche construction: and why it matters." In: *Interface Focus* 7.5, p. 20160157. DOI: `10.1098/rsfs.2016.0157`.

Strogatz, S. H. (2001). "Exploring complex networks." en. In: *Nature* 410.6825, pp. 268–276. DOI: `10.1038/35065725`.

— (2018). *Nonlinear Dynamics and Chaos with Student Solutions Manual: Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and ... Edition (Studies in Nonlinearity) (Volume 1). With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, p. 531. ISBN: 9780813349107.

Stumpo, M. et al. (2021). "Self-Organization through the Inner Heliosphere: Insights from Parker Solar Probe." In: *Atmosphere* 12.3, p. 321. DOI: `10.3390/atmos12030321`.

Sultan, S. E. (2015). *Organism and Environment.* Oxford University Press. DOI: `10.1093/acprof:oso/9780199587070.001.0001`.

— (2021). “Eco-Evo-Devo.” In: *Evolutionary Developmental Biology.* Springer International Publishing, pp. 1165–1177. DOI: `10.1007/978-3-319-32979-6_42`.

Sumida, S. S. and C. A. Brochu (2000). “Phylogenetic Context for the Origin of Feathers.” In: *American Zoologist* 40.4, pp. 486–503. DOI: `10.1093/icb/40.4.486`.

Sun, D. et al. (2019). “Editorial: Horizontal Gene Transfer Mediated Bacterial Antibiotic Resistance.” In: *Frontiers in Microbiology* 10. DOI: `10.3389/fmicb.2019.01933`.

Szathmáry, E. (2015). “Toward major evolutionary transitions theory 2.0.” In: *Proceedings of the National Academy of Sciences* 112.33, pp. 10104–10111. DOI: `10.1073/pnas.1421398112`.

Szathmáry, E. and J. Maynard Smith (1995). “The major evolutionary transitions.” In: *Nature* 374.6519, pp. 227–232. DOI: `10.1038/374227a0`.

Taylor, J. W. et al. (2000). “Phylogenetic Species Recognition and Species Concepts in Fungi.” In: *Fungal Genetics and Biology* 31.1, pp. 21–32. DOI: `10.1006/fgbi.2000.1228`.

Taylor, T. (2004). “Niche Construction and the Evolution of Complexity.” In: *Proceedings of Artificial Life Ix.* MIT Press, pp. 375–380. DOI: `https://doi.org/10.7551/mitpress/1429.003.0063`.

The HDF Group (1997-2022). *Hierarchical Data Format, version 5.* URL: `https://www.hdfgroup.org/HDF5/`.

Thurner, S., P. Klimek, and R. Hanel (2018). *Introduction to the Theory of Complex Systems.* Oxford University Press. DOI: `10.1093/oso/9780198821939.001.0001`.

Torday, J. (2016). “The Cell as the First Niche Construction.” In: *Biology* 5.2, p. 19. DOI: `10.3390/biology5020019`.

Tovar-Moll, F. and R. Lent (2016). “The various forms of neuroplasticity: Biological bases of learning and teaching.” In: *PROSPECTS* 46.2, pp. 199–213. DOI: `10.1007/s11125-017-9388-7`.

Trappes, R. (2021). “Defining the niche for niche construction: evolutionary and ecological niches.” In: *Biology & Philosophy* 36.3. DOI: `10.1007/s10539-021-09805-2`.

Trigos, A. S. et al. (2018). “How the evolution of multicellularity set the stage for cancer.” In: *British Journal of Cancer* 118.2, pp. 145–152. DOI: `10.1038/bjc.2017.398`.

Turing, A. M. (1952). “The chemical basis of morphogenesis.” In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 237.641, pp. 37–72. DOI: `10.1098/rstb.1952.0012`.

Turner, J. S. (2002). *The Extended Organism. The Physiology of Animal-Built Structures.* Harvard University Press, p. 256. ISBN: 9780674009851.

Unwin, D. M. (1998). “Feathers, filaments and theropod dinosaurs.” In: *Nature* 391.6663, pp. 119–120. DOI: `10.1038/34279`.

van Kampen, N. G. (2007). *Stochastic Processes in Physics and Chemistry, Third Edition.* North Holland, p. 464. ISBN: 9780444529657.

van Valen, L. (1973). “A new evolutionary law.” In: *Evol theory* 1, pp. 1–30. DOI: `https://doi.org/10.7208/9780226115504-022`.

Venner, S., C. Feschotte, and C. Biémont (2009). “Dynamics of transposable elements: towards a community ecology of the genome.” In: *Trends in Genetics* 25.7, pp. 317–323. DOI: `10.1016/j.tig.2009.05.003`.

Vetsigian, K., C. Woese, and N. Goldenfeld (2006). “Collective evolution and the genetic code.” In: *Proceedings of the National Academy of Sciences* 103.28, pp. 10696–10701. DOI: `10.1073/pnas.0603780103`.

Volterra, V. (1927). *Variazioni e fluttuazioni del numero d’individui in specie animali conviventi.* Memorie della Classe di Scienze Fisiche, Matematiche e Naturali. Societá anonima tipografica ”Leonardo da Vinci”.

von Neumann, J. (1966). *Theory of Self-Reproducing Automata.* Ed. by A. W. Burks. University of Illionois Press. ISBN: 978-0252727337.

Vrba, E. S. and N. Eldredge (1984). “Individuals, hierarchies and processes: towards a more complete evolutionary theory.” In: *Paleobiology* 10.2, pp. 146–171. DOI: `10.1017/s0094837300008149`.

Waddingtion, C. H. (1942). “CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS.” In: *Nature* 150.3811, pp. 563–565. DOI: `10.1038/150563a0`.

Wagner, A. (1994). “Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization.” In: *Proceedings of the National Academy of Sciences* 91.10, pp. 4387–4391. DOI: `10.1073/pnas.91.10.4387`.

Wagner, A. (1996). “DOES EVOLUTIONARY PLASTICITY EVOLVE?” en. In: *Evolution* 50.3, pp. 1008–1023. DOI: `10.1111/j.1558-5646.1996.tb02342.x`.

Wang, Y., H. Liu, and Z. Sun (2017). “Lamarck rises from his grave: parental environment-induced epigenetic inheritance in model organisms and humans.” In: *Biological Reviews* 92.4, pp. 2084–2111. DOI: `10.1111/brv.12322`.

Watson, J. D. and F. H. C. Crick (1953). “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.” en. In: *Nature* 171.4356, pp. 737–738. DOI: `10.1038/171737a0`.

Watson, R. A., M. Levin, and C. L. Buckley (2022). “Design for an Individual: Connectionist Approaches to the Evolutionary Transitions in Individuality.” In: *Frontiers in Ecology and Evolution* 10. DOI: `10.3389/fevo.2022.823588`.

Watson, R. A. and E. Szathmáry (2016). “How Can Evolution Learn?” en. In: *Trends in Ecology & Evolution* 31.2, pp. 147–157. DOI: `10.1016/j.tree.2015.11.009`.

Watson, R. A., G. P. Wagner, et al. (2014). “THE EVOLUTION OF PHENOTYPIC CORRELATIONS AND “DEVELOPMENTAL MEMORY”.” In: *Evolution* 68.4, pp. 1124–1138. DOI: `10.1111/evo.12337`.

Weinberg, W. (1908). “Über den nachweis der vererbung beim menschen.” In: *Jahresheft des Vereins für Vaterländische Naturkunde in Württemberg* 64, pp. 369–382.

Weismann, A. (1892). *Das Keimplasma: eine Theorie der Vererbung.* Fischer.

Werner, M. (2008). *Information und Codierung. Grundlagen und Anwendungen.* Vieweg, p. 213. ISBN: 9783834802323.

West, S. A. et al. (2015). “Major evolutionary transitions in individuality.” In: *Proceedings of the National Academy of Sciences* 112.33, pp. 10112–10119. DOI: `10.1073/pnas.1421402112`.

Westra, E. R. et al. (2012). “The CRISPRs, They Are A-Changin’: How Prokaryotes Generate Adaptive Immunity.” In: *Annual Review of Genetics* 46.1, pp. 311–339. DOI: `10.1146/annurev-genet-110711-155447`.

Whitehead, H. (2017). “Gene–culture coevolution in whales and dolphins.” In: *Proceedings of the National Academy of Sciences* 114.30, pp. 7814–7821. DOI: `10.1073/pnas.1620736114`.

Whitehead, H. et al. (2019). “The reach of gene–culture coevolution in animals.” In: *Nature Communications* 10.1. DOI: `10.1038/s41467-019-10293-y`.

Whiten, A. (2017). “A second inheritance system: the extension of biology through culture.” In: *Interface Focus* 7.5, p. 20160142. DOI: `10.1098/rsfs.2016.0142`.

— (2019). “Cultural Evolution in Animals.” In: *Annual Review of Ecology, Evolution, and Systematics* 50.1, pp. 27–48. DOI: `10.1146/annurev-ecolsys-110218-025040`.

Wilecki, M. et al. (2015). “Predatory feeding behaviour in Pristionchus nematodes is dependent on a phenotypic plasticity and induced by serotonin.” In: *Journal of Experimental Biology*. DOI: `10.1242/jeb.118620`.

Wilensky, U. (1999). *NetLogo*. Tech. rep. Northwestern University, Evanston, IL: Center for Connected Learning and Computer-Based Modeling.

Williams, G. C. and R. Dawkins (2018). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton Science Library. Princeton University Press. ISBN: 9780691185507.

Williams, H. and R. F. Lachlan (2021). “Evidence for cumulative cultural evolution in bird song.” In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 377.1843. DOI: `10.1098/rstb.2020.0322`.

Williams, H., I. I. Levin, et al. (2013). “Three decades of cultural evolution in Savannah sparrow songs.” In: *Animal Behaviour* 85.1, pp. 213–223. DOI: `10.1016/j.anbehav.2012.10.028`.

Wintersberger, E. et al. (2019). *h5cpp*. DOI: `10.5281/zenodo.3573739`. URL: `https://github.com/ess-dmsc/h5cpp/tree/v0.3.3`.

Wiser, M. J., N. Ribeck, and R. E. Lenski (2013). “Long-Term Dynamics of Adaptation in Asexual Populations.” In: *Science* 342.6164, pp. 1364–1367. DOI: `10.1126/science.1243357`.

Witt, U. (2008). “What is specific about evolutionary economics?” In: *Journal of Evolutionary Economics* 18.5, pp. 547–575. DOI: `10.1007/s00191-008-0107-7`.

Woese, C. R. (2002). “On the evolution of cells.” In: *Proceedings of the National Academy of Sciences* 99.13, pp. 8742–8747. DOI: `10.1073/pnas.132266999`.

Wolf, Y. I., M. I. Katsnelson, and E. V. Koonin (2018). “Physical foundations of biological complexity.” In: *Proceedings of the National Academy of Sciences* 115.37. DOI: `10.1073/pnas.1807890115`.

Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media Incorporated, p. 1197. ISBN: 9781579550080.

Wright, S. (1931). “Evolution in Mendelian Populations.” In: *Genetics* 16.2, pp. 97–159.

Young, H. P. (2015). “The Evolution of Social Norms.” In: *Annual Review of Economics* 7.1, pp. 359–387. DOI: `10.1146/annurev-economics-080614-115322`.

Zachar, I. and G. Boza (2020). “Endosymbiosis before eukaryotes: mitochondrial establishment in protoeukaryotes.” In: *Cellular and Molecular Life Sciences* 77.18, pp. 3503–3523. DOI: `10.1007/s00018-020-03462-6`.

Zaitsev, D. and Z. Li (2018). “On simulating Turing machines with inhibitor Petri nets.” en. In: 13.1, pp. 147–156. DOI: `10.1002/tee.22508`.

Zaitsev, D. (2012). “Petri Net Paradigm of Computation.” In: p. 7.

Zhang, J. (2003). “Evolution by gene duplication: an update.” In: *Trends in Ecology & Evolution* 18.6, pp. 292–298. DOI: `10.1016/s0169-5347(03)00033-8`.

Zheng, J., J. L. Payne, and A. Wagner (2019). “Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks.” In: *Science* 365.6451, pp. 347–353. DOI: `10.1126/science.aax1837`.

Zhu, H., G. Wang, and J. Qian (2016). “Transcription factors as readers and effectors of DNA methylation.” In: *Nature Reviews Genetics* 17.9, pp. 551–565. DOI: `10.1038/nrg.2016.83`.

Zuppinger-Dingley, D. et al. (2014). “Selection for niche differentiation in plant communities increases biodiversity effects.” In: *Nature* 515.7525, pp. 108–111. DOI: `10.1038/nature13869`.