Evolvability: a Formal Approach

Alexis Gallagher
St Peter's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2009

Acknowledgements

The gift which has allowed me to complete this work is the support of other people.

I would like to thank my thesis supervisor Alan Grafen. His criticism has taught me the rare and delicate art of making speculative arguments without merely speculating. His reminders to scrupulously cite supporting evidence have kept me honest. And I would also like to thank him for always holding the faith that my work was going somewhere, even in the days when there was little evidence to cite supporting that claim.

I would like to thank Adam Wilkins, Johannes (Yogi) Jaagu, and Richard Watson for feedback and vital encouragement. A man in the wilderness is starved for company, even if his manners have grown wild. I would like to thank my undergraduate mathematics professor, Benedict Gross, for teaching me proof, and my physics professor, Hitoshi Yamamoto, for teaching me theory.

I would like to thank Mark Rendel for invaluable conversations and Jotun Hein for welcoming me and Mark into his group while Alan was on sabbatical.

I would like to thank the Bodossakis Foundation for providing the financial support that made this work possible and that permitted me to join friendly St Peter's College.

I would like to thank the Maydays for keeping me sane by embracing the insane.

I would like to thank my parents for encouraging me and supporting me. Parents are marvellous. It is wonderful that everyone gets a pair but I've been especially lucky with mine.

And last I would like to thank my wife Ringae Nuek, who endured so many inconveniences for her husband's mad pursuit – living in a cinder block, spending hours in a punishing daily commute to London, and patiently tolerating both my glooms and my grandiosities.

Evolvability: a Formal Approach
D.Phil, Trinity 2009
St Peters College
Alexis Gallagher

ABSTRACT

This dissertation clarifies the concept of evolvability, the increased capacity of some organisms or systems to support evolution, especially the evolution of life-like complexity. I survey the literature, which is spread over the fields of population genetics, developmental biology, artificial life, and microbial and molecular evolution. Finding that researchers have often used the term vaguely and incompatibly I identify five distinct kinds or senses of evolvability. I also identify five key constituent ideas, which I discuss in the context of *organismic evolvability*, a sense of evolvability with deep roots in the traditional fields of animal development and macroevolution. In these fields research into evolvability has historically been hampered by an insufficiently detailed knowledge of development. Research in molecular evolution has produced a thorough knowledge of the folding of RNA into secondary structure, which can be regarded as a model of development. This has motivated new approaches to evolvability based on representing development via a single genotype-phenotype mapping function. I build on these approaches to invent new mathematical methods to formalise the traditional ideas. I create an exact model illustrating a classic example of evolvability, the capacity for repeated segmentation and simple modularity. I analyse this with two new formal approaches. First is the *genospace algebra*, a propositional calculus based on graph theory. It is a formal language for describing genotype-phenotype maps. It provides a system for making calculations, proofs, and diagrams about mutational structures in genotype space, and it is flexible enough to allow description at arbitrary degrees of resolution. Second is a pair of concepts, the *genetic leverage* and the *genetic fulcrum*. The leverage provides a crude numerical measure of evolvability, and the fulcrum provides a heuristic for identifying the genomic and developmental causes of evolvability. Besides its specific relevance to diversification and development, evolvability is also crucial to the fundamental question of how evolution produces ordinary biological life. Simulation systems that implement only a conventional textbook model of evolution – systems possessing only variation, inheritance, and selection – fail to evolve anything resembling the complexity of the biological world. Research into evolvability is our best bet to illuminate the "missing ingredient" for life-like evolution.

Contents

Preface

A question with half an answer

An animal – any animal – is a miraculous thing. Composed of many parts fitted together in an exquisite harmony, it eats, grows, reproduces, moves, and acts with every appearance of purpose. In all these regards an animal stands out from the non-living world of stuff. Although every organism is really a chemical machine at the microscopic level, the living world is so far beyond any man-made mechanism that it appears, still, quite different in kind.

So it is natural to wonder what created the living world, what explains it. This is the fundamental question of biology, so basic it is easily taken for granted. What explains the living world? That is, what explains the unique complexity that distinguishes the living world?

And the fundamental answer is evolution by means of natural selection. This process has transformed our planet, over billions of years, from an inanimate rock into a teeming greenhouse. Seen at the fundamental level, as a logical process, it requires only three conditions: variation, selection, and inheritance. This is our most basic explanation for the existence of the natural world.

However, a couple problems make this explanation alone unsatisfactory.

First, it omits details. This explanation describes a process of microscopic change but does not say how those changes add up to macroscopic changes. No doubt they do, but how? If the evolution of every complex organism is a unique story, then there is nothing to say beyond cataloguing these stories in all their particularity. However, it seems plausible that these stories share themes, that they fit into categories – in short, that there are principles that describe the evolution of complexity in general. What are they?

Macroevolutionary patterns suggest that other forces besides natural selection play an important role. Some organisms seem to survive longer, to diversify more widely, or to develop over the long run according to an internal logic. What are the features of an organism that facilitate its evolution? How much are these features also an explanation of complexity?

A second problem is that it is easy to create artificial systems that possess the conditions for evolution – variation, selection, and inheritance – but then this evolution does not produce anything like the complexity of the biological world. This shows that these conditions are necessary but not sufficient for life as we know it. Biological evolution depends on some extra condition that we do not understand but, presumably, that we could understand. This claim is either painfully obvious or bizarrely radical, depending on what kind of specialist you ask. In disciplines where it is not experimentally feasible to create artificial systems, the ideas those systems illustrate can come to seem unthinkable.

All these problems point to the fledgling idea of evolvability, the capacity to evolve. A theory of evolvability would speak to a variety of questions. What are the conditions that promote evolution – especially evolution of the sort of complexity taken for granted in the biological world? Is it something special about certain organisms or their developmental systems? About certain substances, like carbon or water? What can we say about the evolution of complexity, in detail and in general? If systems vary in their evolvability, then why is that? How do we measure it? How do we predict it?

These questions have been studied for decades in a variety of disciplines, but usually in an ad hoc and informal manner and without much communication between researchers. This dissertation reviews past work on evolvability, sorts out different senses of the term, and then develops a mathematical framework for precisely modelling the most promising ideas.

All of this, it is hoped, will shed light on the question of evolvability. This question is certainly ambiguous and hard to specify. It is also, however, fundamental and important and worth tackling head-on.

The history of "evolvability"

The literature on evolvability is quite fragmented so before discussing the idea it is worth outlining the varied usage of the term.

The term "evolvability" is first used by Dawkins (1989) in a short paper presenting his Biomorphs model. This model illustrates how an organism's developmental system might increase its ability to evolve, how this ability itself could evolve, and how this might affect the organism's evolutionary history.

But this term merely brought focus to a cluster of ideas with an older and more diffuse history. For instance, Maynard-Smith et al. (1985) present a lucid review of these issues under the rubric of "developmental constraints". Even earlier, Riedl (1978) presented similar ideas in the *Order of Living Organisms* (discussed by Wagner and Laubichler (2004)). Conrad (1979) discussed "amenability to evolution" more narrowly in regards to chemical evolution, and Rechenberg (1973) promptly encountered related issues in the first work in genetic algorithms.

In the 1990s the term starts appearing too frequently to be traced in detail.[1] However, one can identify a distinct school of researchers all now using the term in roughly the original sense as they study the same cluster of issues. The key work in this area is the work of Gunter Wagner, Marc Kirschner, and John Gerhart. There have also been significant contributions by Yang, von Dassow, Poole, and others. The sense of evolvability which is studied in that work I will refer to as *organismic evolvability*, because it is about kinds of organisms. Chapter 1 will explain that concept by walking through its key ideas and the main publications of these authors.

During this period other researchers start using the term in new senses which, while related, are narrower. These researchers typically stand outside of evolutionary biology and they modify the idea as they translate it to different disciplines – population genetics, microbial biology, or evolutionary computation. These new senses often emerge from efforts to clarify the original idea so it can be applied to experiments or exact models. Chapter 2 will review this work and distinguish these alternative senses of evolvability.

[1] As of 4 February 2008 searches on the biological and medical references services PubMed and Biological Abstracts give over 400 and 200 hits respectively. The multi-disciplinary service Google Scholar gives over 8,000 hits, with 2,200 in biology and life sciences, and over 3,600 in computer science and other engineering disciplines. (Engineering disciplines may be overrepresented because of fuller computerisation of their literature, and because of an unrelated usage of "evolvability" in systems engineering referring to tolerance for changing technology and usage scenarios (e.g., Rowe et al., 1998).)

I call them *trait evolvability* (the most prominent), *individual fitness evolvability*, *individual mutational evolvability*, and *substrate evolvability*.

The word also falls into vogue for a few years, resulting in a scattering of publications which use the term casually and without a technical meaning (Cowell et al., 1999; Matsuura et al., 2002; Suzuki et al., 2003; Woodruff, 2001; Yamauchi et al., 2002). I do not review this work.

Plan of this dissertation

Chapter 1: Organismic evolvability

Chapter 1 reviews the literature on organismic evolvability, the original sense of the term, and the one with the longest research history and most immediate explanatory promise. The earliest work began with research on the body plan, the effect of development on the body plan, and the consequent effects of developmental bias on biological diversity. This work provides the concrete sense of evolvability that underlies nearly all later intuitions. Much of this work connects with older and more mainstream discussions about the evolution of complex features (e.g., the "problem" of half a wing, etc.).

Using Dawkins's model of biomorphs as a simplified illustration, this chapter describes five key ideas that have organised discussions of evolvability: the structure of variability, the developmental architecture, the effects on macroevolutionary patterns, the incompatibility with Darwinian selection, and the effects on the evolution of complexity. A central idea in work on organismic evolvability is the idea of a *core component*, a stable part of an organism's developmental architecture which modifies the expression of less stable parts in a way that increases the organism's phenotypic variability and biases it towards complexity. This chapter also describes the ways in which these discussions fit into the traditional research literature.

(The technical parts of this thesis – chapters 4 through 6 – are dedicated to trying to clarify organismic evolvability by expressing it in simple, heuristic formal models.)

Chapter 2: Taxonomy of evolvabilities

Organismic evolvability is about organisms and is rooted in the study of how their developmental systems shape their evolution. Chapter 2 names and distinguishes four other notions

of evolvability which have emerged independently or as a result of efforts to formalise the original notion of organismic evolvability.

These are the notions of trait evolvability, substrate evolvability, individual fitness evolvability, and individual mutational evolvability. These notions come from research in population genetics, molecular evolution, microbial evolution, and evolutionary computation. Many discussions of evolvability have been thought-provoking but also somewhat vague or mutually contradictory. Clarifying the distinctions and relations between these various notions of evolvability is essential to evaluating them.

In addition, considering these other notions of evolvability and the research around them provides valuable insights for how to analyse organismic evolvability. Research using the notion of individual fitness evolvability has shown clearly how the evolution of evolvability is compatible with Darwinian selection, and sheds light on the key idea of variability. The notion of substrate evolvability, which underlies much work in evolutionary computation, highlights the connection between evolvability and deeper questions about the fundamental conditions necessary for evolution.

But most of all, research on evolvability in the molecular evolution of RNA shapes has stimulated the development of a new formal methods for analysing the genotype-phenotype map. This work is the foundation for the modelling approaches that I will develop in the following chapters.

Chapter 3: The topological phenotype

Chapter 3 recapitulates past efforts to use the mathematical construct of a topological space in order to represent and analyse the genotype-phenotype mapping. These efforts rely on a chain of intermediate formal constructs – from the mapping, to the accessibility measure, to the accessibility relation, to the set of neighbourhoods, and finally to the topological space. The topological space describes the mutational relationships between phenotypes, as implied by the mapping.

However, in reviewing these efforts, it is shown that the notion of a topological space has been applied incorrectly or only as a kind of inspiration. Applied strictly, a topological space destroys so much information that it renders very different biological systems identical, making it a poor basis for models. Furthermore, the intermediate construct of the *accessibility measure* does not accurately measure the probability it is supposed to measure and it conceals

structural information which might be crucial to evolvability.

However, aspects of this approach seem very worthwhile. It develops other worthwhile ideas like the *phenotypic pre-image*, and the accessibility measures and relations are themselves powerful aides in thinking clearly about mutational relationships. This chapter offers a new accessibility measure, κ', which corrects the probability calculation. And considering the structural blind spots in the measure leads to a clearly definable contra-indication of evolvability, the idea of a *dead end genotype*, which is a genotype lacking the mutational possibilities of its phenotypic peers.

Chapter 4: Evolvable segmentation and the genospace algebra

Chapter 4 presents a model and a new kind of analysis. The model illustrates one of the most frequently discussed mechanisms supporting evolvability in the biological literature focused on development: the capacity for repetition of a phenotypic feature, such as a body segment. It compares two genotype-phenotype maps of differing evolvability, a naive mapping ϕ which expresses genes directly and an evolvable mapping ϕ_E where a regulatory mechanism provides an abstracted control over the number of body segments.

Analysis shows that the capacity for segmentation in the evolvable mapping eliminates all dead end genotypes. Analysis also shows that the evolvable mapping always allows certain direct mutational transitions between phenotypes, transitions which are not available in the naive mapping. Both of these results are consistent with the basic idea of organismic evolvability – the idea that the mapping with a capacity for abstracted control is indeed more evolvable.

To perform the analysis this chapter introduces a new mathematical formalism, the *genospace algebra*, a propositional calculus based on graph theory. It provides a system for making calculations, proofs, and diagrams about mutational structures in genotype space. This formalism is flexible enough to allow description at arbitrary degrees of resolution. It is a formal language for describing genotype-phenotype maps.

Chapter 5: Genetic leverage

Genetic leverage is another formal tool for measuring the evolvability of a genotype-phenotype mapping, cruder than the genospace algebra but simpler to apply. Roughly, it measures how much certain parts of the genome enjoy an amplified influence over phenotypic change,

under certain genetic configurations. This kind of amplification is indicative of a high evolvability. The idea of genetic leverage works alongside the idea of the *genetic fulcrum*, a formalisation of the genetic configuration that enables this amplification. Chapter 5 discusses how to define the *natural* genetic fulcrum for a given genotype-phenotype mapping, and how to study a mapping by trying different fulcra and interpreting the results.

These constructs all formalise the basic idea that evolvability is enabled by the core features of an organism's developmental system, which remains relatively constant even as it influences the nature of evolutionary change outside of the developmental system.

Chapter 6: Evolvable modularity

Chapter 4 presented a model of evolvability produced through the abstracted control of the number of repetitions of a single phenotypic feature, a body segment. The model in chapter 6 incorporates abstracted control of which type of feature is expressed and of the ordering of the features in the body. This represents a richer kind of modularity than mere repetition.

The genetic leverage is calculated, confirming that the evolvable mapping is indeed more evolvable.

This model is only slightly more complex than the segmentation model. However, even this slight change introduces new facets to the genotype-phenotype mapping: module arrangement, gene discrimination, and regulatory addressing. These new facets enable new modes of abstracted control, by allowing new forms of abstraction beyond the simple abstraction of repetition number in the segmentation model. This illustrates the idea of abstracted control more fully.

In this way, the model suggests how evolvability can support the evolution of biologically realistic phenotypic modularity, just as the facets themselves better resemble biological development systems, such as hox gene systems.

Formal methods in this dissertation

At certain points this dissertation invents new mathematical constructs to analyse purely illustrative models. This is somewhat unusual and recalls a comment by Francis Crick :

> Elegance and a deep simplicity, often expressed in an abstract mathematical form, are useful guides in physics, but in biology such intellectual tools can be very misleading. For this reason a theorist in biology has to receive much more guidance from the experimental evidence.... (Crick, 1988, p 6)

This warning is quite sensible. Abstract mathematics has a certain rhetorical effect, and it can glorify what are merely more esoteric forms of confusion. Given this, it may be worth explaining how it was in fact a commitment to experimental evidence that led this work in such a mathematical direction.

Work on organismic evolvability springs from a long tradition of empirical research on development and macroevolutionary change. This tradition dates from the earliest efforts in paleontology, taxonomy, and morphology, starting with the first work on the body plan. But as development and genetics have been black boxes for years, this research could only describe an organism's phenotype (which is silent regarding the underlying genetics) and document its history (which is unavoidably unique and contingent). This limited basis of observation left open so many interpretations that arguments for evolvability have always remained a bit speculative, or vague to the point of being unfalsifiable.

But that situation is changing. Recent work on molecular and computational systems has produced much richer, finer-grained data relevant to evolvability, and it has spawned new methods for analysing these data. Also, the rapid progress in evolutionary developmental biology promises such data will become available for the original, traditional system of the organism itself.

In short, conceptual advances have made it possible and empirical advances will make it necessary to update older ideas on evolvability in order to accommodate a more detailed understanding of development. These ideas deserve to be integrated into the mainstream of thinking on development and evolutionary history. A basic first step is simply to spell these ideas out more clearly. This is where mathematics is useful, as this dissertation tries to frame these old ideas without ambiguity. If these ideas are wrong, a formal approach will at least make the error explicit. If these ideas are right, only a formal approach will let us measure how right they are.

Work on evolvability has been around long enough that there are interesting hypotheses leading in every direction. Only greater clarity can show the way forward.

Chapter 1

Organismic evolvability

This chapter describes the literature and key ideas behind the concept of organismic evolvability: evolvability as a property of an organism that facilitates its evolution, especially the evolution of complexity. Organism here means not a single individual but a natural kind, and it can be approximately translated as a species.

The following captures the essence of the idea, as it is loosely held in common by a variety of researchers:

Definition: organismic evolvability

> Evolvability is an organism's propensity to generate useful phenotypic variation from random genotypic variation – especially phenotypic variation which leads to complexity.
>
> As it is the organism's developmental system that translates genotypes to phenotypes, evolvability is fundamentally a property and a result of that developmental system.[1]

What reason is there to believe this concept is important? What does it explain? Evolvability is suggested by three lines of argument.

First, it is suggested by empirical patterns in the macroevolutionary record. Certain organisms seem unusually abundant or persistent – for instance, beetles. There are more species in the order Coleoptera than in any other animal or plant group (Farrell, 1998). This might be because they are especially fit, or because this order is fragmenting in parallel

[1] This definition is intended broadly, so that it includes behaviour under the rubric of phenotype.

with its ecological niche. Or it might be because beetles are especially evolvable, because they are intrinsically more predisposed to generate new variants, new species.

Second, evolvability is suggested by mechanical intuition. An organism is a kind of machine and this suggests that certain organisms, like certain machines, are more easily tweaked, varied, and re-arranged. The various segments of a beetle seem so neatly self-contained that it seems like one could modify a single segment without disrupting the others, just as replacing an old car's cassette player with a CD player does not also require upgrading the engine. The design of some animals, and features of their developmental systems, suggest this very strongly.

Last, evolvability is suggested by some speculative thinking regarding the nature of biological complexity. Suppose we accept that certain organisms are more evolvable because they are more easily varied and re-arranged. The speculation is this: perhaps this same kind of ready re-arrangeability might also facilitate the evolution of a complex organism. While complexity is hard to define, it certainly seems to involve the coordination of semi-independent parts. This might be the same kind of semi-independence that makes the parts re-arrangable. In this way, the idea of evolvability could help explain the evolution of complexity. It does not contradict the conventional account of incremental change by natural selection, but it adds a layer of conceptual detail. This is either an important addition, or a modest clarification, depending on how satisfactory one finds the conventional account.

In other words, theories of evolvability are essentially treating two very old questions: what are the origins of animal form (i.e., phenotypic complexity), and what produces the broad patterns of evolutionary history. While these questions are old, what is new in the idea of evolvability is a few features of the intellectual approach – the effort to consider development and the phenotype in mechanistic terms, and the interpretation of that mechanistic detail from a more abstract or systems-oriented perspective. The hope is that this perspective will expose an overall property of the developmental system, evolvability, which sheds light on these two questions.

Given the traditional focus of the questions, it is no surprise that organismic evolvability has attracted researchers working in the traditional studies of development, anatomy, and the historical evolutionary record, as well as their modern offshoots such as evolutionary developmental biology (or evo-devo). All these areas also tend to treat the organism as a whole – as a single, integrated, functioning physical being. This "whole organism"

perspective encourages an interest in complexity as a subject in itself, how it develops in the individual, how it changes through time, and the connections between these issues. The idea of evolvability brings these themes together. Let us now review this idea in detail.

1.1 Biomorphs: an illustration

We will review organismic evolvability by introducing five key ideas that surround the concept. These are as follows:

1. the structure of variability,
2. the role of the developmental architecture,
3. the traces of evolvability in the macroevolutionary record,
4. the role of evolvability in phenotypic complexity, and
5. the concern that evolvability is unDarwinian

All of these key ideas may be seen in one of the earliest discussions of evolvability, Dawkins's biomorphs model. Therefore we will briefly review this model, using it as a skeleton to introduce the ideas in a simplified context. Afterwards, we will discuss the ideas in more depth, relating them to key publications in the research literature on evolvability.

Biomorphs are simple stick figures vaguely resembling animals (introduced in Dawkins (1986), discussed with respect to evolvability in Dawkins (1989)). The figures are drawn by a recursive computer algorithm, which is seeded with a small set of integers (re-implemented in section 1.4). The stick figure shape itself is regarded as the organism's body or phenotype. The seed integers are its genotype. And the drawing algorithm is the developmental system or, as Dawkins calls it, the embryology.

The user plays the role of natural selection – out of every generation, she selects a single biomorph. The system then generates the next generation of biomorphs, which inherit their parent's genotype but modified by mutation – this represents inheritance and mutational variation. Through this selective breeding, Dawkins set out to evolve strains of biomorphs to resemble certain target shapes, the different letters of alphabet. But despite the presence of selection, inheritance, and variation, he describes how it was nevertheless impossible

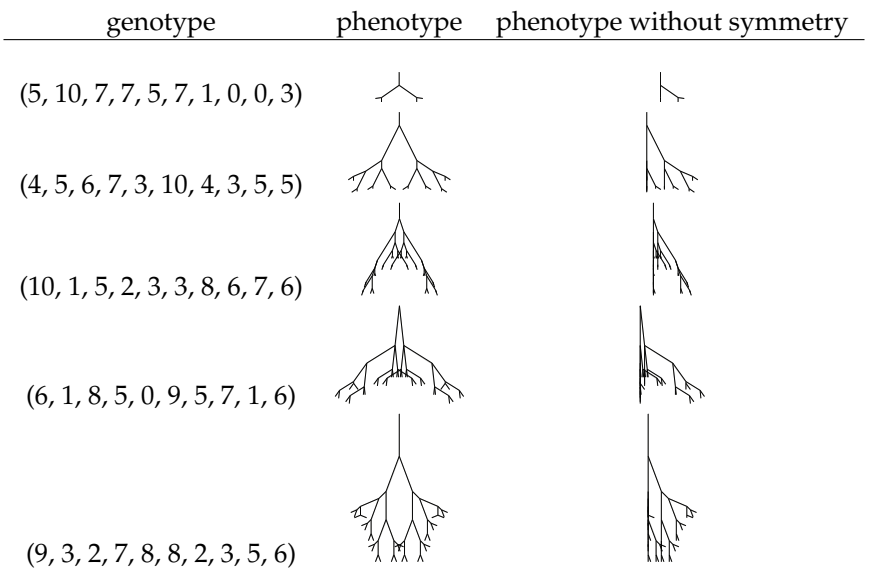


Figure 1.1: Biomorph phenotypes are shapes, drawn programmatically from a genotype of integers. Even when innumerable forms are possible, the drawing algorithm – like the developmental system – profoundly influences the general pattern of those forms, illustrating the idea of evolvability.

to evolve biomorphs resembling the letters of the English alphabet. In every generation, mutation produced random new phenotypes, but the set of phenotypes was always random in the wrong way. Not all variation is equal. This illustrates the first key idea in evolvability – the *structure of phenotypic variability*.

This failure led him to experiment with different drawing algorithms, or embryologies, trying ones that seemed likely to reproduce features typical of biological organisms, such as bilateral symmetry or a segmented body plan. The conclusion is that not just variation, inheritance, and selection, but also the embryological system, is crucial to making it easy or indeed possible for evolution to produce certain results. This illustrates the second key idea – the importance of *the developmental architecture*.

This idea leads Dawkins to draw the distinction between ordinary evolutionary changes and what he calls "evolutionary watersheds". These are rare events which "open floodgates to future evolution", not by increasing fitness but by increasing evolvability. He goes on to argue that, like the changes he made in experimenting with different drawing algorithms, certain changes in biological developmental systems may represent such evolutionary watersheds. Their main impact is not in how they improve immediate fitness, but in how they make other kinds of future innovation easier. He gives as an example the invention of

segmented body plans, hypothesising that it resulted from the evolution of a mechanism like a "repeat" command in a computer program. The first segmented creature probably did not itself enjoy a great increase in fitness, but possessing the embryological potential of segmentation – of repeating body components – was a powerful engine for evolving new phenotypes with different numbers of segments. In the long run, this creature produced additional descendants because it was a "champion evolvor" (sic).

It may seem paradoxical that evolvability should result in additional descendants without affecting immediate fitness. Is it consistent with Darwinism to suppose that natural selection could favour a property such as evolvability, which is fitness neutral and apparently forward-looking? Dawkins finally supposes that some kind of higher level selection might do this work.

These last reflections foreshadow three more key ideas.

Many have repeated the worry that the evolution of evolvability is logically incompatible with Darwinian natural selection, that it is in principle ineligible for natural selection. Let us call this the idea that evolvability is *unDarwinian*.

In the idea that greater evolvability promotes more numerous descendants, one sees the the idea that evolvability should produce characteristic *patterns of diversification in the macroevolutionary record*.

Last, the particular example of segmentation foreshadows later discussions around evolvability and *phenotypic complexity*. Segmentation is the most basic instance of the general idea of modularity – the division of the body into loosely self-contained parts – which itself can be taken as the most basic aspect of phenotypic complexity. In this way, the biomorphs discussion is in fact a modest speculation on the developmental origins of complexity. The style of the discussion is also noteworthy. By proposing that segmentation results from a kind of "repeat" command, Dawkins treats the connection between development and complexity from an abstract and computational point of view, a perspective unusual in the biological sciences, but common in thinking on evolvability.

1.2 Key ideas

The biomorph model is simple but it is wonderfully clear. It charmingly illustrates the idea of organismic evolvability – evolvability as a property rooted in the developmental system,

favoured by a higher level of selection, and perhaps crucial to the evolution of complexity. This model also prefigures the simulation-based approaches of later work in artificial life (discussed in section 2.2.3).

Perhaps because the model is so stylised, Dawkins's discussion makes only passing reference to real biological examples, which can create the impression that evolvability is a radical concept disconnected from the traditional concerns and assumptions of evolutionary theory. This is misleading. The *idea* of evolvability has a deep but diffuse history in evolutionary thinking. Later work will take up the new focusing *term* of evolvability while continuing to work within the idea's natural context of traditional evolutionary theory and biological examples. I will now discuss each of the last section's key ideas in more depth and describe how they fit into this later literature. These are the papers of Gunter Wagner, Marc Kirschner, and John Gerhart, although there are numerous other significant contributions by Yang, von Dassow, Poole, and others.

1.2.1 Structure of variability

Wagner and Altenberg (1996) provide an excellent discussion of evolvability, relating it to longstanding questions in development, in the evolution of complexity, as well as work in evolutionary computation.

They emphasise the crucial distinction of *variation versus variability*. Variation is simply "the actually present differences among ... individuals". This is relatively easily observed, is basic to population genetic models, and is reflected in common measures such as heterozygosity. The observed variation in a natural population will depend on sampling error, on the effects of selection on the various phenotypes, as well as the underlying variability. The variability, by contrast, is the fundamental "potential or propensity to vary". It might be described as the potential range of variation, as that range is defined only by the original process that generates individuals. It is a commonplace that selection can act only on available variation, and this fact is crucial for understanding basic phenomena such as the long-lasting effects of evolutionary bottlenecks. However, variation itself is circumscribed by the underlying variability.

Variability exists prior to all considerations of selection.[2] This is true not only in the

[2]Contrary to this standard view is the idea of "directed mutation", that "cells may have mechanisms for choosing which mutations will occur" perhaps after assessing which mutation environmental conditions would favour (Cairns et al., 1988). Lenski and Mittler (1993), in a later review, argued that evidence does not support this.

temporal sense, since mutations affect every offspring before it matures and has the opportunity to reproduce, but also in the logical sense, since the processes determining variability are separate from those driving selection. Considering variability thus de-emphasises the role of selection. It focuses attention away from the fitness of phenotypes and onto the fundamental biological processes that produce the phenotypes, the processes of mutation and development. These processes determine the exact variability of a given organism – what I will call the *structure of variability*.

Wagner and Altenberg (1996) describe how the distinction between variability and variation is easily overlooked by traditional evolutionary theory. While this is a profound point they present it in a verbal argument typical of traditional evolutionary theory. It drives home the point further to describe the structure of variability formally. This requires only two constructs: the set of possible phenotypes following a mutation, and the probability distribution over that set giving their likelihood. This formalisation appears in other work (e.g., Fontana and Schuster, 1998b). It forces one away from the vague statement that "mutation is random" toward a more exact approach, an approach which distinguishes the random process of mutation, which drives variability, from the combined random process of mutation, sampling, and selection, which determine variation. It is because this formal description completely characterises phenotypic variability that we may say simply that the probability distribution of mutant phenotypes *is* the structure of evolvability.

As evolvability is determined by variability, there can be no understanding of evolvability without an understanding of this structure. Quantifying the idea opens the door to clear questions about its cause and effects. So what determines the structure of variability of an organism? The answer is the developmental architecture.

1.2.2 Developmental architecture

Core components

The structure of variability will depend on the developmental system, the machinery which actually creates the body of the adult organism and which translates a particular genotype to a particular phenotype. Some developmental systems are more evolvable than others.

This is a basic idea in evolvability. But to substantiate it, we need examples from actual organisms. Exactly what aspects of a developmental system, for instance, increase evolv-

ability? In which organisms? These questions have been explored for about a decade by Kirschner and Gerhart. Kirschner and Gerhart (1998), an important early paper, introduces the main ideas of their work. These ideas appear in their textbook on development (Gerhart and Kirschner, 1997), in a popular book-length treatment (Kirschner and Gerhart, 2005), and in subsequent publications (Gerhart and Kirschner, 2007). Their recent work refers to evolvability as "facilitated variation" but the idea is the same. Facilitated variation is evolvability – the way the developmental system, and the structure of the organism itself, can shape genotypic variability to improve the resulting phenotypic variability.

Their research collects numerous examples of *core processes* from cellular biology and from metazoan evolution and development. Core process are aspects of the phenotype or the developmental system that increase evolvability. These examples are categorised based on a handful of common underlying structural principles which describe how they increase evolvability. The categories have names like weak linkage, exploratory mechanisms, compartmentation, redundancy, etc. Since the categories are based on fundamentally structural principles they have the unfamiliar echo of engineering about them, like headings in a taxonomy of major bridge designs. But this is quite appropriate, since that structural logic is what explains why the component tends to increase evolvability.[3]

What makes for a core process? Kirschner and Gerhart's work is extensive and varied. I will try to summarise the essence of the idea, as it relates to other work on evolvability. One intuitive, important example they give is the system of the hox gene system. As Kirschner and Gerhart (2005, p 191) describe it in *Drosophila melanogaster*,

> All the Hox genes are transcription factors that influence the combinations, amounts, and orders of the various conserved core processes at specific regions in the fly. Because of weak linkage, control by specific Hox genes can be easily imposed on many genes .... The fly is divided into 8 or more large compartments .... Engrailed is a selector gene that makes the posterior compartment different from the anterior. Thus, segments at the front end of the fly are like those in the back in that both contain the same kind of subcompartments ... but they differ in the Hox selector genes expressed.

In other words, the fly is divided into predefined zones or compartments. The hox genes

[3]Having identified the broad categories of core processes, they also show how the same core processes which increase evolvability will also tend to increase phenotypic robustness – robustness in the face of mutations, of shocks during development, and of changes in the adult's environment. This indicates how evolvability might be a byproduct of selection for robustness, which has more immediate fitness benefits. While it is a commonplace to note the tension between mutational robustness (reducing the magnitude of the phenotypic effects of mutations) and evolvability (improving the character of the phenotypic effects of mutation), their work goes into greater depth on this topic and puts this connection into the wider context of other forms of robustness (Wagner, 2005).

occupy such a high position in the logical structure of development that their value alone suffices to determine what kind of body segments gets slotted into each compartment. What is notable is the abstraction and coherence of these effects: changes in hox genes are producing changes in the ordering of body segments, rather than scrambled unrecognisable areas of tissue. (Hox genes are also summarised in Lufkin (2005).)

Kirschner and Gerhart describe the core components like the hox system as being ancient, fundamental, and crucial to evolvability. These components are ancient because they have remained valuable over long stretches evolutionary time. They are fundamental to the organism, in the sense that many other components depend on them. And they are central to evolvability because of the way, as they remain unchanged themselves, they facilitate other kinds of phenotype change.

This is the key point and the hox example illustrates this paradoxical contrast. Mutations in the hox genes can produce substantial phenotypic variation. But we cannot say that the hox genes themselves are solely *responsible* for this variation. Rather than thinking of the hox genes in isolation, we should think of their role vis-a-vis the rest of the developmental system. They are like the levers that control a much larger complex of downstream developmental machinery. The machinery stays constant, producing different results as the levers are set in different positions. This exemplifies the constant but multifarious quality of a core component. The core component is (relatively) difficult to evolve, but it is manipulable in a way that makes other kinds of change (relatively) easy.[4]

Genetic architecture

The simplistic picture of a machine controlled by a lever depicts the idea of a core component in *mechanical* terms, and in the phenotype some core components may actually manifest

[4]In one respect, however, the hox gene system is a poor example of a core component. It is true that hox genes do enable large and coherent phenotypic mutations, thus demonstrating the kind of abstracted control that would be expected from an evolvable developmental architecture. However, these large phenotypic mutations do not seem to parallel the phenotypic variation observed between taxa, as might be expected if evolvability created the variation driving phylogenetic diversification. The hox gene system has been conserved over 600 million years of metazoan evolution, but most phenotypic variation across taxa is due to changes in development material downstream of the hox genes themselves. Hox genes may put the head in the right place for both the bear and the fruit-fly, but it is the downstream machinery that makes those heads so different (Tour and McGinnis, 2006).

A model system that might fit better is the little-understood developmental architecture of ctenaphores, or jellyfish, as early research suggests that relatively few mutations are responsible for a great deal of observed phenotypic variation. Podar et al. (2001, p 229) describe how their analysis "has uncovered the lowest level of sequence variability at the level of the 18S rDNA gene in any metazoan phylum" Moreover, "In the case of *Ocyropsis crystallina*, a two-nucleotide difference separates the two subspecies, *O. crystallina crystallina* and *O. crystallina guttata*." (Podar et al., 2001, p 223). The key question is if this small genetic distance is paralleled in the developmental parts of the genome.

this pattern just so overtly. But the essence of the pattern is fundamentally *logical*. The machinery's operation depends on the setting the lever, not the lever itself. For the purposes of evolvability this setting is always the state of the genotype, since evolvability is about how genetic variability gets turned into phenotypic variability.

In other words, a core component is a feature of the developmental system that causes small genetic changes to produce large phenotypic effects. This lets us easily understand the idea of a core component in terms of a more familiar idea – regulation. All (non-neutral) genetic information has some effect. But we can draw a distinction between genetic information which regulates, and genetic information which codes for whatever is regulated. A core component is a special kind of regulated part which has profound effects on the phenotype as a whole.

Under this conception it requires a rather comprehensive model to assess the evolvability of an organism. It is not enough to understand the developmental system only as a sequence of physical transformations. One must have an understanding of its exact dependencies on the genotype, which is sometimes known as a model of the "genetic architecture" (see footnote 5 for context on this term). And to know the effect of variation it will be necessary to understand the effect of a wide range of genotypes. This must be integrated with a physical understanding of the developmental system and the phenotype. Knowing only which genotype produces which phenotype gives no insight on its own, since it must also be possible to recognise which core process is involved. This will require an understanding of the phenotype *as a mechanism*, in order to appreciate how purely mechanical constraints on its organisational coherence will determine what kinds of modifications will disrupt it.

An untraditionally integrated view

It may seem odd to want such a comprehensive model of an organism, a model which takes a mechanistic understanding of development all the way down to the level of the genetic architecture. von Dassow and Munro (1999) explain why such a systemic and mechanistic perspective is not traditional.

Biological researchers have studied the origin of animal form for centuries (or phenotypic complexity, in the modern idiom). But because they knew nothing of molecular biology they were obliged to treat animal development – the crucial element – as a black box. Mere speculation about the inside of this box would quickly lead to blind alleys off

the road of science, such as the non-explanations of vitalism. Only certain kinds of research methodologies were really viable – comparative and historical methodologies which focused on identifying morphological characters and tracing their phylogenetic histories.

It is as if, while researchers could not see what was happening inside the black box, they could still be rigorous about describing the behaviour of boxes, sorting them into logical groups, and keeping track of their histories. However, these methods can never produce detailed, mechanistic, causal explanations and after a while such explanations even begin to seem alien or superfluous. In other words, the traditional methodological limitations of evolutionary biology have limited the kind of explanations which the discipline now expects.

This situation has changed only recently, as research in evo-devo has overcome old limitations, studying the developmental process down to its genetic and microscopic foundations. This "view from inside the black box" of development creates a new theoretic challenge – integrating this mechanistic understanding of development with older models of morphological change. As von Dassow and Munro (1999, p 307) argue, the concept of evolvability works exactly at this tricky interface:

> Ultimately, evolutionary developmental biology (EvoDevo) expects to articulate how the diversity of organic form results from adaptive variation in development. This ambition demands a shift in the way biologists describe causality, and the central problem of EvoDevo is to understand how the architecture of development confers evolvability.

Evolvability requires an unusually comprehensive model of the developmental system, treated down to the underlying genetic architecture. To capture this specific broad sense of the developmental system, considered all the way down to the exact dependencies in the genetic architecture, I will use the term *developmental architecture*.[5]

[5]The above describes what I mean by developmental architecture – an integrated model of the organism, extending from genetics up to the adult phenotype. But it may be worth also describing the somewhat uncertain context of associations already surrounding the term in order to clarify how it relates to those ideas.

In the genetics literature, genetic architecture refers to the exact structure of genotypic variation. Sometimes this is the variation associated with a specific phenotypic trait like a quantitative trait (Zeng et al., 1999; Mackay, 2001) or a disease variant (Weiss, 2006). Sometimes it is the variation associated with an entire taxonomic group ("The purpose of this study is to investigate the genetic architecture of a highly subdivided species...." (Lynch et al., 1999, p100) and (Rieseberg et al., 1999)). The distinction between variation and variability is relevant here. The term usually implies not just a snapshot of the observed sequence variation, but a rather a more comprehensive model, a characterisation of variability, at least as regards a few measurable traits. For example:

> The genetic architecture of quantitative traits links genotypes to phenotypes: it consists of the number, genomic locations, frequencies and effects of quantitative trait loci (QTL), as well as the interaction of QTL alleles within (dominance) and between (epistasis) loci, pleiotropic effects of QTL, QTL by environment interactions, and so forth. (Zeng et al., 1999, p 279)

1.2.3 Effects on macroevolutionary patterns

Evolvability is a favourable structure of phenotypic variability. Since variability provides the reservoir of phenotypic variation that selection acts upon, it will have an indirect effect on the final results of selection. Over many generations, even a small effect will create large-scale patterns in the evolutionary history of an organism. Therefore, looking backwards, one should be able to see the traces of evolvability in the historical evolutionary record.

What traces would one expect to see? This depends on our claim about evolvability. The modest claim is merely that evolvable organisms are more inclined to generate useful variation.[6] This variation creates more opportunities for diversification. Over a long period of time, more of those opportunities will be realised. In this case, one would expect the historical phylogenetic tree to show that some organisms, the evolvable organisms, have diversified much more frequently than their otherwise identical peers.

The more speculative claim is that evolvable organisms have an easier time evolving complexity because their evolvability predisposes them towards the specific kind of varia-

Within this dissertation, such a comprehensive representation will be expressed via the genotype-phenotype mapping function. That is, the genotype-phenotype map describes the entire genome's genetic architecture (in the genetics sense).

However, work outside of genetics, including work in evolvability, typically uses "architecture" to imply a degree of hierarchical structure. Barabasi and Oltvai (2004); Coffman and Davidson (2001) speak of architecture in describing the hierarchical properties of metabolic networks. In discussing evolvability, von Dassow and Munro (1999) speaks of the "architecture of development" in describing the hierarchical properties of developmental systems. In such works the hierarchy implied can be a hierarchy of evolutionary stability and influence, where relatively stable elements in the organism influence relatively malleable elements. Or it can be the hierarchy of developmental causation, the cascade of developmental pathways by which one process triggers another, which triggers others in turn, and so on until maturity. Or it can refer to a hierarchy of logical dependence – a more general notion which is sometimes narrowly construed as the only intended meaning, and sometimes construed as an abstraction and result of the more concrete forms of hierarchy.

These varied and somewhat ambiguous meanings for "hierarchy" reflect the inchoate status of these ideas. But a detailed treatment of this topic requires first distinguishing them in order to explain their interrelations. For instance, a central argument of Kirschner and Gerhart (1998) (expanded in Gerhart and Kirschner (2007)) is that the hierarchy of stability during development implies a parallel hierarchy of stability under evolution, and that the structure of this hierarchy increases evolvability.

So the usage of developmental architecture in this dissertation draws on both the traditional genetics sense and this hierarchy-oriented sense. It is assumed that certain regulatory, metabolic, and developmental mechanisms are relatively constant over time, that they establish a context shaping future evolutionary change, and that they are sufficiently logically distinct to warrant analysis in their own right. These mechanisms are the core components of the developmental architecture of the organism.

[6] "More inclined to generate useful variations" – it is reasonable to ask, more inclined *than what*? One can look at it comparatively, and then evolvable organisms are simply "more inclined" than other, less evolvable organisms. This view highlights differences between lineages so it makes the idea very concrete and easy to study empirically. This is essentially the view from outside the black box of development. But this is agnostic on the evolutionary and developmental causes of evolvability.

If we believe that evolvability itself evolved, that evolvability itself is therefore an adaptation, then we might expect to find certain features of an organism that specifically promote evolvability. This leads one to view evolvability not as a comparative but as an intrinsic property of organisms. On this view evolvable organisms are simply "more inclined" than would be expected based on a null theory of phenotypic variability – that is, a theory that did not suppose special mechanisms promoting evolvability or a selective advantage driving evolvability.

Unfortunately there is no clear null theory of phenotypic variability indicating exactly what to expect in the absence of evolvability. Thus, any theory of evolvability as a special kind of phenotypic variability must also take the first steps in characterising phenotypic variability in general.

tion that leads to complexity. In this case, one would expect to see that the histories of some organisms showed *kinds* of changes over time that were distinctively conducive toward complexity. This might include, for instance, changes that tended to reproduce coherent anatomical modules. If one lineage, over a long period of time, repeatedly shows this same kind of change much more than its otherwise identical peers, then that might be because of a difference in evolvability.

In addition, one would expect a stratification in the frequency of different kinds of evolutionary changes. The idea of a core component posits two kinds of genetic material – compact and easily modifiable regulatory information, and longer stable sequences coding for regulated parts. These two kinds of genetic sequences imply two orders of evolutionary change – frequent genetically minor changes to regulatory material, and rarer genetically major changes to the core components themselves. Both could produce a major change at the phenotypic level, but one would expect to see a difference in the kind of phenotypic change. One would expect the rare genetically major changes to alter the basic plan of the organism's phenotype (e.g., Dawkins's "evolutionary watersheds"), and the frequent genetically minor changes to generate variant phenotypes within the same plan.

Do these traces exist? As early as 1978, the research of Rupert Riedl sought to explain such unusual patterns in the evolutionary record (Riedl, 1978). Riedl believed that certain body plans were especially prevalent and stable over long periods of time. He explained this in terms of a theory of how the organism's need to maintain the functional integration of a complex phenotype, under incremental change, created a predictable set of internal constraints. The constraints shape its phenotypic variability, making it more or less evolvable, affecting its macroevolutionary history.

Riedl's work does not come out of nowhere. Since the first days of comparative anatomy biologists have sought meaningful patterns in the evolutionary record, and theories of evolvability are just a modern continuation of this effort. Wagner and Laubichler (2004, p 98) comments on how Riedl's work, for instance, is essentially an effort to explain such patterns from a modern perspective – by offering a simple model of the organism as a system, and working within the micro-evolutionary, population genetical framework of the Neo-Darwinian synthesis:

> Riedl thus conceptualized the idea of a body plan, a much maligned concept in the 20th century, as a variational concept, rather than an "abstract principle"

> as it was considered by so-called idealistic morphologists.... All that is added by Riedl is the recognition that different degrees of variability are a serious biological problem. The pattern of variability is a phenomenon that requires close attention to the organismal or "systemic" properties of organisms. The main question then, is what causes the different levels of variability, and how do they originate in evolution.

More recently Yang (2001) compared genealogical histories of different insect taxa, finding greater speciation in the more modular taxa. Evolvability and speciation are different things. But this result is consistent with the intuitive view that modularity is conducive to evolvability (reminiscent of the example presented by Dawkins (1989)), and that evolvability would therefore lead to higher rates of speciation. (Section 1.3 offers further discussion on connections with traditional speciation research.)

The order Coleoptera, beetles, is also a natural example, as there are more species in the order Coleoptera than in any other animal or plant group (Farrell, 1998). On this point McPeek and Brown (2007) argues the contrary, presenting molecular evidence that diversification rates are the same across animal taxa, and that the greater observed diversification of beetles is merely because of their age. They've had more time to diversity. This contradicts the view that some taxa speciate more readily because of their developmental architecture.

Why have there been relatively few investigations? One serious challenge is knowing what to look for. What matters is not just increased rates of diversification, which could be due to fitness, but increased diversification due to an exceptional underlying variability. But the observed history always shows the range of variability *after* it has already been filtered through the action of selection. It is like comparing two paintings to see which has more colors when you can only look at them through a red-coloured lens.[7]

[7]It is also worth noting how the expected macroevolutionary traces of evolvability differ from the traces of selection. Selection is a bias in reproductive success. A long time record exposes the power of selection because a small but consistent selective effect can add up over multiple generations to produce a large net effect. For instance, if it is better to be taller, then generations of small increases in height add up to a large increase in height.

Evolvability is different because it is not necessarily additive. Evolvability is a bias in variability, which means a bias in which mutant phenotypes are possible or in their probability. In each generation, evolvability might create additional phenotypes of various kinds. But in each generation, selection might favour a different one of those kinds, so that it would be hard to see any single total effect. For instance, an evolvable lineage might produce more taller offspring and more faster offspring. In one generation, selection might favour the taller ones; in the next, the faster ones. Evolvability has had an effect on two generations, but the result is visible in the diversification rather than in the accumulation of a single trait.

In short, the advantage of a long time record for observing the effects of selection is that small but consistent selective effects add up over time. The advantage of a long time record for observing the effects of evolvability is simply that it provides more time for detecting patterns of diversification which are driven by the structure of variability. (The only exception to this is if we were talking about the evolution of evolvability itself, with evolvability itself increasing over time. But in that case we are really seeing again selection over multiple generations, just the selection of evolvability.)

Of course, at the margin of very weak effects, selection and evolvability are in interplay. This is because an individual needs both good chances to reproduce (fitness) and good chances of being created through mutation

And even if we could control for the effects of selection and look at the raw variability, that still says nothing of its causes. If no organism were evolvable, then we would not see an increased variability anywhere. But if evolvability were very common, and virtually every complex organism had a variability which has been boosted by evolvable features of its developmental architecture, then we would take that rate of variability for granted and still not see an increased variability anywhere. There's no easy way to distinguish between the variability which is called "normal" because it is ubiquitous and that which is "normal" because evolvability plays no role in it.

This is why it is not enough only to look for effects of evolvability within the macroevolutionary record. One must try to correlate these supposed effects with observations of supposed developmental causes of evolvability. But this brings us back to understanding the developmental architecture, which is a central challenge. (See also "The Identification of Developmental Constraints" in Maynard-Smith et al. (1985) for a discussion of these methodological challenges.)

1.2.4 Incompatibility with Darwinian selection

The last section discussed the traces of evolvability on macroevolution and the problems in observing those traces. A more fundamental problem is posed by the argument that the *evolution* of evolvability is logically impossible because it violates basic tenets of Darwinian natural selection.

The argument is simple: selection only promotes fitness. Evolvability is not necessarily correlated with fitness. You could have an individual with high fitness and low evolvability, or high evolvability and low fitness. Since selection only promotes fitness, there is no reason for selection to promote evolvability. But if evolvability is not an adaptation shaped by generations of selection, then there is no evolution of evolvability. On this view there might be differences in evolvability, but these would likely small and inconsequential since selection cannot act over time to amplify them.[8]

(parent's evolvability). Large differences in fitness could suppress small differences in evolvability, and large differences in evolvability could suppress small differences in fitness.

[8] A more radical argument is that evolvability is illicitly forward-looking. Selection is opportunistic: it cannot see into the future, so it only rewards immediate benefits. But evolvability is defined as an organism's *future* evolutionary potential. Therefore, there can be no selection for evolvability, just as there can be no selection for other forward-looking properties, like resistance to a disease that will not exist for many generations, or the propensity to have descendants that will not be hit by a particular asteroid a hundred years in the future. It is easy to define a property which is advantageous but ineligible for selection, and evolvability is such a property.

This argument slips on in its own subtleties. Evolvability is not merely a synonym for "future evolutionary

A common next thought is that the evolution of evolvability requires group selection or species selection. Since evolvability is not promoted by individual selection then it must be promoted by something else. Evolvability affects why some species reproduce as species (i.e., speciate) and others don't. So maybe species selection is what promotes evolvability. Or, since evolvability is about variation and properties of groups of individuals, maybe it requires group selection.

If this argument seem a bit sketchy that is because it is. In fact, the claim that evolvability is unDarwinian is a bit of a bogeyman: it comes up repeatedly in the literature because researchers are always rebutting it, not because anyone has advanced it in earnest. What seems to happen is that researchers notice, correctly, that evolvability is not promoted by selection in the same direct way as a beneficial phenotypic trait. But they can intuit that the evolution of evolvability is still possible. Thinking about this puzzle calls to mind familiar controversies about group selection, species selection, and inappropriately teleological interpretations of selection.

One can see this process at work in the biomorphs model. Dawkins (1989, p 218–219) notices that the evolution of evolvability would occur through a process different from ordinary natural selection. First he observes that evolvability is not promoted by normal selection:

> As the ages go by, changes in embryology that increase evolutionary richness tend to be self-perpetuating. Notice that this is not the same thing as saying that embryologies that give rise to good, healthy individual organisms tend to be the embryologies that are still with us, although that, too, is no doubt true. I am talking about a kind of higher-level selection, a selection not for survivability but for evolvability.

Then he hastens to distinguish this point from the classic fallacies, urging that "we certainly should have no truck with suggestions that individual animals might forgo their selfish advantage because of possible long-term evolutionary benefits to their species." Finally, he concludes that evolution of evolvability may be a rare instance of species selection, and that this is allowed because evolvability depends on the developmental architecture, which is a "higher-level property" existing at the level of the species.

Later work has settled on the view that evolvability only requires clade selection, a pro-

potential", although certainly such a narrowly conceptual definition would allow for bizarre, unscientific forms of explanations. The idea is that evolvability comes down to phenotypic variability, that over the long run variability will tend to predict future evolutionary potential. However, this argument is responding to the correct intuition that there is something puzzling about how variability is promoted over the long run.

cess which does not depend on evolvability qualifying as a higher-level property. Williams (1992, pp 26–27, section 3.1) discusses this issue in detail.

Later work also repeats this pattern of defending the Darwinian compatibility of evolvability against unnamed critics. Altenberg (1995b, p 208) calls it the "good of the species" problem, and presents a population genetic model showing the viability of evolution of the evolvability, and a simple model of evolvability as a fitness distribution. Turney (1999) concedes "there is no direct selection for evolvability, but there is nonetheless a large-scale trend", while presenting a similar fitness-based model to show how the trend works. Kirschner and Gerhart (1998, p 8426) include a section "Can there be selection for evolvability?" addressing the issue, and concluding that the evolution of evolvability will depend on clade or lineage selection. West-Eberhard (1998, p 8419) gives a short commentary on the history of work on evolvability, especially as regards the tension between Kirschner and Gerhart (1998) and population genetic approaches. This commentary re-iterates the usual concern and the usual answer: "The idea of evolution of evolvability ... might be questioned by some evolutionary biologists because it requires selection above the individual level.... Several evolutionary biologists have argued cogently in favor of clade selection for evolvability....". Brookfield (2001, p 107) argues the same point skeptically, regretting that it depends on a "clade-selection model, which evolutionary theory has traditionally found difficult to deal with". Bedau and Packard (2003) declare that "evolvability has been criticized on the grounds that it involves group selection and it bestows a future rather than present benet on individuals..." and refers to defences of the evolution of evolvability in Wagner (1981); Conrad (1982, 1990); Altenberg (1995b); Wagner and Altenberg (1996); Kirschner and Gerhart (1998); West-Eberhard (1998); Dawkins (1989). However, it is notable that neither they nor these defences ever actually refer to an article attacking the idea.

Perhaps the strongest reason to suppose that evolution of evolvability is compatible with Darwinian selection was pointed out by West-Eberhard (1998, p 8419), where she notes that "Darwin was the first to argue in favor of selection for variability *per se* as part of his 'principle of divergence"'. Evolvability is just a favourable kind of phenotypic variability. Was Darwin himself the first proponent of the evolution of evolvability? The answer seems to be yes, although he makes the argument in passing. In the section on "divergence of character", Darwin (1859, p 111) argues that ecological dynamics will cause small differences between organisms to increase, driving diversification. On page 117 he

describes the roles of variability:

> After a thousand generations, species A is supposed to have produced two fairly well-marked varieties, namely a^1 and m^1. These two varieties will generally continue to be exposed to the same conditions which made their parents variable, and the tendency to variability is in itself hereditary, consequently they will tend to vary, and generally to vary in nearly the same manner as their parents varied. Moreover, these two varieties, being only slightly modified forms, will tend to inherit those advantages which made their common parent (A) more numerous than most of the other inhabitants of the same country; they will likewise partake of those more general advantages which made the genus to which the parent-species belong, a large genus in its own country. And these circumstances we know to be favourable to the production of new varieties.

This passage is muddied somewhat by the way Darwin combines ecological and genetic arguments, obscuring his exact causal claims. However, he is making the key distinctions and claims to argue for evolution of evolvability: (1) that variability ("the tendency to variability") is distinct from immediate phenotypic traits ("those advantages which made their common parent more numerous"), that (2) both variability and phenotypic traits are heritable, and (3) that this tends to promote diversification.

So can evolvability itself evolve under Darwinian selection? The consensus answer seems to be yes, but not in the same direct way as an immediately beneficial phenotypic trait. It depends on clade selection, a process less familiar than individual natural selection. To explain it in detail, Brookfield (2001) is surely right when he argues that a quantitative model of evolvability is needed. The models which expose it best are the simple models which treat evolvability only as a function of individual fitness. An individual's evolvability can then be reduced to nothing more than a measure of dispersion, such as the standard deviation of the fitness of her offspring. In this setting the argument for the evolution of evolvability becomes very clear: multiple computer simulations and microbial experiments have confirmed that, even without a direct fitness benefit, natural selection will indirectly promote evolvability (Altenberg, 1995b; Turney, 1999; Burch and Chao, 2000). This is discussed further in section 2.3.1.

1.2.5 Effects on the evolution of complexity

How is evolvability supposed to support the evolution of complex organisms?

Evolvability describes how the developmental architecture can create systematic biases in the way genotypic variability is translated into phenotypic variability. An evolvable

developmental architecture can therefore, it is argued, translate genetic changes which are relatively "easy" into phenotypic changes which would otherwise be relatively "difficult". The complexity of living things is an archetypal example of something which seems difficult to evolve. Critics have wondered how an incremental process like evolution can produce complexity. Evolvability is one answer.

But how, specifically, might evolvability facilitate the evolution of complexity? The answer to this depends on what is meant by complexity, and it must be admitted that there is no simple single answer to that question. Despite many interesting efforts there is no agreed formal definition of complexity. [9] Consequently, as regards organisms, "complexity" is used in roughly the same colloquial sense one would apply describing a complex machine. It is getting at a number of ideas: that an organism has loosely separate components parts, that those parts are coordinated to serve a common function, that the parts are numerous and delicately coordinated with each other, that the parts are nested hierarchically, etc. Research on the complexity of organisms typically focuses on one of these latent intuitive senses, one of these *aspects of complexity*, taking it as a proxy for complexity in order to produce a viable practical measure. For instance, McShea (1996) has measured the complexity of various metazoans by counting their tissue types. Fusco and Minelli (2000) measures complexity in segmented organisms by looking at "the degree of morphological differentiation of segments".

One aspect of complexity that has received frequent attention is modularity, the idea that an organism is divided into relatively self-contained subsystems or modules (von Dassow and Munro, 1999; Watson, 2006). This aspect of complexity lends itself to a plausible story about how it might be supported by evolvability. At the simplest level, for instance, an evolvable developmental architecture might preserve the integrity of modules by minimising pleiotropic mutations. Wagner and Altenberg (1996) observes that multiple approaches have suggested that a key for the evolution of complex adaptations is that "further im-

[9] Adami (2002) summarises various proposed formal definitions of complexity. Many of these proposals – such as Kolmogorov complexity, statistical complexity (Crutchfield and Young, 1989), and his own suggestion of physical complexity (Adami et al., 2000) are highly abstract. For instance, measures like effective measure complexity (Grassberger, 1986) let you calculate the complexity of a string of characters independently of what that string means – whether it is an English sentence, the genome of an organism, or a DNA sequence of pure gibberish. This abstraction reflects these definitions' origins in disciplines such as dynamical systems theory, as well as their aim to unite diverse phenomena under a single theory.

In theory, this abstraction allows some of these measures to detect complexity anywhere, in any physical mechanism, including an organism. In practise, it is unclear how to apply these measures to real physical systems, like organisms, or whether they would parallel the commonsense ideas about complexity that we are trying to formalise. This problem has dogged work on complexity since early efforts such as Gell-Mann (1995); Gell-Mann and Lloyd (1996).

provements in one part of the system must not compromise past achievements". As they put it, "independent genetic representation of functionally distinct character complexes can be described as modularity of the genotype-phenotype mapping function" (Wagner and Altenberg, 1996, p 971).

It is crucial to note that there are two distinct notions of modularity at work here. The first is the *modularity of the phenotype*, i.e., the notion that the phenotype itself can be divided into "functionally distinct character complexes". This focuses on how a phenotype resembles a well-designed mechanism with independent but coordinated parts, like a watch or a car. This aspect, phenotypic modularity, serves as a stripped-down definition of complexity, a definition which needs to be in place to make sense of the claim of that evolvability facilitates complexity.

The second notion here is the *modularity of the mapping*, the notion that distinct phenotypic parts are represented by distinct genetic representation. This is a quality that makes a mapping more evolvable. It allows you to modify one part of the organism without modifying another, since the mapping defines a set of genes which affects one part but does not another. It is this idea, the modularity of the mapping, which explains how evolvability facilitates the modularity of the phenotype itself.

Any claim that evolvability facilitates complexity needs to have this kind of two-part structure, including first a description of what is meant by complexity, and then a description of how evolvability facilitates it. For instance, in his work Riedl (1978) offers a different model of complexity, focusing on the network of functional dependencies between the different parts of an organism. Since new parts can only depend on pre-existing parts, it becomes progressively harder to modify older parts, creating what he calls "burden", reducing the evolvability of the organism. He needs a separate concept, the "imitative epigenotype", to describe how the structure of phenotypic variability can be favourably matched with the needs of fitness and the constraints of burden. This describes how evolvability supports complexity, as he defines it.

Such narrow, relatively concrete definitions of complexity are easier to understand, to apply experimentally, and to relate to evolvability. However, it is worth noting that the more general, abstract definitions of complexity, even if they are harder to apply, suggest interesting intuitions. For instance, various general formalisations of complexity, such as effective complexity (Gell-Mann, 1995), depend on the idea of the minimal possible descrip-

tion of a system. Complex systems are supposed to have a kind of deep-seated regularity that makes it possible to describe them with unusual brevity simply by describing that regularity. (For instance, only a body with repeated segments can be succinctly described by saying "repeat the segment".) One can think of the genotype as a kind of description of the phenotype. Part of what makes a genotype-phenotype mapping evolvable is that it requires smaller number of mutations for big changes, that is, a smaller genetic description for the phenotype as a whole. What this suggests is that complex phenotypes, just as a result of the nature of complexity, lend themselves especially well to evolvable genotype-phenotype maps. These are in fact the maps that leverage their complexity. If this were true, then it would be very difficult to create a complex organism without using an evolvable genotype-phenotype map.

This is an interesting suggestion but it is hard to take it beyond being a suggestion. The lack of a consensus practical formal definition of complexity, the first part of any claim that evolvability promotes complexity, makes it hard to translate such ideas into a clear, concrete, falsifiable argument. In chapters 4, 5, and 6, I develop a technical model which works on a very modest sense of complexity based on different numbers and types of body segments.

1.3 Relation to traditional research topics

Work on evolvability is continuous with traditional research topics in evolution so it is worth reviewing how these topics address issues of evolvability but under a different name.

First, work on "developmental constraint" is about how the developmental system and basic biophysical limitations affect the structure of phenotypic variability (Gould and Lewontin, 1979; Maynard-Smith et al., 1985; Calow, 1997). In an early review, Maynard-Smith et al. (1985, p 265) defines a developmental constraint as "a bias on the production of variant phenotypes or a limitation on phenotypic variability caused by ... the developmental system". This paper is a pellucid and well-grounded discussion of evolvability, missing only the word evolvability. It clearly distinguishes variation from variability, explains the key contrast between genotypic and phenotypic accessibility, and compares experimental methods for distinguishing evolvability from selection. The final section broaches the topic of the evolution of evolvability, noting that "One final question concerns the extent to

which genetic mechanisms themselves have evolved so as to serve an evolutionary function" (Maynard-Smith et al., 1985, p 283), arguing that there is strong evidence for it in the case of prokaryotes, and raising the usual concern of Darwinian compatibility. More recently, Gould (2002) has offered a book-length work which treats developmental constraints as a major force in evolution and discusses evolvability in that context. In comparison with evolvability research, traditional efforts have focused more on constraints due to basic biophysical constraints rather than the genetic underpinnings of the developmental architecture. This is understandable: morphology is open to commonsense reasoning, visible in the palaeontological record, and a traditional object of study.

Second, the literature on speciation is obviously relevant to understanding the diversification of the species. Work on evolvability focuses on the intrinsic propensity of an organism to generate phenotypic variants, i.e., on the structure of variability and its intrinsic roots in the developmental architecture. This variability furnishes the original small divergences that may grow to define new species. Work on speciation surveys the various isolating mechanisms, both intrinsic and environmental, that might create the reproductive isolation necessary to amplify these small divergences (Wood and Rieseberg, 2005).

Where do these different focuses of research overlap? It would seem to be in speciation research focused on genetic isolating mechanisms, or postzygotic barriers. Presumably, studying the genetic barriers separating two emerging lineages is quite related to studying the intrinsic structure of difference that produced those lineages. For instance, the standard "complementary loci" model of postzygotic barriers describes a genetic architecture where fitness interactions among just a few loci produce divergent lineages with reduced hybrid fitness. While noting that "only a handful of well-designed studies have examined the genetic basis of prezygotic barriers", Ungerer and Rieseberg (2005) conclude that initial studies indicate that "reproductive barriers can result from the combined effects of many loci (as envisioned by the neo-Darwinian theory) but often require few genetic changes". This parallels the idea that a few mutations, in evolvable organisms, can create large (thus perhaps incompatible) differences in phenotypic variants.

(One difficulty in consolidating this research overlap is that models of speciation rely on the biological species concept, which defines species as an interbreeding population. They thus assume sexual reproduction and build on details of the mating and fertilisation system, such as ploidy. Models of evolvability rely on a phylogenetic species concept, ignore sex,

and say little about the process of divergence, perhaps assuming that a new phenotype is always different enough to occupy a new ecological niche.)

Third, the idea of "complexity" also appears in various traditional research areas such as the literature on evolution of innovations and complex organs. In the chapter on theoretical difficulties with evolution, a famous section of the *Origin* is dedicated to the problem posed by "organs of extreme perfection and complication" (Darwin, 1859, p 186). Darwin gives the example of the human eyeball, acknowledging how "absurd" it seems to think it could be formed by natural selection. But, he argues, even the most complex organs can develop incrementally. A modern understanding of inheritance, unavailable to Darwin, allows us to appreciate that the fundamental requirement here is that *genetic* change be incremental, leaving the possibility that an evolvable developmental architecture amplifies the effect on phenotypic change. Traditional research does not tackle complexity in general presumably because no formal theory of complexity-in-general has been a great success. Still, it seems foolish to deny the powerful intuition that biological organisms are uniquely complex or, as Darwin called it, "the vague yet ill-defined sentiment, felt by many paleontologists, that organisation on the whole has progressed" (Darwin, 1859, p 345).

Last, the idea of a core component – a part of the developmental architecture which persists over many generations, but is constantly expressed in different ways – is just another instance of the commonplace rule that evolution by natural selection is conservative and incremental. It is more likely to modify old features than to invent new ones. This is seen at the level of the phenotype, in the way new organs tend to emerge as adaptations of older structures. It also appears at the genetic level, as new genes tend to emerge from the duplication and modification of existing genes. Old features are re-used; nothing appears out of nowhere. Research trying to identify core components is merely looking for the same pattern in the developmental architecture as a whole. Sometimes this will be obvious, as when these processes leave gross physical signatures, as with the overall body plan of the organism.. Sometimes it will be more cryptic, as with cellular biological mechanisms or the exact network of epistatic dependency among multiple loci.

1.4 Appendix: biomorphs source code

Listing 1.1: Python code for drawing Biomorphs

```
"""This module re-implements, in Python, the Biomorphs algorithm
described in Dawkins1989 p205.

The genome is broken into three blocks as follows::

  g[0] = n/a

  g[1] = affects x
  g[2]
  g[3]

  g[4] = affects y
  g[5]
  g[6]
  g[7]
  g[8]

  g[9] = recursion depth

Every time you run this module in NodeBox, it will draw a new biomorph
and tell you its genome on the console. NodeBox is a 'Mac OS X
application that lets you create 2D visuals (static, animated or
interactive) using Python programming code and export them as a PDF or
a QuickTime movie'. <http://nodebox.net/code/index.php/Home>.

NodeBox version 1.9.2 runs under Mac OS X 10.4 and 10.5. NodeBox is
released under an MIT License, so it should be available freely and
indefinitely. Also, this module uses such simple graphics primitives
that it should be straightforward to port to other systems.
"""

# Set the pen to draw in black
```

```
stroke(0)

def tree(x,y,length,dir,dx,dy):
    if dir < 0:
        dir = dir + 8
    if dir >= 8:
        dir = dir - 8
    xnew = x + length * dx[dir]
    ynew = y + length * dy[dir]

    beginpath()
    moveto(x,y)
    lineto(xnew,ynew)
    endpath()

    if length > 0:
        tree(xnew,ynew,length-1,dir-1,dx,dy)
        tree(xnew,ynew,length-1,dir+1,dx,dy)


def plugin(g, noreflection=False):
    """Take genotype g, and returns inputs for drawing algorithm"""
    dx=[0,0,0,0,0,0,0,0]
    dy=[0,0,0,0,0,0,0,0]

    order = g[9]

    dx[3], dx[4], dx[5] = g[1:4]
    if noreflection is False:
        dx[1] = -dx[3]
        dx[0] = -dx[4]
        dx[7]=-dx[5]
    dx[2]=0
```

```
    dx[6]=0

    dy[2],dy[3],dy[4],dy[5],dy[6]=g[4:9]
    dy[0] = dy[4]
    dy[1] = dy[3]
    dy[7] = dy[5]
    return (order,dx,dy)


def mkTree(g,noreflection=False):
    """Draw a biomorph with genotype g

    This draws all genotypes with a startdir of 10, which produces a
    nice right orientation.

    noreflection - suppresses bilateral symmetry.
    """
    startx=200
    starty=200
    startdir=10
    print "Drawing biomorph with genome: %s:" % g
    order, dx, dy = plugin(g,noreflection)
    tree(startx,starty,order,startdir,dx,dy)

#
# convenience methods
#

def drawRandom():
    """Generate a random genotype and draw it"""
    import random
    g= [random.randint(0,10) for i in range(10)]
    mkTree(g)
```

```
drawRandom()

# the menagerie used to generate the chapter figure
zoo=[[5, 10, 7, 7, 5, 7, 1, 0, 0, 3],
     [4, 5, 6, 7, 3, 10, 4, 3, 5, 5],
     [10, 1, 5, 2, 3, 3, 8, 6, 7, 6],
     [6, 1, 8, 5, 0, 9, 5, 7, 1, 6],
     [9, 3, 2, 7, 8, 8, 2, 3, 5, 6]]

# mkTree(zoo[0],noreflection=False)
```

Chapter 2

Taxonomy of evolvabilities

Chapter 1 discussed organismic evolvability, the most richly developed sense of the term evolvability. This idea has concrete and intuitive connections with animal development, animal morphology, and the historical evolutionary record, and it addresses traditional concerns in those fields. However, it has also been hard to systematise sufficiently to generate a clear research programme.

This chapter will describe three newer notions of evolvability – trait evolvability, individual fitness evolvability, and individual mutational evolvability. These notions have emerged from research in newer disciplines such as population genetics, molecular evolution, microbial evolution, and evolutionary computation. These disciplines better support quantitative methods and they have tended to redefine evolvability to fit those methods. However, these methods are not well-suited to treating evolvability as a property of an organism. As a result, while these redefinitions of evolvability make the idea more exact and tractable, they have also reduced its explanatory scope, cutting it off from the original issue of biological diversification and complexity.

This chapter also discusses a fourth notion, substrate evolvability, which I introduce to describe an unacknowledged *de facto* perspective underlying relevant work in evolutionary computation and astrobiology.

What is the point of these distinctions? It is easy to fall into merely verbal disputes when everyone is using the same word to mean a different thing. I propose these terms for evolvability to highlight how these notions are conceptually distinct. None of these terms captures all of the real and inter-related issues discussed under the heading of evolvability,

but each does illuminate a different cluster of those issues. This classification, it is hoped, gives a map of those issues and discussions.

Together with chapter 1, this chapter provides background to the analysis in chapters 3 through 6. This dissertation takes inspiration from the quantitative methods developed to articulate these newer, narrower notions of evolvability. It then modifies and extends these methods to treat evolvability as a property of the organism, in order to restate the longstanding ideas of organismic evolvability in an exact fashion.

2.1 Trait evolvability

The most reductive definition is what we might call *trait evolvability*:

Definition: trait evolvability

> Within a population genetic model, evolvability is some statistic chosen to measure how rapidly a specific trait responds to selection.

This view treats evolvability as a property not of an organism, but of a trait. It is presented in a widely-cited paper by Houle (1992). He begins defining evolvability generally as "the ability of a population to respond to natural or artificial selection". His analysis, however, is more specific – he is really comparing the merits of a variety of possible statistics for representing, within a population genetic model, how much a given quantitative trait responds to selection. For instance, he argues against the use of "narrow-sense heritability h^2", which relates the selection differential S to the response to selection R in the equation $R = h^2S$. This is unsuitable because, first, you cannot directly compare the heritabilities of traits measured in different units and, second, because it is affected by the selection regime so that it measures a kind of variation as opposed to variability. He goes on to compare h^2 with two other candidate measures on the basis of such issues as dimensionality, comparability, etc. He concludes that "while it is straightforward to define measures of evolvability appropriate to any particular circumstance, it is not possible to define a single measure appropriate for all circumstances".

Houle does not comment on what the impossibility of a single measure of evolvability implies for the status of evolvability as a single theoretical construct, but his analysis suggests a sceptical view: while "evolvability" is a convenient word, it does not name a

coherent idea which has general explanatory power like, for instance, the idea of fitness. However, it is also easy to see that Houle has not really addressed the idea of evolvability as it is understood outside of population genetics. All the measures of evolvability which he compares are in fact means of measuring the "evolvability" of a single trait. But evolvability is supposed to describe large macroevolutionary changes of organisms, like the evolution of new species, of phenotypic innovations, and of complexity, none of which can be understood in terms of a single quantitative trait. While the speed of evolutionary response of a single trait is no doubt important to these issues, it is too simple to stand in for them.

This is because these issues concern seemingly unpredictable responses to selection, responses which are puzzling in how much they seem to demonstrate coordinated, simultaneous changes along many dimensions. Addressing this would require that the measurement be fit into a model of what causes the responsiveness of some traits more than others, and of which traits bear on complexity. Also, in order to see whether differences in evolvability explain features of one species more than other, the model should allow comparisons between populations. Although Houle begins by defining evolvability as an "ability of a population", the population only appears afterwards as the ensemble needed for calculating statistics of traits.

Houle's implicit dismissal of evolvability illustrates how an exclusively population genetic approach forces a narrow analytical focus and excludes all the larger (and admittedly vaguer) macroevolutionary facets of the idea of evolvability. None of this is to minimise the value of Houle's work, which is well-adapted to serve its own purposes. These problems are endemic to efforts to apply population genetic and quantitative genetic theories to questions of phenotypic innovation (Wagner and Stadler, 2003, p 507).

2.2 Substrate evolvability

Another notion of evolvability is the notion of *substrate evolvability*. This is not mentioned in the literature. I propose this idea in order to distinguish it from other senses of evolvability, and because it is an important generalisation of a *de facto* perspective underlying a few concrete areas of research.

Definition: substrate evolvability

> Consider a *substrate*, that is, a model or physical system which defines a set of rules for what entities can exist and how those entities change and interact. A substrate is *evolvable* if its rules suffice to allow those entities, through their interactions, to create rich evolutionary processes.

The general question of substrate evolvability is then, "What makes a substrate evolvable?"

2.2.1 The concept of a substrate

This concept is rather abstract so it is easier to explain it by example.

Consider the question, "What is necessary for evolution?". Stated so generally, one conventional answer might be that evolution only requires three ingredients – variation, selection, and inheritance (or to put it slightly differently, "replication, variation (mutation), and differential fitness (competition)" (Dennett, 2002)). Note that none of these are physical ingredients. This is somewhat unusual. If you asked what is necessary for water, the answer would be hydrogen and oxygen. If you asked what is necessary for hydrogen, the answer would be protons and electrons.

The reason for this difference is that evolution by natural selection is itself not a physical thing. It is the name for a general dynamic process. As such, it is a process that is inevitable once certain basic logical conditions are met – variation, selection, and inheritance. This is why it is possible to speak quite literally of the evolution of memes, of fashions, and of anything else existing under those conditions. Although the particular form of biological evolution on earth has depended on contingent historical events and is discoverable only by empirical investigation, the validity of evolution-the-general-process is an *a priori* truth.

But evolution-the-general-process never happens in a vacuum. It is always realised as evolution-in-some-particular-substrate or world. Some worlds will have more potential for evolution than other. For instance, consider again the example of biomorphs from section 1.1. If we waited long enough, and kept choosing the right biomorph in each generation, should we ever expect to see the evolution of biomorphs that could eat or reproduce? No, that is impossible because the biomorph simulation environment completely defines what kind of biomorphs can exist, and that definition places a ceiling on what is possible. It allows innumerable kinds of biomorphs – but only as long as biomorphs are just stick

figures. Even if we were to design a biomorph by hand, deliberately choosing its genotype, there is no genotype that could produce a biomorph with the capacity for processing information, interacting with other biomorphs, or pursuing resources. Indeed, in the biomorph world, there are no such things a resources. The biomorphs simulation is therefore a substrate which is rich enough to allow genuine evolution (since it allows variation, selection, and inheritance), but also too impoverished to allow evolution of anything but elaborate stick figures.

By contrast, consider the substrate defined by terrestrial chemistry – the chemistry of the elements found readily on earth, such as water and carbon, at the temperature and pressure regimes which prevail on earth. These factors have changed over the history of the planet but always within certain limits. If subtleties of subatomic physics were different but had no effect on this chemistry then, as far as we know, terrestrial biology would be exactly the same. In other words, terrestrial chemistry is the immediate substrate of biology. [1] This set of rules is much richer, and allows the construction of life as we see it around us. Not only does it allow the manual construction of such life, but it has evidently allowed the evolution of such life, which is a harder thing.

These examples illustrate the idea of substrate evolvability, which should clarify the *general question of substrate evolvability*. The specific question, as regards biomorphs, is how much evolution can the biomorph world support (not much). The specific question, as regards terrestrial chemistry, is how much evolution can the earth support (a great deal). The general question of substrate evolvability is, "what, in general, makes a substrate evolvable?"

What is the point of considering such an abstract question? One reason it is important it

[1] In this respect, the idea of a substrate expresses a commonplace about the structure of scientific knowledge, the idea that in some sense biology "reduces" to chemistry, chemistry to physics, etc., an idea discussed in depth in Nagel (1961, chapter 11).

More technically, what I am calling a substrate has been been addressed in the philosophy of science literature on emergent properties. Emergence is "the view that there are features of the world – objects, properties, laws, perhaps other things – that are manifested as a result of the existence of other, usually more basic, entities but that cannot be completely reduced to those other entities" (Humphreys, 2006). The general question of substrate evolvability is just to ask what are the basic features of the world necessary for the emergence of rich evolution?

In the technical literature on emergence, the refinement is in the discussion of the exact sense in which one claims a set of properties "reduces" to another set. Different claims lead to different ideas of emergence, such as ontological emergence, epistemological emergence, conceptual emergence, nomological emergence, etc. Many species of emergence have been described, and I do not propose to delve into these nuances. The key point is that although emergence has been sometimes associated with vague and radical claims about new kinds of science, there are also clear and sober discussions of the idea. It is perfectly compatible with the orthodox scientific metaphysics of reductionist physicalism to suppose that we can identify some properties of a system as emerging from others, creating virtually distinct levels of causal interactions. In fact, that is more or less implied by the basic generalisation of treating evolution as a single process that has applied to various different species over time.

is that it is really not too abstract. It is no more abstract than evolution itself. Since evolution by natural selection is fundamentally a logical process, it is quite reasonable to ask about the logical preconditions for the process. These are variation, selection, and inheritance, and also an evolvable substrate – whatever that might require.

Another reason is that the question of substrate evolvability is the *de facto* perspective underlying a few areas of research which consider specific concrete substrates – origin-of-life research, astrobiology, and evolutionary computation. The following sections will briefly review work in these areas in order to show what has been learned about substrate evolvability in these specific cases, and what this suggests about substrate evolvability in general and for less radical notions of evolvability.

2.2.2 Origin-of-life and astrobiology

All life on earth is built on terrestrial chemistry. Knowing this, it is easy to imagine more exact questions. First, we can ask about the *degree of evolvability*. Terrestrial chemistry is evolvable but is it very evolvable? In other words, was life a likely thing or just a lucky thing? Second, we can ask about *comparisons of evolvability*. For instance, is terrestrial chemistry uniquely evolvable compared to all other chemistries? Would life be more likely, less likely, or possible at all under different chemical circumstances?

The question of the degree of evolvability has been touched on indirectly in research on the origin of life. This work aims to reconstruct the physical and chemical conditions of the primeval earth and then to deduce the history that starts from those ingredients and leads to self-replicating molecules and then onward to life as we know it. How evolvable was the primeval earth? It is of course extremely different to prove anything about ancient events. But there are a few reasons to believe the original origin of life was quite likely. For one thing, bioactive compounds and even cell-like structures are produced surprisingly easily from basic precursors (Miller, 1953; Fox, 1965). Second, research has suggested that life appeared as soon as it conceivably could do so (Dartnell, 2007). Also, origin-of-life theories have suggested ways that even the most basic chemical self-replication can lead to complex evolutionary phenomena such as strategic cooperation, paving the way for higher levels of organisation (Maynard-Smith and Szathmary, 1995). All of this suggests that early conditions were highly evolvable and that life of some kind was likely.

On the other hand, it is also true that for most of earth's history life was comparatively

simple. It took at least two thousand million years for the emergence of complex multicellular life, which arrived relatively late on the scene. It may be that while life itself is likely and relatively common in our galaxy, complex life is quite rare, since it depends on certain critical innovation like multicellularity that are themselves very improbable (Bounama et al., 2007; Ward and Brownlee, 1999; Dartnell, 2007).

As regards comparisons of the evolvability of terrestrial chemistry with other alternatives, the potential uniqueness of terrestrial chemistry has been analysed in the nascent field of astrobiology. This field studies the possibilities for life in environments outside earth. It may be there is nothing unique about terrestrial biochemistry. Bains (2004) argues that life need not be based on the familiar terrestrial biochemistry of carbon, hydrogen, oxygen, and nitrogen. Instead, the key requirement for a biochemistry is that it allows stable physical structure for bodies, complex coding structures for inheritance, and, for the chemistry itself, a highly flexible and specific system of catalysts. Assuming that this requires a liquid chemistry, and considering the galactic abundances and thermodynamic properties of various elements, he argues for the viability of a few alternative biochemistries, such as one based on silicon and liquid nitrogen or methane.

Yet another reason to doubt that terrestrial chemistry is special is how frequently one would need to expand the parameters of "terrestrial" for it to remain true. The last few decades have seen the unexpected discovery of many *extremophiles*, organisms that thrive at extremes of temperature, pressure, and acidity – at temperatures as low as -20 C in the Antarctic, or as high as 150 C near deep sea vents, or even buried in rock kilometres underground without access to light, oxygen, or organic molecules of any kind (Dartnell, 2007). Some of these environments are shared with other planets in the solar system. Few would have been anticipated to support life before the discovery. (And indeed, who would have guessed there were such things as microbes before the invention of the microscope?)

While it may seen incredible that there are alternative biochemistries, the argument that such alternatives are impossible is mainly based on how hard it is to imagine them. But in fact, little work has been done to evaluate these possibilities. And given the years it took to unravel the principles of terrestrial biochemistry, even with the convenience of being able to observe it, it seems overconfident to infer much from how hard it is to describe an alternative biochemistry when guided by imagination alone. Lack of imagination is not in itself an argument. Or if it is an argument, it is the venerable "argument from incredulity",

usually used to argue against evolution itself, as memorably described by Dawkins (1986, p 38).[2]

These two areas, astrobiology and origin-of-life work, represent research into concrete cases of the general question of substrate evolvability. Clearly, both areas face enormous methodological challenges. They address very remote circumstances which cannot be experimentally reproduced, varied, or observed over appropriate durations. But in both areas the problem is that the question of substrate evolvability is extremely difficult to approach empirically – not that it is self-contradictory, meaninglessly vague, or otherwise incoherent.

These methodological challenges do not apply in another research area, evolutionary computation.

2.2.3 Evolutionary computation

Evolutionary computation includes the fields of genetic algorithms and artificial life.

The idea of genetic algorithms is to harness the creative power of evolution to solve difficult engineering problems. The first step is to create a representation that can describe every possible solution (for instance, a set of parameters describing every possible shape of an aeroplane wing). A fitness function is then defined to assign a measure of quality to every possible solution (for instance, the lift coefficient of a wing). Then, over successive generations, solutions are mutated, sexually recombined, and reproduced in proportion to their fitness, until an adequate solution is found. Evolution acts as a kind of search for finding the best solution.

This perspective is not as restrictive as it sounds, since almost any problem can be framed as a search process. A solution can be as simple as a handful of parameters over clearly defined ranges (such as those specifying a wing shape), or as complex as a set of instructions for building a structure.

Since their invention three decades ago (Rechenberg, 1973), genetic algorithms have improved to the point where they are used in a variety of commercial applications, such as airline scheduling, drug discovery, and flavor synthesis (Walter, 2003). However, current best practices depend not on mere variation, inheritance, and selection, but on a great

[2] In fact, opponents of evolution are not strangers to this topic. After Darwin undermined the argument for God based on the perfection of the species, ingenious natural theologians like James Maurice Wilson, the archdeacon of Manchester, could still make the same arguments based on the providential qualities of water (Wilson, 1881).

deal of artful fine-tuning (Mitchell and Taylor, 1999). For instance, much depends on the particular data structure chosen to represent solutions. Some data structures make some solutions easier to evolve, just as the dominance or recessiveness of a gene has effects on its prevalence which are independent of its associated phenotypic trait. This fine-tuning amounts to manipulating the substrate on which the evolution occurs, trying to make that substrate more evolvable. Evolution alone is not automatically productive.

Artificial life is a more speculative field which uses simulations to explore the process of evolution itself. It has produced similar results. Ray (1992) introduced the Tierra system, which established the prototype for much later work. Small snippets of code in a small computer language represent the genes of "digital organisms". These organisms compete to acquire resources within their computational world or to perform some predefined task. The more successful organisms produce more offspring, subject to mutation, leading to evolution within the virtual world. This design has been continued in Tierra's closest successor, Avida, a mature piece of software which has provided a common basis for a coherent body of research (Adami and Brown, 1994; Ofria and Wilke, 2004; Lenski et al., 2003). Such a system can be thought of as microscope for the controlled study of general dynamics in evolutionary theory (Wilke and Adami, 2002). However, none of this work has realised the original, bolder hope of the field of artificial life, which was to reproduce "automatically" the more singular and dramatic evolutionary phenomena such as complex organisms and persistent, open-ended innovation.

Instead of a constant change producing increasing variety and complexity, the systems show an evolution that proceeds incrementally and then stabilises. One can identify simple evolutionary phenomena in retrospect, such as the parasitism that appears in Tierra, but these results are of a different order from the rich evolution of biology. In other words, the evolution in alife (artificial life) has never reproduced the phenomena that make biological evolution a topic of fundamental scientific interest. Rennard (2004) provides an excellent summary of these limitations and difficulties. (See appendix 2.5 for further discussions.)

This is despite that these systems meet the three conditions supposedly sufficient for evolution and despite significant experimentation in the basic genetic language and in the simulated world of their organisms (Spector and Robinson, 2002; Pargellis, 1996a,b). Work in artificial life is now centred on the more modest goal of describing this failure exactly and determining which conditions beyond mere selection, variation, and inheritance are

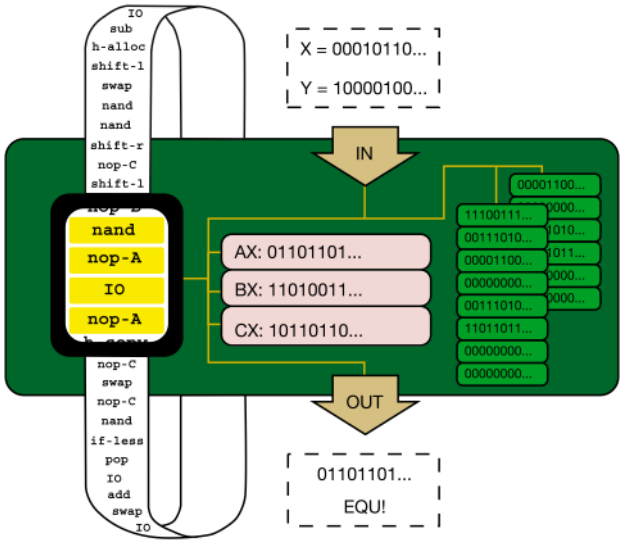


Figure 2.1: "Each individual organism has a circular genome and a virtual CPU with two stacks and three registers that hold 32-bit strings. Execution of the genomic program generates a computational metabolism...." (Lenski et al., 2003, also, figure) Organisms eat numbers from the environment and excrete computed outputs into the environment. This paper traces the genealogy of evolution under artificial selection for a particularly complex computation requiring many subcomputations, the 32-bit equals function EQU.

necessary to produce biological variety. For instance, Ofria et al. (2002) attempt a direct quantitative comparison of the evolvability of different alife languages. He considers language features such as template matching rules and size of symbol alphabet. This kind of work represents a concrete step in measuring substrate evolvability in avida-like systems. Other efforts have proposed quantitative measures of the degree of innovation demonstrated by an evolutionary system (Bedau et al., 1997, 1998; Bedau and Brown, 1999; Channon, 2001). These may be thought of as external metrics of evolvability. More evolvable systems should produce evolutionary activity that scores higher.

2.2.4 Implications

In summary, research in astrobiology has raised the tantalising possibility that evolution might not require physical ingredients we take for granted, like water. But work in evolutionary computation indicates that evolution does require logical conditions which we take for granted, but which it is extremely hard to specify. What does all this work imply about substrate evolvability in general, or about how to approach the less radical notion of organismic evolvability?

Substrate evolvability is important, neglected, challenging

The general question of "what makes a substrate evolvable?" can seem vague, since the idea of a substrate occupies a higher level of abstraction than is typical in biology. But as these concrete cases illustrate, the question is coherent and valid. It is just ignored by existing theory.

This is peculiar, considering that the question "what is necessary for evolution?" is of profound scientific interest. The scientific community has spent millions exploring the most fundamental basis of physical matter but has ignored the fundamental basis of evolution, which lies in the set of logical conditions that define evolution as a dynamic process. Yet without understanding those conditions, we do not really understand evolution itself or the living world that evolution has produced. Instead we have only half an explanation – a story which ties together numerous observations under the heading of "evolution", but then treats that in some ways as a magic word, never going deep enough to describe what causes the wonderful process. Of course every explanation must have some stopping point, some claim which is taken for granted. But as regards evolution, what is odd is that there is no consensus regarding where the explanation stops – nor even an avowed awareness of the fact that the theory does give up at a certain point.

One reason for this, surely, is that an explanation of logical conditions for evolution requires an axiomatic or *a priori* way of thinking which is unusual in biology. It requires not only measuring the evolvability of one particular substrate or another, but a theory describing which features of a substrate produce evolvability. In addition, this axiomatic analysis of the causes of evolvability will have to gibe with empirical research on particular substrates, research that itself poses significant practical difficulties as described above. Astrobiology is faraway, origin-of-life is ancient, and in both it is impossible to manipulate initial conditions over repeated trials.

This is what is especially interesting about the work in evolutionary computation, which is not subject to those constraints. As Wagner and Altenberg (1996) describes it, work in evolutionary computation exposes how traditional biology works within imaginative limits inherited from its methodological limits:

> Biologists are not confronted by this problem [of choosing an organism's genetic representation] because they study the end products of evolution, which are prima facie evidence that the favourable mutations have occurred at a suffi-

> cient rate. Furthermore, a biologists wanting to study this question faces great methodological hurdles; comparative and experimental approaches to the problem are blocked because one cannot simply pick alternate genetic systems that produce the same phenotype and compare their capabilities to produce adaptive variation. In evolutionary computation, however, this is possible.

Computational insights need biological inspiration

If the methods of evolutionary computation are uniquely suited for studying evolvability, then what has that work taught about evolvability? It is hard to find a clear lesson.

The main problem with the alife models is that they are too general. Those models started with the assumption that evolution requires only variation, selection, and inheritance. As a result, they did not try to model specific features of specific biological organisms. The failure of these models has shown, instructively, that this was a false assumption. For alife systems to produce more biological results and insights, they will probably need to draw on more biological designs and inspirations.

For instance it seems significant that it was a biological idea which provided the crucial breakthrough in Tierra's design (developing ideas that appeared earlier in coreworld, in Rasmussen et al. (1990)). This was the choice of how one part of the genome would refer to other genes in the genomic program language. In a typical computer assembly language, a piece of code refers to another piece of code by its location in memory (i.e., memory-addressing). In Tierra, a piece of the genome refers to another piece by providing an inverse template of it (i.e., template-matching). This was directly inspired by the sequence-matching that underlies the action of DNA and RNA. Ray, who was trained as a biologist, describes how this more flexible method quickly jump-started the evolution within the system (Ray, 1992).

Similarly, it seems significant that later systems like avida, which have failed to produce evolution much richer than what is found in Tierra, have not taken more ideas from biology. Biology provides the only actual example of rich evolution so it is the natural place to look for ideas, but this does require a deeper engagement with the biological literature.

And of course the whole idea of genomic program code – the genotype as a literal set of computer instructions, and the phenotype as the resulting set of behaviours – is quite unlike a typical biological organism, where the phenotype is a physical structure. To make them more biological, it might be useful to try alife models where the phenotype was not so

purely logical, where it was the result of an observable process of development. The more closely that developmental process resembled known features of biological development, the more it could teach us about how biological development affects evolution.

The challenge here would be to do this in a way that preserved the flexibility of the genomic program codes, which can in principle produce any kind of computational behaviour, i.e., an unlimited range of possible evolutionary products. The danger would be of fixing too rigid a system of development or allowing too little scope for the resulting phenotypes to interact with each other or the environment. This would produce a model like the biomorphs, which can never illustrate the evolution of complexity since it is too limited to allow even the manual construction of complexity. (This issue is discussed in more depth in appendix 2.5.)

But again, even in approaching this challenge, the most natural place to start looking would be in existing biological systems. In short, simulation does not provide a royal road to evolvability. General truths will not come from general principles, but from studying specific, existing, biological examples.

Focusing on the genotype-phenotype map

What are the best biological examples to start with? As discussed in chapter 1, the richest set of concrete arguments about evolvability lies in the literature on organismic evolvability. This literature discusses how specific mechanisms in the developmental architecture increase evolvability. So those mechanisms are the most concrete starting point for developing formal or computational models of evolvability which could lead to a more general understanding.

Also, arguments about development can be analysed in terms of the idea of the genotype-phenotype map. In discussions of biological development., the idea of the map is familiar as a loose simplification. In evolutionary computation, it is concrete idea which can be defined precisely. The map is a formal construct which can therefore serve as a natural bridge connecting work in formal and computational models and work with actual biological systems. Wagner and Altenberg (1996) argue extensively for its importance, and point out how the map is central both to the biological problem of evolvability, and to the computational work on genetic algorithms, which uses evolution as a kind of search process.

The map provides a clear formalisation for comparing the evolvability of different

organisms and understanding organismic evolvability. In contrast, there is no such clear formalisation at all for comparing the evolvability of different substrates and understanding substrate evolvability. What sort of formal construct could be used to represent terrestrial chemistry on one hand, an exotic extraterrestrial chemistry on the other, as well as the basic dynamics of an alife simulation? While the idea of a "substrate" is coherent, it is not easy to specify a substrate explicitly. The map is just a mathematical function – two sets, and a set of associations. Formalising a substrate would require a specification of all the entities a substrate allows, as well as a complete description of the dynamic laws governing how they interact, combine, separate, etc. Such a thing is not unimaginable but it is not easy.

The closest thing to this is the interesting work on artificial chemistry. In this area, workers try to understand the ramifications of very simple chemistry-like systems designed to reproduce the essential potentialities of biological chemistry while omitting unnecessary details. Dittrich et al. (2001) provides an excellent review of this work, which is very interesting but not mature enough to present clear implications for life processes on top of those chemistries.

Last, it should be noted that the significance of substrate evolvability will depend greatly on which substrate is being analysed. By definition, a substrate is the most basic set of rules defining what things can exist within a system and how those things interact. In the real world, organisms are complex structures built on top of organic chemistry, which is in turn built on top of nuclear physics, etc. So to speak about an organism's genotype-phenotype is already to describe things at a level of abstraction substantially removed from the underlying physics.

However, with the exception of work in artificial chemistry, models in evolutionary computation are designed to abstract away the physical basis of biology. Alife simulations are so simple that the genotype-phenotype mapping – the rule which translates the genes of a digital organism to its phenotype – is part of the bedrock definition of the simulated reality itself. In the avida system, for instance, there is no model of a chemistry underlying the digital organisms. It is a world made out of "genes" and with organisms, but no atoms. As a result, the substrate evolvability of an alife system might tell us more about organismic evolvability in the real world than about substrate evolvability in the real world.

So the genotype-phenotype map, in addition to being the best tool for analysing real-world organismic evolvability, will also be the best option for analysing many simulation

systems. This is one more reason why it seems like the most promising focus of formal efforts.

2.3 Individual fitness evolvability

Fitness is a fundamental idea in parts of evolutionary theory, and a great deal of work has gone into considering the subtleties of the concept (Dawkins, 1982; Rosenberg and Bouchard, 2002; Maclaurin, 2001). Equally fundamental is the methodological focus on the individual organism as a principal unit of selection. So it is natural to try to define evolvability by focusing on the fitness of individuals. This leads to what one might call *individual fitness evolvability*:

Definition: individual fitness evolvability

> Evolvability is a property of an individual organism. It depends entirely on fitness, in the sense that statements about evolvability reduce to statements about the fitness of individuals.
>
> Along these lines, the evolvability of an individual might be defined via a variety of measures:
>
> 1. the expected variance of the fitness of its offspring, prior to selection;
> 2. the expected maximum fitness of its offspring, prior to selection;
> 3. the expected fitness of its offspring after selection (or of some future generation). (Turney, 1999)
> 4. the likelihood that the offspring generation includes an individual of higher fitness (Altenberg, 1995a; Smith et al., 2002)[3]

These definitions are different on the surface but very close in spirit. Not only are they all centred on the fitness of individuals, but they all spring from the same implicit story of the evolution of evolvability. When that story is true, all these measures track each other.

Recall that one key idea in evolvability has always been the concern that evolvability is unDarwinian – that is, that any story of the evolution of evolvability will require mechanisms which are implicitly teleological, violating the scientific physicalism of Darwin's

[3] To be exact, Altenberg defines it as the likelihood of an individual reaching a phenotype with improved fitness, a definition which tracks with the ones above. But this is in the context of hill-climbers on fitness landscapes, which amounts to the same thing.

theory (see section 1.2). Considering the story which motivates these definitions is an excellent way to address this concern.

In fact, the main value of these definitions is that they clarify the key ideas of evolvability by presenting them in the familiar context of individuals and fitness. The following sections will review these definitions by revisiting some of those key ideas, starting with the argument of Darwinian incompatibility, then addressing the structure and meaning of variability.

2.3.1 Key ideas of evolvability, formally revisited

Darwinian compatibility

All these definitions correlate with each other if the expected story about the evolution of evolvability holds true. The story has one protagonist, fitness. Fitness is conceived simply as fecundity – i.e., the expected number of offspring, as determined by heritable properties of the phenotype. That story goes as follows.

Under a typical distribution of offspring fitness before selection, a greater variance in fitness will imply a greater maximum in fitness (definition 1 implies 2). Since selection will remove the low fitness individuals, a higher maximum fitness of offspring before selection (or just higher proportion of high fitness offspring) will imply a higher average fitness after selection (2 implies 3). As a result, a greater expected variance in the fitness of an individual's offspring before selection will lead to a greater expected fitness of those offspring after selection (thus, 1 implies 3). Over time, as this effect accumulates, higher evolvability will tend to produce higher fitness.

Since we are discussing evolutionary change the assumption is that we are considering conditions of directional rather than purifying selection. That is, some mutations might increase fitness. Under such conditions, the higher the maximum offspring fitness, the more likely that the maximum offspring fitness is higher than the parent's own fitness. (So 1-3 imply 4.)

To put it simply, evolvability is about the *offspring fitness distribution*. Claims for the importance of evolvability are basically claims that the expected offspring fitness distribution matters. This is because selection itself, working over generations, will interact with such distributions to increase the absolute level of future fitness.

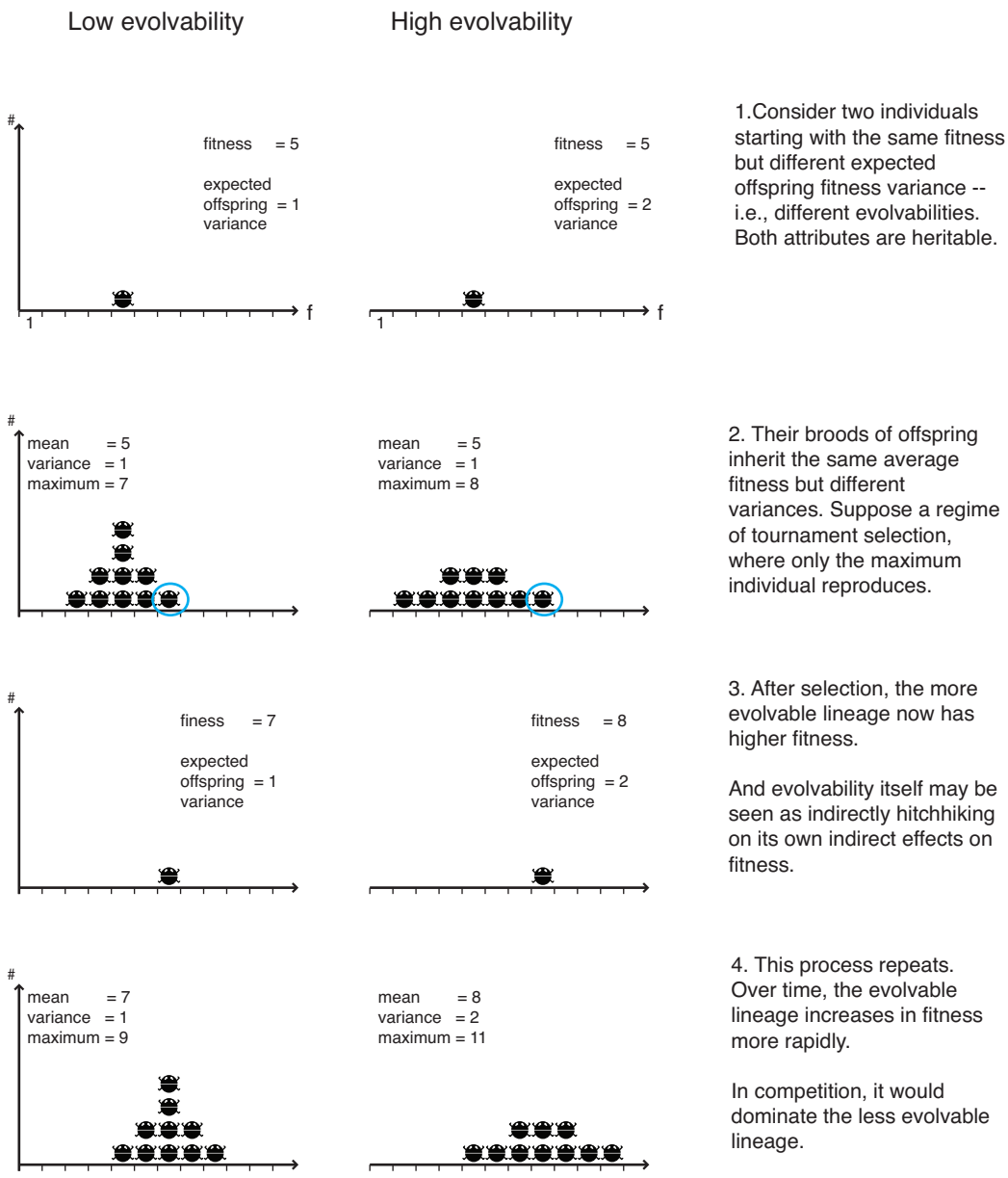


Figure 2.2: Models of evolvability as a property of individuals clarify the basic logic of the evolution of evolvability: if fitness and offspring fitness variance – evolvability – are separate heritable properties, then selection itself will indirectly promote evolvability by selecting for fitness. In a sense, evolvability hitchhikes on fitness.

This story makes many assumptions. The evolution of evolvability requires that evolvability is heritable, so this story assumes that expected offspring fitness variance is heritable. In addition, it assumes the selection regime is strong enough and the typical range of a fitness distribution is large enough to support the evolution of evolvability. And it assumes innumerable other factors can be safely neglected: interactions between individuals, dif-

ferential impact of the environment or changes in the environment, sexual recombination, purely stochastic variation in survival, etc.. However, the argument that the evolution of evolvability is incompatible with Darwinian selection is not based on arguing that this story is too simple, by challenging one of these assumptions. The argument is that even with these assumptions the evolution of evolvability would still be impossible, because the basic story is logically incorrect.

Although this story is quite straightforward the argument against its Darwinian compatibility can also seem irrefutably clear, so it is worth identifying where it goes wrong. The nub of that argument is that, since Darwinian selection sees only immediate fitness, it will not in general tend to promote a property like evolvability which is independent of immediate fitness. But equating evolvability with offspring fitness variance resolves this problem. It is true that an individual's evolvability is not correlated to its own fitness by any kind of definitional *logical necessity*, since offspring fitness variation need have no particular relation to an individual's own fitness. Yet evolvability will still lead to a greater fitness over time because of how selection favours the tops of fitness distributions – because of *predictable indirect effects* of selection itself.

Another way to see this dynamic is in terms of genetic hitchhiking. This was understood as early as 1979, when Conrad (1979) argues that "the structural features which increase evolutionary amenability [i.e., evolvability] ... accumulate in the course of evolution as a consequence of hitchhiking along with desirable traits whose evolution they make possible". Or yet another perspective is to say, following Sober (1984), that direct selection *for* higher fitness will produce an indirect selection *of* evolvability, since individuals that inherit an unusually high fitness will also inherit the high evolvability that led them to have a high fitness (Sober, 1984; Buller, 1999).[4]

[4]It is instructive to consider the subtleties in applying Sober's distinction here. In his original distinction of selection-of versus selection-for a trait, Sober emphasises that what matters is the trait's weight in the causal history of the selection event. We say there was selection *for* a trait if the trait was responsible for an individual's reproduction. And could we not say that evolvability is responsible for the parent producing the high fitness offspring?

The trouble here is that Sober was thinking of the evolution of normal phenotypic traits, which have immediate effects on an individual's own fitness. It is hard to apply this rule to evolvability because evolvability affects fitness in various *indirect* ways which make it unclear how much causal "credit" it should get. First, evolvability does not simply increase fitness but rather affects the distribution of fitness, thus causing high fitness as well as low fitness individuals. Second, evolvability does not impose these effects on the fitness of the evolvable individual itself, but on the offspring generation. And last, evolvability does not work alone – it also requires the action of selection for its effect on offspring distributions to produce an increase in fitness over generations.

In other words, selection promotes fitness directly, acting on one generation at a time. Selection promotes evolvability indirectly, as a side-effect of the direct selection for fitness on the offspring generation. All these complications mean that evolvability is not a direct cause of fitness like an ordinary phenotypic trait. As a result, it is more accurate to say this is a result of a selection *of* evolvability.

The key assumption here is that fitness and evolvability are separably heritable. This sheds light on another key idea in evolvability – the structure of variability, and the distinction between variability and variation.

Structure and meaning of variability

In this story the *structure* of variability becomes very concrete – it is merely the shape of the offspring fitness distribution. A favourable structure of variability is, simply, just to have a distribution with a high variance, since this guarantees higher fitness survivors after selection. This concreteness lets us see the relevance and also the challenges of this idea more clearly. Characterising offspring with just one variable, fitness, lets us characterise the structure of variability with a single variable distribution, a familiar construct. If we characterise offspring phenotypes more fully, it will be harder to describe the variability and to identify which of its aspects promote evolvability.

Even more illuminating, however, is what is shown about variability as distinct from variation. In this story variability is the *expected* distribution of offspring fitness, which is summarised by its variance. Of course an individual's actual offspring fitness variation will define the material available for selection in the next generation, so variation always affects the make-up of an individual's ultimate descendants. This is obvious. But as long as this offspring variation is unpredictable, and not a property of the parent individual, then it is always a contingent and accidental factor on the make-up of descendants, and thus seemingly not worth mentioning in any explanation.

Once we shift to a conception of an individual which makes it meaningful to speak of an individual's expected offspring fitness variance (i.e. variability) as a real, distinct, heritable property of the individual, then it can ground a deeper kind of explanation of the individual's evolutionary history. Then one can say, looking forward, that *because* individual A has a higher variability than B, the later descendants of A will probably have a higher fitness – or, looking back, that the *reason* existing individuals have such a high fitness is because their ancestor A had such a high evolvability. If evolvability cannot explain things in this way, then it does not say much. That is why the crucial requirement here is that an individual's offspring fitness variance is a heritable property of its own, separate from fitness.

If this requirement seems arbitrary, that is probably because the abstraction of the

individual fitness perspective obscures the essential issue, which is the role of development. The idea that an individual has an expected offspring fitness variance which is real, heritable, and distinct from its fitness, corresponds to the commonsense idea that an organism has a developmental architecture which is real, heritable, and distinct from its phenotypic traits. Multiple possible developmental architecture could produce the same phenotypes. But these developmental architectures will produce different distributions and variances of offspring fitness, and we need to capture those differences to describe evolvability.

2.3.2 Hypermutation: a case study

This conception of individual fitness evolvability tends to be implied in nearly ever discussion of evolvability, as it is conjured up by any casual use of the word fitness (e.g. Dawkins, 1989; Altenberg, 1994; Wagner and Altenberg, 1996). However, explicit treatments of the idea, like explicit models of evolvability in general, are more rare. The idea is sometimes applied in artificial life, but it has appeared most in the context of computer models and microbial models which deliberately address the evolution of evolvability. These models study what is perhaps the simplest possible case of evolvability – hypermutation.

Hypermutation is the evolution of the mutation rate itself. A higher mutation rate will lead to a greater variety of offspring phenotypes and thus a higher offspring fitness variance. So any mechanism that affected the mutation rate – such as, for example, proofreading mechanisms in the gene transcription machinery – is a kind of developmental feature that drives offspring variance, i.e., evolvability. In this case the proposed story about the evolution of evolvability reduces to two specific claims. First is the claim that this evolvability mechanism is beneficial – that increasing the effective mutation rate in response to certain conditions speeds the process of adaptation. Second is the claim that this evolvability mechanism itself has evolved – that when an organism starts mutating more it is *because* selection has promoted a specific mechanism in the organism, like bad proofreading, for boosting the effective mutation rate.

Various microbial and computational studies of hypermutation address this first claim (see True and Lindquist (2000); Pál (2001); Burch and Chao (2000); and Poole et al. (2003), who provides a very rich comparison of prokaryotes and eukaryotes). Computational models validate the story of the evolution of evolvability by demonstrating in detail that the proposed dynamics work as described. The empirical studies, in yeast and other

microbes, demonstrate that this dynamic is in fact visible in real organisms. Turney (1999) presents an explicit computational model of the evolution of evolvability. Admitting that "It is difficult to define evolvability beyond saying it is the capacity to evolve", he proceeds as follows:

> We suggest the following sufficient (but not necessary) condition for evolvability: If individuals A and B are equally fit but the fittest child of A is likely to be more fit than the fittest child of B, then A is more evolvable than B.

In laboratory work Burch and Chao (2000) coincidentally use Turney's formulation almost exactly in their study of the the RNA bacteriophage $\phi 6$. Burch compared the evolvabilities of two homogeneous populations A and B. Both populations had repaired a deleterious mutation in their common ancestor. However, each population had repaired the damage differently – A through a back mutation, B through a new mutation. Burch documents that after one hundred generations, under identical selective pressures, the ancestors of A achieved a higher fitness than the ancestors of B, showing that A had greater evolvability.

Burch's system reproduces the essence of Turney's theoretical conditions. By ensuring that A and B had both repaired the same ancestral damage, she establishes that they are identical in phenotype and presumably fitness. But because they used different repairs, she knows they have different genotypes. By repeating the experimental with clones, she establishes that the result was a consequence of this genotypic difference, rather than a random effect. It is true that Burch measured the fitness of remote descendants of a homogeneous population, while Turney and Smith discussed the fitness of the immediate offspring of an individual. But this is beside the point. Burch used a large population over multiple generations as a practical necessity, in order to amplify an individual, single-generation effect to the point of visibility. The core idea is still variation of offspring fitness. (See also Landry et al. (2007), discussed in section 2.4.4, for another use of mutational accumulation lines.)

These models of hypermutation support the first claim that evolvability is beneficial, and they should completely suffice to rebut the argument that evolvability is unDarwinian. However, it is more problematic to say whether they support the second claim that hypermutation is a case where evolvability itself has evolved.

To show the evolution of evolvability, it is necessary to show that the hypermutation

response itself is an evolved capacity. But this is very hard, since it requires distinguishing between an increased mutation rate which indicates a sophisticated evolved capacity to facilitate evolution and an increased mutation rate which is merely the accidental (but possibly beneficial) effect of damage. One way to make this distinction would be if the increased mutation rate were localised to particular areas of the genome, those less likely to disturb critical function and more likely to produce adaptations specific to the form of stress. Another way would be to perform a detailed analysis of the internal metabolic mechanism of the hypermutation itself, but such an analysis would spoil the original appeal of a model which requires looking at nothing but fitness and mutation rates.

But the deeper problem with hypermutation – like individual fitness evolvability itself – is that it is too simple to imply much about the connection between evolvability and organismic complexity.

2.3.3 Ignores phenotypic development and complexity

Defining evolvability in terms of individual fitness provides an exact form for tackling a couple of key ideas – the structure of variability, and the concern of unDarwinism. But this simplification has a price. Two other key ideas become impossible to approach – how the developmental architecture promotes evolvability, and how the same mechanisms might promote the evolution of complexity.

One cannot study the issues of phenotypic development and phenotypic complexity by comparing individuals. While development drives differences in individuals, the differences relevant to evolvability are between kinds of organisms. One is interested in comparing earthworms to fruit-flies, not one earthworm to its brother. This does not make the idea of individual fitness evolvability incoherent, but it indicates it is an unsuitable approximation for studying these issues.

Both these issues fundamentally concern the phenotype, the organism as a functioning physical system, while a modelling approach based on fitness has no notion of the phenotype. Instead of considering how development defines a mapping from genotype to phenotype ($\phi : G \to P$), and then how the intrinsic qualities of the phenotype and its interactions with the environment define another mapping from phenotype to fitness ($f : P \to [0, \infty]$), this approach short-circuits the key details by jumping directly to the composed map from genotype to fitness ($(f \circ \phi) : G \to [0, \infty]$).

In that respect this approach to evolvability is in the tradition of the fitness landscape models which assume only an idea of fitness and of possible genotypes. In fact in later work, Smith refers to his measures as "local evolvabilities", acknowledging their natural interpretation as local averages of fitness within a fitness landscape (Smith et al., 2003).

Landscape models can provide a valuable intuition for understanding probable evolutionary dynamics. But by ignoring the development of the individual's phenotype, the structure of the phenotype, the complexity (or non-complexity) of the phenotype, these models ignore all the factors that determine the shape of the landscape. Ignoring what conditions produce an evolvable landscape amounts to ignoring the *causes* of evolvability. So these models cannot pursue the widespread intuition, mentioned in chapter 1, that the causes of evolvability lies in a theory of organismal self-organisation (Dawkins, 1989; Wagner and Altenberg, 1996; Kirschner and Gerhart, 1998; Yang, 2001; Hansen and Wagner, 2001; Hansen, 2003; Pepper, 2003).

What this shows is that it is inadequate to *define* evolvability in terms of individual fitness. Evolvability will produce fitness consequences, and looking only at those consequences is a helpful approximation for understanding the role of Darwinian selection. But starting with fitness ignores the mechanisms that lead to those effects on fitness. Evolvability is connected to fitness, but it cannot be reduced to the idea of fitness.

2.3.4 Landscapes introduce euclidean distortion

It might still be hoped that fitness-based models allowed us to describe evolvability in terms of fitness landscapes, as introduced by Wright (1931) – that is, a three dimensional plot of a fitness function, resembling a physical landscape of hills and valleys.

The visual appeal of a fitness landscape has made it an enduringly popular, intuitive depiction of evolutionary processes (Dawkins, 1996). At the same time, its origins as a mathematical plot has made it a natural basis for abstracted, mathematical models of evolution, spawning variants such as rugged landscapes (Kauffman, 1993) and neutral holey landscapes (used for modelling speciation in Gavrilets (2004)). So just as hills, saddles, and other shapes give an intuitive picture for different kinds of evolutionary dynamics, might other shapes give an abstract but intuitive signature of types of evolvability?

Unfortunately, this is unlikely. The landscape visualisation is unavoidably misleading about facts crucial to evolvability. This is because it must squeeze the set of all possible

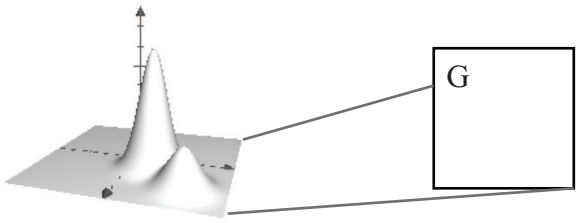


Figure 2.3: A fitness landscape is a function graph of fitness on the z-axis against genotype on the xy-plane. Every fitness landscape implies some *embedding*, a rule for placing every genotype within a patch of the Euclidean xy-plane. But the Euclidean plane has a different structure from genotype space, so nearly every embedding introduces a substantial *Euclidean distortion*.

genotypes, with their multidimensional mutational relationships – the genospace – into the flat two-dimensional patch of the Euclidean plane under its contours. Except for trivial cases, doing this will always misrepresent these relationships – introducing what one might call, *Euclidean distortion*.

Severity and inevitability of euclidean distortion

The description of the landscape as a "metaphor" indicates how these difficulties are casually appreciated, but it is worth explaining why this distortion is so severe and mathematically impossible to avoid.

The most obvious problem is that embedding the genospace into the plane suggests too few dimensions. It suggests that a genotype has only two independent axes of variation much like the north-south and east-west of a geographical landscape. But a mutation cannot be expressed as a combination of two basic types of mutation. Since a genotype is a string of symbols, every position is an independent "direction" of mutation. A string of length 10 has 10 dimensions of variation. As early as 1938, in a letter to E.B. Ford, Fisher recognized this and observed that "... so far as individuals are concerned, there is only a discontinuous aggregate of lattice points, each having its own selective value. There is no continuum of possible values in which we might speak of peaks or maxima" (Bennet, 1990, p 201). He restated the point in the published literature Fisher (1941).[5]

[5]This point that genotypes form a lattice- or graph-structured space is an essential problem with Wright's use of a landscape to depict the set of all genotypes.

However, this is separate from the criticism that Wright's own account was internally inconsistent, and that he "treated two different conceptions of the fitness surface as equivalent and interchangeable, to the extent of using the same diagram for one version as perfectly appropriate for the other" (Provine, 1986, p 314). In addition to treating landscape as representing different "gene combinations", i.e., genotypes, he also treated it as respresenting different

One might hope that an informed reader just remembers this caveat as she reads the landscape. But this solves nothing. The value of the fitness landscape is that it connects evolutionary dynamics to shapes like fitness valleys, hills, and plateaus. Those shapes will depend on the embedding, on the particular arrangement of the genotypes within the plane. So for a fitness landscape to mean anything, a reader cannot just remember the landscape is an approximation. She must make positive assumptions about the embedding – for instance, about what it means when two genotypes are near on the landscape.

This is the second problem, the fatal one: there is no set of assumptions a reader can make which meets even the most basic requirements. For instance, one essential requirement is that an embedding must preserve mutational relationships. When there are mutations that link one set of genotypes with another, then the corresponding areas on the plane should touch each other. This allows the shapes to show which phenotypic transitions are possible. Without this requirement, small mutations could produce large jumps from one part of the space to another. In the centre of every fitness valley, there might be an invisible tunnel leading directly to the peak of a distant fitness hill.

Unfortunately, as soon as we treat cases of more than four phenotypically distinct classes of genotypes, it is impossible to meet even this essential requirement because. In other words, it is not possible to embed the mutational structure of a genospace into the Euclidean plane. (I prove this in section 2.6. While this conclusion is not exactly surprising, it is odd that there is no published proof of it anywhere.) As a result fitness landscapes, which are based on this embedding, must misrepresent the mutational structure of genospace. They must misrepresent certain neighbouring genotypic regions as separated and give no hint about whether the route between them is one of ascending or descending fitness.

It may seem quibbling to belabour the flaws in what is acknowledged to be a metaphor. But this is a metaphor which begs to be read as something more substantial, as a plain functional plot. Yet functional plots deserve close study since they faithfully represent underlying constructs. The fitness landscape not only fails to represent higher dimensional structural features, it positively blinds us to them by focusing attention on a much simpler image. It has aided intuition about evolutionary dynamics through the ideas of local optima, hill-climbing behaviour, fitness saddles, shifting balance theory, but it has also tended to

"sets of gene frequencies" in the population. Later Simpson (1944) introduced a third conception, interpreting the landscape as representing different combinations of phenotypic character values.

create the impression these are the only issues in the field, as when Wright (1931) states that saddle-crossing is the "central problem in evolution".

If there is anything worth knowing about the true detailed multidimensional structure of genotype-phenotype mapping, if that structure has any effect on evolution, then the fitness landscape has passed from being an aid to clear thinking to being an obstacle. This is of course due to its enormous success – as the only widely known tool for visual reasoning about the dynamics of fitness, it has a monopoly on our imagination.

2.4 Individual mutational evolvability

Models of individual fitness evolvability clarify those issues that are specifically tied to fitness, such as how selection for fitness could indirectly promote evolvability. But by ignoring the phenotype, they obscure the fundamental question of how the development architecture could facilitate phenotypic change.

By contrast, focusing directly on phenotypic change leads to what we may call *individual mutational evolvability*.

Definition: individual mutational evolvability

> Evolvability is a property of an individual organism. It is defined by the mutational relationships between its possible mutant phenotypes.

This approach treats evolvability as a property of an individual, an assumption which is rather misleading but which allows very clear models (as argued in section 2.3.3). It then uses that clarity to address explicitly the idea of the phenotype, and the role of the developmental architecture in facilitating phenotypic change. While fitness can only go up or down, a phenotype can change in a number of ways. This is why focusing on the phenotype does not allow a single, univariate metric of evolvability (like, for instance, offspring fitness variance) and requires a more nuanced kinds of quantitative description. I will now review some of the work in this area and the concepts and methods this work has generated.

2.4.1 Evolvability and the genotype-phenotype mapping

The conceptual basis for this work is the *genotype-phenotype mapping*. Wagner and Altenberg (1996) argue for the value and centrality of this concept, which can be applied to biological development, to mathematical models, and to computational simulations.

The genotype-phenotype mapping is simply a formalisation of development in terms of a mathematical function, or map. Technically, a function, $\phi : G \to P$, is just a rule or procedure which associates each element in one set G, called the domain, with some element in another set, P, called the codomain or the range.[6] In the genotype-phenotype map, the domain G is the set of all possible genotypes and the codomain P is the set of all possible phenotypes. The map thus determines how genetic change translates to phenotypic change. Wagner and Altenberg (1996) argue that many traditional issues in evolutionary biology ("dissociability in development, morphological integration, developmental constraints, biological versatility, fluctuating asymmetry, the Baldwin effect, epistasis, canalization, heterochrony, genetic variance/covariance matrices, quantitative trait loci, and the adaptive landscape") may all be understood as investigations of the map, which can serve as a unifying conceptual framework for these efforts, and they relate this idea to the literature.

Thinking in terms of a genotype-phenotype map departs from the normal procedure in quantitative biology. The set P does not represent the possible values of a single measurable phenotypic trait, but literally all possible phenotypes; just as the set G does not represent the different allelic values of a particular gene, or set of genes, but instead all possible genotypes, all possible total genetic configurations. This is rather different from using a function to describe how a single trait, like height or disease resistance, depends on the allelic value of a particular set of protein-coding genes. By taking in all genetic information, the mapping can capture any kind of epistatic interactions. By taking in all possible phenotypic information, it may also be suitable for addressing dynamics in the evolution of complexity, such as the dependencies and relations between different parts or modules of the phenotype. All of this makes the map an appealing construct for analysing evolvability and the evolution of evolvability.

[6] Some writers use *codomain* to refer to the set of potential targets of the mapping, and *range* to refer to the subset of those elements actually addressed by the mapping. I do not observe this distinction.

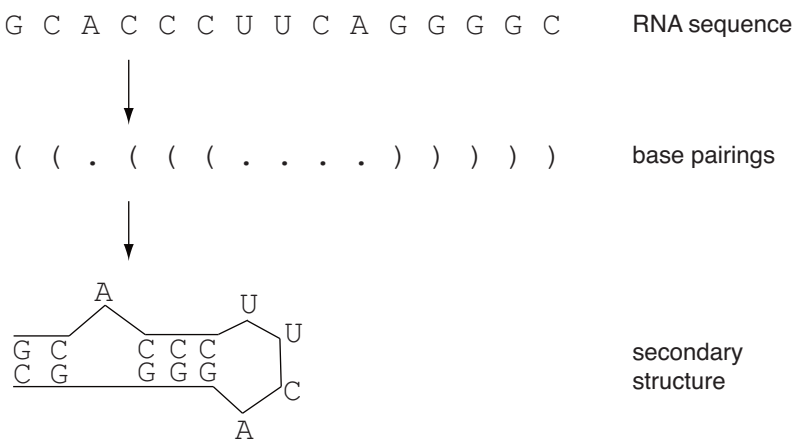


Figure 2.4: Nucleotides in RNA sequences form chemical bonds with each other, distorting the molecule from a straight strand into a folded shape known as its secondary structure. Since this shape itself is biologically active, it may be regarded as a kind of phenotype.

2.4.2 RNA secondary structure as a test system

The difficulty with the genotype-phenotype map is that it presumes total knowledge of all possible genotypes, all possible phenotypes, and their relations. This assumption makes it profoundly difficult to model any actual organism, since current research is far from understanding any organism's development in such detail.

However, there is one biological system – not exactly an organism – that can be analysed at this level, and studying this system has driven the conceptual advances in the detailed quantitative analysis of genotype-phenotype maps. This system is RNA. The idea is to consider all the RNA molecules of a certain length. The genotype of an RNA molecule is its actual nucleotide sequence. The "phenotype" of a sequence of RNA is the physical shape it folds into at a given temperature – or to be exact, just the secondary structure of its shape.

An RNA molecule's secondary structure is the aspect of its physical structure defined by normal electrostatic attraction bringing together nucleotides to form base pairs. These pairings produce typical structural motifs in the molecule itself, such as stacks, loops, joints, free ends, etc. Weaker chemical bonding forces also affect the final three dimensional folded shape of the molecule, the so-called tertiary structure. While tertiary structure remains unpredictable, a variety of probabilistic, thermodynamic, and experimental work has led to robust algorithms for predicting RNA secondary structure from sequence alone (Hofacker (2005) summarises this work; original research include Waterman and Smith (1978); Nussinov and Jacobson (1980); Zuker and Stiegler (1981); Zuker and Sankoff (1984)).

Therefore, if one is willing to see RNA secondary structure as a kind of phenotype, then RNA provides a truly biological model system – perhaps the only one – where one can claim a complete knowledge of all possible genotypes, all possible phenotypes, and the mapping that determines which genotypes lead to which phenotypes.

But what rationale is there for seeing RNA shape as a kind of phenotype? Much like the knots of string they resemble, RNA shapes are rather different from living things. They do not grow, repair themselves, or maintain a homoeostatic balance. They do not seek out mates or physical resources in the environment. They do not have a physical metabolism that transforms surrounding material into energy and useful products. Unlike even the virtual organisms of artificial life simulations, they do not compete with each other, they do not reproduce on their own, and they do not – to recall the original, underlying sense of organism – constitute a "a whole with interdependent parts" (OED, 2008).

Biological relevance of RNA

While the RNA molecule is not itself a kind of organism, it is crucial to organisms. In addition to relaying information in normal gene translation as messenger RNA (or mRNA), it is has recently been discovered to play a variety of roles regulating gene expression without simply coding for a protein, in the form of non-coding RNA (or ncRNA). ncRNA seems increasingly important, given the discovery that the human genome contains relatively few traditional, protein-coding genes (only about twenty thousand, fewer than some kinds of yeast). Perhaps ncRNA makes up a larger and more influential part of the human genome, and the genomes of other advanced metazoans, than was suspected.

The many kinds of ncRNA and their modes of action are far from completely understood. However, it is clear that in many of these non-protein coding roles, the RNA molecule's secondary structure is biologically crucial (even as this presents new computational challenges for gene-finding (Rivas and Eddy, 2000; Eddy, 2002)). One particularly interesting area is recent work on micro RNA, or miRNA.

miRNAs are very short strands of RNA, a little over twenty nucleotides in length, which regulate gene expression, including genes affecting development in plants and animals. They inhibit or promote normal gene expression through a variety of mechanisms which are still coming to light, e.g., loosely annealing with target mRNA to arrest translation directly, or directing exact cleavage via dedicated enzymes. miRNAs were first identified

as a distinct class of genetic elements in 2001 from their role in regulating development in *C. elegans* and *Drosophila*, as well as from parallels with the mechanism of RNA interference, or RNAi (Ruvkun, 2001; Baulcombe, 2002). Dozens of miRNAs were quickly identified in *C. elegans* alone. Many miRNAs are conserved between nematodes, flies, and mammals. They are relevant to development – Baulcombe (2002) for instance, includes an illustration of a mutant plant phenotype *Arabidopsis PHB* caused by mutations in a miRNA gene target.

How is RNA shape relevant? Mature miRNAs are generated from multiple processing steps, starting with longer so-called primary transcript RNA strands. This processing depends on the secondary structure of the intermediate RNA strands (Denli et al., 2004). Recent work on miRNA emphasises the importance of miRNA to development and human disease (Su and Zhang, 2008; Staton and Giraldez, 2008). Thus, within the organism, the large class of miRNAs presents one example where the RNA secondary structure is biologically consequential. It determines the specificity of interaction. Insofar as those reactions affects the fitness of the organisms, one may consider the evolution of RNA secondary structure under the same light as the evolution of any other organismal trait. Even if most mutations of RNA shape make a relatively minor contribution to the fitness of an organism as a whole, as long as that contribution is real then the folding algorithms present us with an example of a genotype-phenotype mapping which is empirically grounded, well-understood, and expressive of real biology.

The very short length of miRNAs gives them additional relevance for enquiries into evolvability. That length – roughly twenty nucleotides – puts them within the range of what may become feasible to study via existing methods of exhaustive computational folding studies, assuming modest progress. For instance, Rendel (2008) folds all sequences of length 16. The computational load increases exponentially with length, but further improvements in efficiency are not unthinkable. Also, the shortness of miRNAs might make them the perfect example of small regulatory mutations that have large phenotypic effects.

2.4.3 Conceptual advances from RNA-based research

As mentioned, work in the 1970s and 1980s developed algorithms for predicting the secondary structure of a given RNA sequence (Waterman and Smith, 1978; Nussinov and Jacobson, 1980; Zuker and Stiegler, 1981; Zuker and Sankoff, 1984; Turner et al., 1988; Jaeger

et al., 1989; He et al., 1991). In the 1990s researchers applied these folding algorithms to large numbers of sequences, calculated the the statistical properties of the resulting ensembles (Fontana et al., 1993), and began to consider the folding algorithm as a kind of genotype-phenotype map (Schuster et al., 1994). Workers used both sampling methods and exhaustive surveys for different kinds of statistical analysis (Schuster et al., 1994; Gruner et al., 1996a,b; Göbel, 2000).

Regarding the RNA folding algorithm as a genotype-phenotype map is a significant step. That perspective transforms RNA secondary structures from an esoteric topic in biochemistry into a springboard for a variety of general reflection on evolution and evolutionary dynamics, all warranted as empirical since RNA is a biological system. Researchers have used RNA to study the effects of neutrality on evolutionary dynamics (Huynen et al., 1996; Weber, 1997; van Nimwegen et al., 1999), phenotypic plasticity and genetic canalisation (Ancel and Fontana, 2000), and mutational robustness (Schuster and Fontana, 1999; Wilke, 2001). One especially ambitious line of research uses the RNA folding map to argue that traditional analytical methods are unsuitable for describing evolutionary change, which requires new mathematical formalisms based on topological spaces (Fontana and Schuster, 1998a,b; Cupal et al., 2000; Stadler et al., 2001, reviewed in chapter 3). Fontana (2002) reviews the findings from this research on RNA, arguing that the resulting concepts may prove useful for thinking about evolutionary developmental biology (evo devo) in general.

It is reasonable to question how much RNA implies about the evolution of organisms. However, whatever the correct scope of its implications, the availability of this map as an exact, computable, and convenient artifact has obliged people to think about the general idea of genotype-phenotype maps in greater detail, which has spurred a number of conceptual advances. Many of these concepts follow simply from the basic idea of a genotype-phenotype map, but were not explored in the absence of the concrete example provided by RNA.

These concepts are as follows.

Degeneracy of the mapping

The first point is that there is a very high degree of *degeneracy* (or redundancy) in the genotype-phenotype map. That is, there are many more RNA sequences (genotypes) than RNA shapes (phenotypes), since many different RNA sequences fold into the same RNA

shapes.

This fact was implied by, although largely peripheral to, the discussions in the earlier literature on genetic "neutrality" (reviewed in Takahata, 2007). When two sequences which yield the same shape differ by a single locus then one says that the locus is neutral. Since a mutation on that locus would have no effect on the phenotype, it would be a neutral mutation. Much original work on neutrality aimed to determine the prevalence of neutral loci in the genomes of mature organisms, in order to gauge the significance of drift compared to selection, and to develop methods for estimating the ages of lineages based on the accumulation of neutral mutations (molecular clocks). This work tended to treat neutrality as a property of the locus itself.

In contrast the idea of degeneracy has come out of research on the genotype-phenotype map as a whole, which looks at the genome as a whole. This work has tended to see degeneracy as an inevitable property of genotype-phenotype mappings as such, a property which would apply to entire organisms as well as RNA shapes. Why is degeneracy the norm? The set of possible genotypes is large. In essence you generate the set just by counting upward in a nucleotide number system. But the set of resulting phenotypes is not just a matter of counting. It is circumscribed by the actual constraints which shape phenotypes in the real world, whether those are the biochemical rules of RNA folding, or the developmental and engineering constraints that define what kinds of organisms are possible. This mismatch will always mean there are more genotypes than phenotypes.

When one sees this degeneracy as a feature of the relationship between genotypes and phenotypes, it naturally invites statistical questions that summarise this relationship. How many sequences go to a certain shape? Which shapes have the most sequences or the fewest? What does the frequency distribution look like? All of the early statistical investigations consider these questions.

Mutational relationships

But genotypes also have relationships to each other, via mutation. This defines a kind of spatial relationship, by which one sequence can neighbour another, or be "farther" away from it by various degrees in a notional space of genotypes. These two structures together – the mutational structure within genotype space, and the relational structure of the genotype-phenotype mapping – produce a rich structure that can ground many kinds

of analysis.

This leads to the second main finding – the value of thinking about genotypes and phenotypes in terms of their detailed mutational relationships, relationships which it would not be possible to define without a total knowledge of the genotype-phenotype mapping. For instance, of a given genotype, you can now ask which phenotypes it can mutate into directly – or in other words, what is the *phenotypic neighbourhood* of that genotype? Moreover, the mutational relationships between genotypes produce indirect relationships between phenotypes. So you can also ask about the neighbourhood of a given phenotype. The papers studying the RNA genotype-phenotype map are led very naturally to develop these concepts. They allow a kind of structural, qualitative characterisation of the genotype-phenotype map. This kind of detailed description of the mutational opportunities around a given genotype is referred to as the evolvability of that genotype. This is basically a phenotype-oriented analogue to the fitness-oriented definitions of evolvability that also focused on an individual's mutational environment.

Neutral networks

The last main concept to come out of the RNA investigations is the idea of a *neutral network*. This is a particular kind of set of genotypes, i.e., a particular kind of structure in genotype space. A neutral network is a set of genotypes, which all map to the same phenotype, and which all form a connected set with each other under mutation.

Why is this concept important? Since the genotypes of a neutral network all share the same phenotype, fitness provides neither obstacles nor inducements for evolutionary change within the neutral network. Since the genotypes all form a connected set, the neutral network could act as a kind of a highway providing paths of connection between otherwise distant regions in genotype space. Researchers have found that the neutral networks do have such a shape, in the case of RNA. The so-called "shape space covering" describes the fact that neutral networks are so widely dispersed throughout the genotype space, that a randomly chosen sequence is never very far from a neutral network which maps to any of the more common shapes (Fontana, 2002, p 1171).

Neutral networks could in principle take on a wide variety of structures. Every structure could have subtly different effects on the dynamics of evolutionary change under that genotype-phenotype mapping. These structures are virtually impossible to visualise since

they occupy the mathematical space of all possible strings, which has as many dimensions as the length of string under consideration. This leads to a highly theoretical kind of research, where researchers explore various ways to characterise these structures and suggest the different ways they might influence evolution.

In a recent detailed investigation, Rendel (2008) explores the neutral networks in the genotype-phenotype map resulting from exhaustive folding of all RNA strings of length 16. He studies different measures of network connectivity, and considers evolution on these networks under drift versus directional selection, and under different regimes of assumed correlation between phenotype and fitness. The main finding is that "degeneracy does not increase the accessibility of adaptive mutants" (p 171).

Implications for formalism

RNA sequences and advanced organisms are different enough that research into the RNA map is no substitute for thinking directly about the problems of organismic evolvability.

However, the RNA work is still extremely useful. Not only are new and important roles of miRNAs currently being discovered frequently, but the methodologies developed for RNA will be ready to be extended to the more complex case. For this reason, in chapter 3 I will review the RNA-based work focused mainly on these methodologies. This introduces a few ideas – the idea of an accessibility relation, of the graph defined by phenotype accessibility relations, and the idea of modelling those relations using topological spaces.

2.4.4 Research not based on RNA

RNA work is not the only empirical research into mutational relationships defined by genotype-phenotype mappings. Other research has investigated the evolvability of microbes. The fact that this work is studying real organisms gives it more direct implications for basic questions of evolvability, but also makes it impossible to have the kind of global knowledge of the genotype-phenotype mapping over all possible genotypes.

Dichtel2004: empirically measuring micro-evolvability

Dichtel et al. (2004) points out that an individual's possibilities offspring phenotypes are determined by its genotypic mutational neighbourhood, and that the genotype-phenotype

mapping structures this relationship. It defines as "the phenotypic neighborhood" the set of mutationally accessible phenotypes. These are familiar and basic points in theoretical discussions of evolvability. In its focus on the individual's neighborhood, it does not consider the importance of the global genotypic structure or provide a way for identifying the role of the developmental system. Perhaps in recognition of this, it is titled as a discussion of "micro-evolvability".

However, the main contribution of this paper is in methodology. It reviews methods for exploring phenotypic evolvability in the laboratory – genetic screens, mutations accumulation lines, and selective breeding. It considers two examples – C. elegans body size (a single quantitative phenotypic trait), and nematode vulva development (which features qualitatively varying phenotypes, and allows considerations of development).

Landry2007: evolvability of gene expression

Landry et al. (2007) is an example of exciting, current, empirical research on regulatory evolution and evolvability. Like Dichtel et al. (2004), this paper uses mutation accumulation lines to study the phenotypic neighbourhood of an individual genotype under mutation. But instead of studying *C. elegans* body size and structure, the model organisms is yeast and the phenotype is represented by gene expression levels.

Given the past work on yeast, this paper is also able to examine how the phenotypic neighbourhood is affected by the number of genetic and regulatory elements that a given gene depends on. The research shows, as expected, that a larger mutational target size (due to dependencies on other elements like TATA boxes or trans-regulatory elements) is associated with greater phenotypic variation following mutation.

It is instructive to note how this paper illustrates the current status of evolvability as a research topic. This experiment measures phenomena central to evolvability theory, phenomena that have become accessible to empirical research only relatively recently: the phenotypic neighbourhood (or "effects of spontaneous mutation with the confounding effect of natural selection"), and the question of how epistatic dependencies in the genotype-phenotype mapping affect phenotypic variation (through the study of the global mutational target size). The paper also uses the word evolvability occasionally, recognising its relevance.

But unlike Dichtel et al. (2004), the paper never defines the term, and it never frames the

empirical results in terms of basic ideas like phenotypic neighbourhood, or variation versus variability. This is not atypical. Although the subject of evolvability attracts widespread interest, and the term may sound familiar, the basic theoretical foundations are unknown and frequently re-invented.

2.5 Appendix: relevance of evolutionary computation

Conventional thinking is that evolution requires only variation, selection, and inheritance. But as discussed in section 2.2.3, work in evolutionary computation has shown that biological evolution needs more than this. This is because computational simulations with those three ingredients have reproduced the process of evolution but have failed to reproduce the richness and complexity that is distinctive of the biological world.

This is a startling result if one really expects that these three ingredients are all that evolution needs. This expectation may sound simplistic, but that is the point: when you state the conventional thinking plainly, it does sound simplistic. It sounds simplistic because it is.

Nevertheless, it is reasonable to be conservative about revising such fundamental ideas. One way to hold onto the conventional account is to object that computer simulations are not a valid test of the conventional account. This appendix will answer a few forms of this objection.

2.5.1 Objection based on misunderstanding scope

One objection is that computer simulations are not a valid test because they do not adequately reproduce necessary conditions of biological evolution that, one might argue, are so obvious they are taken for granted. This objection can have two senses: it can refer to physical conditions or to what we might call process conditions. Let us take these senses one at a time.

First, it could mean that the simulations do not adequately reproduce particular physical conditions of biological evolution, such as the existence of carbon, of DNA-based genes, etc. For instance, if the systems do not simulate DNA, then they cannot show that DNA-based evolution fails to become rich and complex. This is certainly true. But the problem is that this objection misunderstands the intended scope of the original argument about what is

"necessary for evolution". These physical factors do not fit under under the same heading as the three logical ingredients – they are not the same *kind* of thing.

The goal of alife systems is to test a more general claim about evolution as a logical process, the claim that it only needs variation, inheritance, and selection. This claim pretends to be indifferent to such physical details. Objecting that alife work does not reproduce physical conditions fundamentally misunderstands the nature of this argument.

Of course if rich evolution in fact requires not just inheritance, but inheritance mediated by DNA molecules, then that would be an important finding. It would also be a substantial revision of our current understanding of evolution, and the next step clearly would be to try to understand why this is the case. But the goal of alife systems is not to test such narrower claims about the importance of familiar physical conditions. For one thing, these claims are rarely made. Also, to test them, an alife system would need to be like a model train set, a virtual replica of every physical detail of biological life. It would be so complex that it would be impossible to implement, and uninformative even if it were possible to implement.

2.5.2 Objection based on timescale

Second, it is also possible to object that the simulations do not reproduce necessary process conditions for evolution. Let us consider two possible process conditions: timescale, and the maximum supported complexity.

As regards timescale, the objection is that variation, inheritance, and selection are enough to produce rich evolution, but that the process of evolution needs enough time to operate. It is clear that modern life evolved over billions of years (Dartnell, 2007). Alife simulations, by contrast, never run more than a year. Maybe the three ingredients are enough, but the alife simulations do not let them "cook" long enough to produce rich evolution. This objection sounds familiar since it echoes a typical supposition in evolutionary biology – the argument that some process, modest in incremental steps, will produce dramatic results over longer timescales (Darwin, 1859, p 189).

But this objection fails in the case of simulation systems. What matters is not the actual time required to run the simulation (so called "wall time"), but the time elapsed within the world of the simulation itself as measured, for instance, by the number of generations. As Rasmussen (1992) puts it, "An artificial organism must perceive a reality R_2, which for it

is just as real as our 'real' reality R_1 is for us" (quoted in Rennard (2004)). It is the reality within the simulations that tells the story.

And within the alife simulations, organisms have had ample time to evolve. It takes only hours to run thousands of generations (Ofria et al., 1988). The simulations fail to generate more complexity not because experimenters shut them off while they are still flourishing, but because they stabilise and stagnate. This is why the alife literature, which is discussed and referenced in section 2.2.3, is not about how to speed up the simulations to give them more time to run; it is about improving how they run. Perhaps the strongest evidence comes from what is not published. I do not refer to any articles on how to speed up an alife simulation that is manifesting ever increasing complexity but running out of computational resources. I do not refer to such an article because, in reviewing the alife literature, I have not found a single one. This includes all articles in the journal *Artificial Life*, since it started publication in 1989.

It is of course possible that the existing alife systems, which seem to be stagnating, would in yet more time suddenly manifest unexpected evolutionary innovation. It would take a mathematical proof to exclude this possibility absolutely (an exercise which might be an informative). But even lacking such a proof there seems to be no reason to believe it.

All of this rebuts the objection that simulations do not reproduce the necessary process condition of adequate time.

2.5.3 Objection based on simulations being "too simple"

As a kind of agreement rather than an objection

Another kind of "process condition" objection might be that variation, inheritance, and selection are enough to produce rich evolution, but that the simulations cannot produce rich evolution because they are just "too simple".

This objection has great intuitive appeal. But on closer inspection it looks less like an objection to the original claim and more like agreement. This is because it is in fact an argument from substrate evolvability: the objection is essentially that substrates need to be sufficiently complex themselves in order to support the evolution of complex forms – i.e., that the simulations are not evolvable substrates because they themselves are too simple.

The original claim was merely that the failure of the simulations shows we need an idea

of evolvability in addition to the ideas of variation, selection, and inheritance. Obviously, you cannot rebut this claim by saying the simulation failures do not count since those simulations lacked evolvability. You cannot use the idea of evolvability as a premise to argue that there is no need for an idea of evolvability.

As an argument about the maximum allowed complexity

However, let us try to reformulate a stronger version of this objection by drilling into what might be meant by the idea that the alife simulations are "too simple".

The objection might be that all the alife simulations are, like the biomorph model, too simple to support lifelike complexity *under any circumstances*. That is, they are too primitive for any structure of lifelike complexity to be built on them, even if it were constructed "by hand"; so, *a fortiori*, they are also too primitive for evolution to build a complex structure. The simulations do not test if evolution can produce lifelike complexity, because the simulations have an absolute complexity ceiling below the level of lifelike complexity.

(Why is this itself not also a statement about the evolvability of these simulations? Maybe the need for a high enough complexity ceiling is taken as so obvious that there's no need for a neologism like "substrate evolvability" for discussing it. And maybe it is further assumed that beyond this ceiling evolution proceeds so effortlessly that neither is there any need to discuss how *easily* evolution proceeds.)

This view shifts all the weight onto the idea of how *much* complexity a substrate can support. Well, do the simulations support sufficient complexity? The problem is, as described in section 1.2.5, there is no consensus definition of complexity, so there is no obvious way to measure how much complexity a substrate can support. Without more rigour, any argument downplaying a simulation's maximum possible complexity is at risk of becoming suspiciously vague, and of just using "complexity" as a magic word to dismiss unwelcome simulation results.

Relevance of Turing-completeness

However, there is in fact one rigorous idea that is relevant to this objection, and it seems to undermine it fatally. This is the idea of computational universality, or Turing-completeness. The concept of Turing-completeness is rooted in the foundations of logic and computer science. A language is said to be Turing-complete if it is sufficiently rich to simulate a

Turing machine. A Turing machine is itself an exact, extremely simple idealisation of a computing device. If you write a program for a Turing machine, it will execute it and produce some result. Alan Turing defined the formal idea of a Turing machine while trying to clarify the informal idea of an effective procedure, which is, very roughly, a well-defined procedure, based on a finite number of instructions, which could be carried out by any diligent person without special insight, and which will produce a desired result in a finite number of steps. (Turing (1936) is the original publication. Countless works now summarise this result, such as, e.g., Herken (1988).)

What is important about the idea of an effective procedure is that it is a very general notion for almost any kind of explicit, formalisable, step-by-step logical process. And what is relevant to alife about the idea of a Turing machine is the surprising finding that, despite the extreme simplicity of a Turing machine, it seems "that every effective procedure can be carried out by a Turing machine". This is the *Church-Turing thesis*. It has not been absolutely proved, but it is supported extremely strongly (Copeland, 2002). What this implies is that almost the simplest imaginable computing machine, a Turing machine, can in principle calculate anything which can be calculated by much more elaborate computing machines. It may calculate it more awkwardly or more slowly, but it can calculate it. The only fundamental computational limit on any Turing machine is storage and time, rather than some other feature of the language or the computing device.

Turing-completeness of alife systems

Why does this matter for alife? In many alife systems every organism is a little program, and the set of possible organisms is defined by a little programming language. This language is effectively the substrate of evolution. Most of those languages are Turing-complete. For instance, Ofria et al. (2002) describes avida, Tierra, and coreworld as Turing-complete. Pargellis (2001) compares a couple other systems as well, such as Amoeba.[7] Therefore, in principle, those languages can produce anything which could be produced by a different or more complex programming language. If evolution fails to produce some kind of complex result in a relatively simple alife simulation like avida (and if this is not because of limits of storage and time), then we cannot say this is because such a result requires a more complex

[7] It may be that all the main alife systems are Turing complete. I am unaware of any complete surveys of the matter. Wolfram (2002) describes extensive new work based on two-dimensional cellular automata, showing that almost the simplest imaginable set of transformation rules can be Turing complete.

language to be expressed at all. In the strictest and clearest sense available, these simple languages are complex enough.

In other words, a simple language can in principle do anything a complex language can do. So the problem is not that a simple alife language cannot support complex life, but that evolution does not in fact produce complex life with a simple language. Evolution might produce complex life with more complex alife languages. We do not know. All we know is that evolution produces complex life in the physical world.

In general, we have assumed simulations are valuable because they can play out the dynamics implied by our theories of the physical world. This is an informal idea. A more formal argument comes from the claim that the physical world is itself perfectly simulatable. Since simple languages can in principle simulate anything that is simulatable, this claim would make even stronger the argument that even the simple alife systems should, in principle, be able to produce complexity similar to what is observed in the physical world. The claim that the physical world is itself perfectly simulatable comes down to the question of whether the laws of physics can be described as the carrying out of an effective procedure. This is the assumption of the so-called *physical Church-Turing thesis*, which is widely credited, and asserts that a Turing machine can simulate any "finitely realizable physical system" (Deutsch, 1985).[8]

In short, alife simulations are computationally universal, i.e., Turing-complete. Turing-completeness defines a maximum of the possible supported complexity in a computational process. Anything which is Turing-complete can simulate anything else which is. So it is not enough to claim that the alife simulations provide too simple a substrate. They are simple, but they are complex enough to be in principle equivalent to any other computing system (Church-Turing thesis). As long as we believe that the physical world can in principle

[8]Copeland (2002) summarises the Church-Turing thesis. He takes pains to distinguish Turing's original thesis, which concerned Turing machines and effective procedures, from a different claim that now erroneously goes by the same name, the so-called physical Church-Turing thesis that a Turing machine can simulate any actual physical machine. He is particularly concerned about how the physical Church-Turing thesis underwrites a naive computationalism in issues of philosophy of mind, by implying that the brain and mental processes can be simulated by a Turing machine. Why does this matter? Similar concerns apply to the computationalism that underwrites the significance claims of artificial life, which assumes that real-world biological process could in principle be simulated by Turing machine, and to the argument I present here that relatively basic alife systems like avida could still in principle show unlimited complexity.

Rennard (2004) also provides a notable discussion of the key role of computationalism, summarising the work of Rasmussen (1992), who states this computationalism explicitly.

The physical Church-Turing thesis would be false if Turing machines could not simulate all physical process, that is, if there were physical processes that displayed super-Turing computability, making it possible to build what is known as a hypercomputer. Douglas (2003) reviews the literature on super-Turing computation. At present, the idea remains a mathematical idea which seems impossible given the actual physics of the world.

be simulated by a computing system, and most observers do (the physical Church-Turing thesis), then the alife simulations are complex enough in principle to support the same level of complexity as the biological world.

Since simulations are substrates which support the maximum potential complexity, what we are seeing is that evolution is failing to realise the potential complexity allowed by the substrate. That is, the simulations do still illustrate the question of substrate evolvability. They show that evolution needs a substrate not merely that makes complexity possible, but that makes complexity possible *for evolution*. Rennard (2004) provides an excellent summary of related difficulties of artificial life work, discussing both the ontological claim that alife *is* a kind of life, and the failure of past work to reproduce the rich phenomena of biological life.

I emphasise again that the equivalence of different alife systems, as regards their maximum supported complexity, is only true subject to computational restraints of storage and time. My survey of the literature suggests that these restraints have not in fact stymied evolution in these systems, so that this proviso does not undermine the conclusion. However, it would take a more formal argument to establish this absolutely.

2.6 Appendix: proof of euclidean distortion

Here we prove that is impossible generally to represent the mutational structure of a genotype space by embedding it in the Euclidean plane. This is relevant because it means that the shapes in fitness landscapes must misrepresent the actual possibilities of evolutionary change, as discussed in section 2.3.4. We will first establish a few lemmas, and then note their effect on a counterexample.

First, we observe that: to draw some graphs in the plane with more than four nodes, we must cross edges. Or more technically, not every graph is a *planar graph*. Kuratowski's Theorem proves this in detail (Kuratowski, 1930), but we can also see the fact directly:

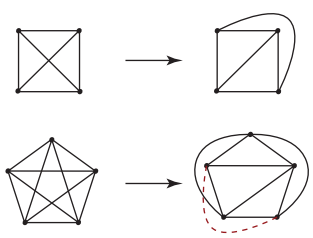


Figure 2.5: The complete graph K_4 (above) can be redrawn without edge crossings. But K_5 (below) cannot. Only K_4 is a *planar graph*.

Second, we observe that: every tiling in the plane implies a corresponding graph with no edge crossings. To construct this graph, we define a vertex for every tile in the space and an edge for every boundary between two adjacent tiles.

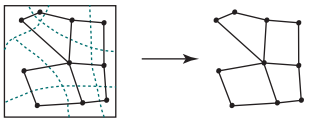


Figure 2.6: Every tiling of an area implies a corresponding planar graph.

Third, since every tiling implies a corresponding a graph without edge crossings, it follows by the contrapositive that a graph with edge crossings is not implied by any tiling.

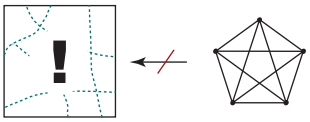


Figure 2.7: Thus, a nonplanar graph is not implied by any tiling.

Now consider how these points apply to the biology of genotype-phenotype mappings. Imagine a mapping to five possible phenotypes $p_1, \ldots, p_5$, where each one can always mutate into the others. This is diagrammed as a complete graph of five nodes, where every edge represents the mutational link between a pair of phenotypes. The question is, if we wish to represent the genospace as an area in the plane, how do we cut up that area into five regions so that each region neighbours all the other regions? This is the area that would underlie a traditional fitness landscape plot of this system, even though that plot ignores the realities of phenotypes and maps directly to fitness.

The answer is there is no way to do this. Since the graph of the mutational links between the phenotypes will always have edge crossings, there is no tiling which can imply that graph.

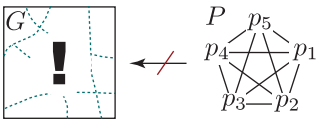


Figure 2.8: No partitioning of the plane produces five genotypic regions corresponding to five mutationally linked phenotypes.

This completes the proof.

Chapter 3

The topological phenotype

Chapters 1 and 2 reviewed past thinking on evolvability. One conclusion of that review was that a theory of evolvability will need to be a theory of phenotypic change – specifically, a theory of how certain kinds of developmental systems make certain kinds of phenotypic change easier. This will require new modelling approaches which focus less on fitness and more on the genotype-phenotype map and on the phenotype itself.

This chapter looks at one candidate approach, a body of work that has aimed to describe evolution using the mathematics of elementary topology (Fontana and Schuster, 1998a,b; Schuster and Fontana, 1999; Bornberg-Bauer and Chan, 1999; Cupal et al., 2000; Stadler et al., 2001; Stadler and Stadler, 2002; Wagner and Stadler, 2003). This work argues that the conventional approach of quantitative genetics judges phenotypes too much by their appearances. It treats phenotypes as fully characterised by physical attributes, so that they differ only insomuch as those attributes differ. This is a result of framing the theory in a *metric space*. Instead, the work argues for using a more general construct, a *topological space*, which will let us express the mutational relationship between phenotypes as they are defined by the genotype-phenotype map. A model built in this way will see under the surface of the phenotype and describe the invisible genotypic structures that guide visible phenotypic change.

Another thread in the argument is that using topological spaces will let us apply standard topological ideas to biological problems. For instance, one idea is that topological continuity can formalise the notion of continuity in evolutionary change (Fontana and Schuster, 1998a). Another is that pretopological local factorisability can formalise the notion of phenotypic

character (Wagner and Stadler, 2003).

This chapter evaluates just the conceptual core of this proposal – the idea of representing phenotypes with topological spaces. The main finding is that it does not work. When applied in a consistent manner, this mathematical construct destroys too much information to be useful for models (as is summarised in figure 3.5, and worked out in section 3.7.5).

However, this work also developed some simpler ideas which seem very promising, such as the *phenotypic accessibility relation*. These ideas allow one to frame new ideas like the idea of the *dead end genotype*, a possible contra-indication of evolvability.

3.1 Critical review

Past work has aimed to apply topological ideas to illuminate a variety of biological topics: computational studies of the genotype-phenotype map defined by RNA folding; the discussion of evolutionary continuity; the definition of phenotypic character; and the role of neutral networks. I do not address these applications of the topological approach, but will reconsider the conceptual foundations of the approach itself.

This past work may be divided into three phases. This section briefly summarises this work and criticisms of it. Technical discussions appear in the following sections.

3.1.1 Informal phase

In the first phase (the informal phase), Fontana and Schuster (1998a) introduce the term "topology" to describe the relations between phenotypes. This is in the context of studying continuity in the evolution of RNA sequences, but the topological ideas are not restricted to this system and set the mold for later conceptual work.

To treat phenotypes as making up a topological space they rely on the notion of a *neighbourhood system*. This is one of the five standard, logically equivalent structures for defining a topological space (Sugakkai, 1993, p 1606). They define a phenotype's neighbourhood in terms of an *accessibility relation* between phenotypes, a relation which is in turn defined in terms of an *accessibility measure*. The accessibility measure gives a real number for every ordered pair of phenotypes. If the measure from α to β exceeds a threshold, then the accessibility relation holds and we say α can access β, indicating that it is relatively easy for α to mutate into β. This measure is itself based on the *genotype-phenotype map*. In other words,

this work implies a chain of formal constructs: genotype-phenotype map, to accessibility measure, to accessibility relation, to neighbourhood, to topological space. This is effectively a recipe for turning one construct, the genotype-phenotype map, into another, the topological space. The point of this is to make use of the rich set of pre-existing ideas for analysing topological spaces, such as ideas of continuity.

This topology-construction recipe or variations of it is used in later work. Fontana and Schuster (1998b) discuss the same issues at greater length, sampling the RNA genotype-phenotype map to generate a "statistical topology". Bornberg-Bauer and Chan (1999) and Schuster and Fontana (1999) continue in this vein, also using the term "topology" in studies of neutral networks and RNA folding.

What is a topological space?

But there is a problem. While Fontana and Schuster (1998a) originally describe treating genotypes and phenotypes as a "topological space", this is not strictly accurate. To define a topology, a neighbourhood system must meet certain definitional criteria. They never show that a phenotype's neighbourhood, as defined by an accessibility relation, is guaranteed to meet these criteria. The papers above do not resolve this either. In other words, the implied recipe for building a topological space is flawed, since there is nothing about it which guarantees that it will produce a true topological space. This is shown in section 3.3.1.

What is the reason for this oversight? It might be due to some unfortunate ambiguities of usage. "Topology" can refer to a large branch of mathematics, dedicated to the study of invariant properties of deformable objects. More narrowly, "topology" is used as a shorthand for the foundational ideas within this branch, the speciality of elementary topology (also, *general topology* or *point-set topology*), which is the study of particular objects called topological spaces. Even more narrowly, a topology is also the key construct in the most common of several equivalent definitions of a topological space. In this sense, a topology is the family of special subsets of a space, known as its "open sets" (details in section 3.7.2).

Separately, "topology" is also used in the mathematical discipline of graph theory or network studies. A graph is simply a set of points (or nodes, or vertexes) together with a set of lines connecting them (or edges). In this context "topology" is just the name for the particular structure of a graph, the particular pattern defined by all its edges. This is in fact the kind of structure defined by an accessibility relation, where the relation defines the

edges and the phenotypes the nodes, but this is a totally different thing from a topological space.

The point is, unfortunately, "topology" does mean a lot of things. But every one of those things is an exact concept. Some of those exact concepts, like the topological space, are the foundation of other exact concepts, such as continuity, homeomorphism, etc. While Fontana and Schuster (1998a) may aim to "properly frame continuity in the spirit of topology", the spirit of topology is properly inextricable from these details. Ultimately, the problem with this phase of work is that it claims to be using these exactly defined topological concepts (phenotypes define *topological space*, based on *neighbourhood systems*, allowing analysis in terms of *continuity*) while it is actually using the terms only informally, possibly mixing in an unrelated sense from graph theory (so neighbourhood just means mutational neighbourhood, topology just means structure, and topological space just means a set plus some structure).

3.1.2 Formal phase

The second phase of work on topologies (the formal phase) is represented by Cupal et al. (2000), who are the first to engage the mathematical definitions. They summarise the two different accessibility relations proposed by Fontana and Schuster (1998a) and offer a new one. They summarise relevant results from the mathematical literature, such as simplifications that apply to all finite topological spaces. They also spell out a procedure for building a valid topological space from the graph defined by an accessibility relation. This is also the first contribution to mention pretopologies (also known as generalised topological spaces), in the course of a passing reference to the mathematician Hausdorff.

This work resolves the errors implicit in past uses of the term "topological space". However, the clarified approach exposes a more fundamental flaw: the revised topology-construction recipe destroys so much information that the resulting topological space is unlikely to be useful for modelling. This is shown in sections 3.3.2 and 3.3.3.

3.1.3 Pretopology phase

Stadler et al. (2001) inaugurates the third phase of work (the pretopology phase), reviewing past work, and summarising results in the mathematics of pretopologies, which they assert

are preferable to topologies. Stadler and Stadler (2002) continue in this vein, considering how that formalism handles genetic recombination. Wagner and Stadler (2003) propose that the biological concept of character, originally proposed by Lewontin (1978), can be equated with local factorisability of the phenotype space treated as a pretopological space. I comment on this in section 3.5.

3.2 Four key constructs

The procedure for using topological spaces to analyse genotype-phenotype maps depends on four quantitative constructs: the mutational boundary of a set of a genotypes, the pre-image of a phenotype, the accessibility measure between two phenotypes, and the accessibility relation between two phenotypes. I will now describe these constructs to lay the groundwork for my technical criticism of this procedure. These constructs are also worth understanding in their own right. They provide an exact way of thinking about aspects of evolution, and I will use them in my own analysis in chapters 4 through 6

An *accessibility measure* aims to quantify how "easy" it is for one phenotype to evolve into another. It targets a particular kind of "easiness" – easiness as determined purely by the genotype-phenotype map, prior to choosing any regime of selection which might favour one phenotype or another. This is not in order to deny the obvious importance of selection, but to speak more precisely about the constraints which exist prior to selection and within which selection operates.

The accessibility measure is justified by a probabilistic argument. A transition from one phenotype to another is easy if it is probable. As usual, we calculate probabilities by counting alternatives. Since changes in the phenotype are due to underlying changes in the genotype, the measure is essentially an attempt to count the number of possible underlying genetic changes that will produce a given phenotypic change. This tally will differ for different genotype-phenotype maps, indicating that a given phenotypic change – a given evolutionary change – will be easier under some genotype-phenotype maps than others.

How do we make this count? A few accessibility measures have been proposed. All these measures use variations of the same idea: start with two phenotypes, and look at how their *pre-images* abut one another in genotype space. A pre-image is a standard mathematical concept. In general, the pre-image of a point α under a mapping $\phi : G \to P$ is the set of all

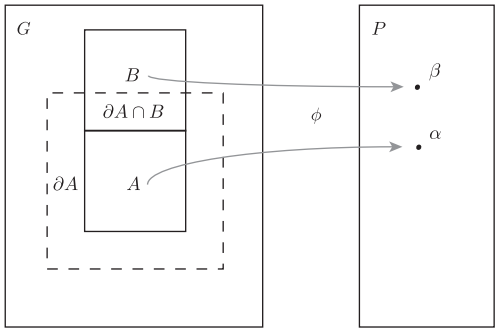


Figure 3.1: Calculating the accessibility measure κ. The genotype-phenotype map $\phi : G \to P$ associates regions in genotype space (such as A and B) with particular phenotypes (α and β). The accessibility measure, $\kappa(\alpha \to \beta) = |\partial A \cap B|/|\partial A|$, looks at the relative size of the overlap of B and ∂A, the annular region which is the mutational boundary of A.

points A that map to α. That is, $A = \phi^{-1}(\alpha) = \{a \in G \,|\, \phi(a) = \alpha\}$. A phenotype's pre-image is merely the set of genotypes which will map to that phenotype under the genotype-phenotype mapping. Looking at how the pre-images of two phenotypes abut one another is a way of estimating the chance of a mutation that will take a genotype from one phenotype to another. The measures are all defined based on the size of the overlap between these two regions, between their mutational boundaries, or between some combination of the two.

For instance, figure 3.1 illustrates the quantities used to calculate the accessibility measure proposed by Cupal et al. (2000, p 4). The phenotype α has pre-image $A = \phi^{-1}(\alpha)$ and similarly $B = \phi^{-1}(\beta)$. The accessibility measure, κ, is as follows:

$$\kappa(\alpha \to \beta) \stackrel{\text{def}}{=} \frac{|\partial A \cap B|}{|\partial A|} \quad \text{where } A = \phi^{-1}(\alpha), \; B = \phi^{-1}(\beta) \tag{3.1}$$

Here ∂A is a shorthand for $\partial_C A$, the *mutational boundary* of A. This is the set of all points which are not themselves in A but which may result from mutations of points in A. That is,

$$\partial_C A \stackrel{\text{def}}{=} \{a \,|\, a \notin A \text{ and } \exists a' \in A \text{ where } a \text{ is mutant of } a'\} \tag{3.2}$$

Under this definition, κ is the probability, given that a mutation is not neutral, that that mutation will take α to β.

Once we have an accessibility measure, an *accessibility relation* is defined by choosing

a threshold value κ_0 of the accessibility measure. One phenotype is accessible to another if the accessibility measure from one to another is above the threshold. In addition, the measure is reflexive by stipulation. So:

$$\alpha \hookrightarrow \beta \stackrel{\text{def}}{=} \kappa(\alpha \to \beta) > \kappa_0 \text{ or } \alpha = \beta \tag{3.3}$$

Now if there's a high probability that α can evolve into β, we find that $\alpha \hookrightarrow \beta$ and we say that α can "access" β. Otherwise, $\alpha \not\hookrightarrow \beta$.

3.3 Technical criticism

Using the constructs defined above, this section explains the technical problems with the literature arguing for topological spaces. It is not essential reading but is relevant for understanding the subtleties of topological mathematics, which turn out to be treacherous. Readers wishing to skip these technicals detail may jump directly to section 3.4. The main points of this section may be summarised as follows:

1. Under the correct definition of a topological space, it is not possible to define a topological space from a genotype-phenotype map in the way suggested by the topology literature in its informal phase (Fontana and Schuster, 1998a,b; Schuster and Fontana, 1999; Bornberg-Bauer and Chan, 1999).

2. However, it is possible to build a topological space using the procedure described in Cupal et al. (2000). But this procedure destroys so much representational information that it is not suitable for modelling.

3. Merely building an accessibility measure, the first step on the way to building a topology, also destroys some representational information. For instance it becomes impossible to detect *dead end genotypes*, genotypes that because of their position in genotype space lack the mutational opportunities that their phenotype would suggest.

3.3.1 An accessibility relation alone is not equivalent to a topological space

Here I review definitions in order to show that an accessibility relation alone does not define a topological space. They are not equivalent constructs. This is relevant because Fontana and Schuster (1998b, p 514) and Fontana and Schuster (1998a, p 1451) speak of topologies and "topological spaces", but they only discuss how to define accessibility relations and neighbourhood functions.

Neighbourhood functions versus neighbourhood systems

As discussed, an accessibility relation is a reflexive binary relation between phenotypes which summarises how easily one can evolve into another. From an accessibility relation we can define a *neighbourhood function*, $N : X \to \mathcal{P}(X)$ where $N(x) = \{y \in X | x \hookrightarrow y\}$. This function gives the set of points accessible from x, known as x's *neighbourhood*. An accessibility relation and a neighbourhood function are equivalent in the strict sense that they are different formulations of the same idea and are freely inter-convertible without loss of information. For instance, the stipulation that every point is accessible to itself (so that $x \hookrightarrow x$ for all x) translates into the requirement that every point be in its own neighbourhood (so that $x \in N(x)$ for all x). So much for neighbourhood functions.

One way to define a topology is in terms of a *neighbourhood system*, a collection of neighbourhoods. A neighbourhood system is defined by certain conditions, as described in Sugakkai (1993, p 1606):

> Let X be a set. A *neighborhood system* for X is a function $\mathcal{U}$ that assigns to each point $x \in X$, a family $\mathcal{U}(x)$ of subsets of X subject to the following axioms:
>
> 1. $x \in U$ for each U in $\mathcal{U}(x)$
> 2. if $U_1, U_2 \in \mathcal{U}(x)$, then $U_1 \cap U_2 \in \mathcal{U}(x)$
> 3. if $U \in \mathcal{U}(x)$ and $U \subset V$, then $V \in \mathcal{U}(x)$
> 4. For each U in $\mathcal{U}(x)$, there is a member W of $\mathcal{U}(x)$ such that $U \in \mathcal{U}(y)$ for each y in W.

The point to note here is that these conditions apply to a *family* $\mathcal{U}(x)$ of subsets – a family of neighbourhoods – which is defined separately for *every point* in X. This is different from $N(x)$, which merely defines a single subset – a single neighbourhood – for every point in X. Or to put it formally, $\mathcal{U}(x) \in \mathcal{P}(\mathcal{P}(X))$ while $N(x) \in \mathcal{P}(X)$. These are very different structures.

Implied equivalences do not work

Fontana and Schuster (1998a,b) describe how to build an accessibility relation, and how an accessibility relation defines a single neighbourhood function. They also talk about defining a "neighborhood system" (Fontana and Schuster, 1998b, p 503) or a "system of neighborhoods" (Fontana and Schuster, 1998a, p 1451). But they never describe the procedure to generate a valid neighbourhood system from a single neighbourhood function. For this reason Fontana and Schuster (1998a) and subsequent work is not in fact applying the concept of a topological space.

Their omission suggests the procedure is not worth mentioning. Perhaps there is a single obvious way to do it? One procedure might be as follows. Since a neighbourhood system is a family of sets, and since a neighbourhood function gives one set at a time, then maybe the neighbourhood system can be defined simply as the range of the neighbourhood function, $\{N(x) \mid x \in X\}$, i.e., the family of neighbourhoods that results from applying the neighbourhood function N to every point in X. Unfortunately this procedure does not meet the basic definitional requirement that the neighbourhood system give a family of neighbourhoods for *every point*. It is hard to be certain, but this procedure seems to be what is intended by Fontana and Schuster (1998a, p 1451):

> A set of entities is organized into a (topological) space by assigning to each entity a system of neighborhoods. In the present case, there are two kinds of entities: sequences and shapes, which are related by a thermodynamic folding procedure. The set of possible sequences (of fixed length) is naturally organized into a space because point mutations induce a canonical neighborhood. The neighborhood of a sequence consists of all its one-error mutants.

But defining the neighbourhood of a sequence as all its one-error mutants is not the same as assigning a family of neighbourhoods to each point.

Another possibility is that they intend $\mathcal{U}(x) = \{N(x)\}$ – i.e., that the neighbourhood system of a point is the singleton family consisting of only one neighbourhood, the result of the one neighbourhood function evaluated at that point. This satisfies axiom 1 trivially. Furthermore, if $|\mathcal{U}(x)| = 1$, then $U_1, U_2 \in \mathcal{U}(x)$ implies $U_1 = U_2 = N(x)$, so $U_1 \cap U_2 = N(x) \cap N(x) = N(x) \in \mathcal{U}(x)$, satisfying axiom 2. However, a counterexample shows axiom 3 does not hold in general.

Suppose we have a space $X = \{1, 2, 3\}$, with a neighbourhood function as follows:

$$M(1) = \{1, 2\}$$

$$M(2) = \{2\}$$

$$M(3) = \{1, 2, 3\}$$

We see that $M(1) \in \mathcal{U}(1)$, and $M(1) \subset M(3)$, but $M(3) \notin \mathcal{U}(1)$, violating requirement 3. This example proves it is also inadequate to suppose that a neighbourhood system evaluated at a point can be simply the value of one neighbourhood function evaluated at that point.

So it seems there is no obvious way to produce a valid neighbourhood system from a neighbourhood function.

Missing data are not the problem

The problem is logical and definitional, so neither does it solve the problem to refer to the gap between the true genotype-phenotype and what we know experimentally. For instance, Fontana and Schuster (1998b) analyse what they refer to as a "statistical topology". All accessibility relations are based on the genotype-phenotype map. The statistical topology results from an accessibility relation calculated by randomly sampling the genotype-phenotype map, rather than documenting it exhaustively. They concede that "a rigorous topology is invariably spoiled by the complexities of folding". But it is unclear if they mean that it is sampling error which makes the relation unsuitable for building a rigorous topology, or the structure of the RNA folding map itself.

If they mean sampling error, they should explain what topology-spoiling property is possessed by sampling error but not by the RNA map itself. If they mean it is the RNA map itself which has this property, then why originally describe it as a topology? (Or at least, they should specify what a non-rigorous topology is.) Without a statement on when and how an accessibility relation can create a rigorous topology, it is impossible to disqualify any particular relation, whether it's formed by the RNA folding map, polluted by sampling error, or chosen entirely at random by rolling a pair of dice. And no statement is forthcoming because an accessibility relation alone does not define a topology when you consider the definition of a topology in detail.

3.3.2 Constructing a finite topology from an accessibility relation destroys information

Cupal et al. (2000, p 6) do offer a procedure for creating a topology from an accessibility relation. Instead of trying to build a neighbourhood system, this procedure works by using the accessibility relation to construct a family of subsets called a *topological basis* (definition review in section 3.7.2). Since a basis set defines a single topological space, this procedure suffices to construct a topological space from an accessibility relation.

The basis they construct is also special – it is the *non-redundant basis* of the topology, and it is unique to that topology (proof in section 3.7.3). Although every basis defines a single topology, a topology may be defined by many different bases. However, every finite topology is associated with only a single non-redundant basis. Since we assume biological spaces are finite, we can use that non-redundant basis as a standard representation of any topology we encounter.[1] It becomes a convenient shorthand for designating, comparing, and depicting these topologies.

Cupal et al. (2000, p 6) highlight the uniqueness of the basis when they introduce it as part of their procedure for constructing a topology:

> Given an accessibility relation $\mathcal{N}$ (that is, a subbasis[2] of a topology τ on V), we

[1] Is it reasonable to assume all biological systems can be represented by finite mathematical spaces? There are two questions here. The first is whether biological systems are *actually* finite in their possible variation. The second is whether it is a good approximation to *represent* them as such with finite mathematical structures. If biological systems are finite, then they can certainly be represented as such. If they are infinite, then a finite model might still suffice if their biological properties did not depend on their infinitude.

There are two mathematical spaces in the models we discuss, the set of genotypes G and the set of phenotypes P. All known genetics is based on finite genome strings built from finite alphabets, which produces only a finite number of combinations, so it seems reasonable to suppose the set of possible genotypes G is finite. If we are neglecting random and environmental factors on development, as these models do, then the corresponding phenotype space must be finite as well so there is no problem.

However, even acknowledging such factors, it still seems reasonable to assume finitude. It is true that a traditional quantitative model will represent a physical feature – say, the length of a limb – as a real number falling between two extremes, suggesting an infinite number of possible lengths and therefore an infinite number of possible phenotypes. Yet that is an assumption of convenience and a radical one to take literally. To argue that such a feature cannot be approximated by a finite set of lengths is equivalent to arguing that differences in length are biologically significant even at the sub-atomic scale, where it is not clear what length mean. It also requires assuming that space itself is infinitely divisible – that is, continuous – an idea made universal by calculus, which requires it, but which is not known to be true ("Particle physics experiments have shown that space acts as a continuum down to distances of around 10^{-20} meters .... But there is absolutely no reason to think that discrete elements will not be found at still smaller distances." in Wolfram (2002, p 472, and "History of discrete space", p 1027)).

[2] Every subset of a set X constitutes a subbasis of some topology on X (Willard, 1970, p 39, theorem 5.6), so this parenthesis only clarifies how this procedure will use the accessibility relation. It is not an inference based on some property of the accessibility relation.

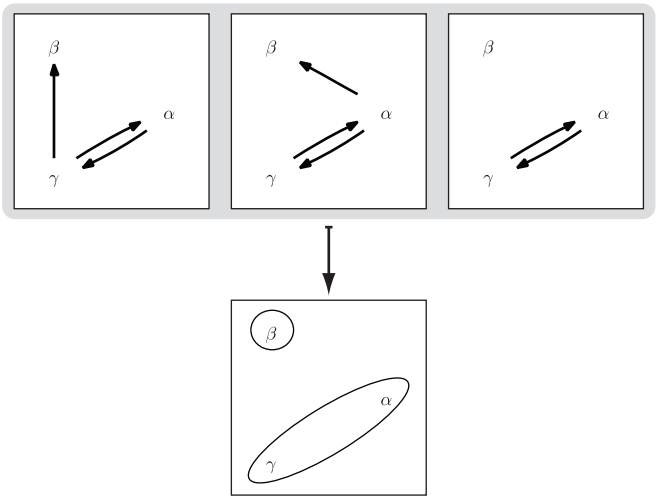


Figure 3.2: Three accessibility relations which yield the same topological space, given the topology-construction procedure prescribed in Cupal et al. (2000). An arrow shows when one phenotype is accessible to another. An ellipsis denotes an element of the (unique, non-redundant) basis set of the topological space.

can explicitly construct the unique non-redundant basis of τ:

$$\mathcal{B} = \left\{ B(x) = \bigcap_{y|x\in N(y)} N(y) \,\middle|\, x \in V \right\} \tag{3.4}$$

Here $N(y) = \{q \,|\, y \hookrightarrow q\}$ is the neighbourhood of y, the set of points accessible from y according to the accessibility relation, and $\mathcal{N} = \{N(x) \,|\, x \in V\}$ is the set of all neighbourhoods.

However, there is a problem. While the non-redundant basis is unique to the topology it generates, in the sense that the topology defines only one such basis, this basis is not unique *to the accessibility relation used to construct it*. In fact, two different accessibility relations will generate the same basis. In other words, following this procedure, two biological systems which have different accessibility relations will have identical topologies.

Figure 3.2 illustrates a simple example (calculations in section 3.7.5). In a space consisting of three possible phenotypes (α, β, and γ), three different accessibility relations all produce the same basis and hence the same topological space. But the relations are very different. They represent systems where only phenotype α can easily evolve into β, where only γ can easily evolve into β, and where none can easily evolve into β – presumably systems that a useful model could distinguish.

All of this can be put more succinctly. The procedure outlined by Cupal et al. (2000) defines a mapping, c, from accessibility relations to topologies. This procedure succeeds in

building valid topologies. However, this mapping is not an *information-preserving* mapping (i.e., it is not *injective*, or it is *many-to-one*). So this transformation of an accessibility relation to a finite topology destroys information about the system. From looking at the topology, you cannot tell much about what kind of accessibility relation produced it. In this sense, a topological space is a significantly less expressive formalism than an accessibility relation alone.

What is the reason for this problem? The algebraic reason is simple enough, and is visible in the calculations underlying the example above (in section 3.7.5). A certain possible conceptual reason, however, also suggests itself. It is as if this procedure for constructing a topology is inspired by the same conflation of "neighbourhood system" and "neighbourhood function" that was evident in Fontana and Schuster (1998a,b) and discussed in section 3.3.1.

Standard notation would make this mistake easy. A neighbourhood system is usually designated in calligraphic script. And if $\mathcal{N}$ were a neighbourhood system then it would be possible to convert it into a topology without loss of information, just as it is possible to interconvert among all five of the topological structures. In fact, if $\mathcal{N}$ *were* a valid neighbourhood system, and $N(y)$ *were* the family of neighbourhoods of y, then the procedure, as described by equation (3.4), *would* be the correct procedure for converting the neighbourhood system to a topology (proof in section 3.7.4). But $\mathcal{N}$ is defined not as a neighbourhood system, but just as the range of a single neighbourhood function. This all suggests that the topology-construction procedure was built on a correct understanding of how to find the basis of a topology, but a misunderstanding of how neighbourhoods define a topology.

3.3.3 Constructing an accessibility relation from a genotype-phenotype map destroys information

So far my criticism has been that past work either uses topological spaces incorrectly or in a way that destroys too much information. This suggests forgetting about topological spaces and just using the accessibility relation to construct the simpler construct of a directed graph. It is a more intuitive structure: it simply depicts which phenotypes can mutate easily into which others.

However, it is worth noting how this also yields an incomplete picture, since the ac-

cessibility relations also destroy some of the information that was used in their construction. In this case, it is destroying information in the original genotype-phenotype map. Like the mapping from accessibility relations to topological spaces, we can say that the mapping from genotype-phenotype maps to accessibility relations is *many-to-one* (i.e., not information-preserving).

This is not damning in itself. The question is whether the lost information is relevant to the biological phenomena we hope to understand. In the case of topological spaces we lost the power to distinguish the most basic kinds of relationships between phenotypes (e.g., $\alpha \hookrightarrow \beta$ versus $\alpha \not\hookrightarrow \beta$). Are accessibility relations "pitched" more correctly? Let us review what is lost.

First, an accessibility relation is built by imposing a threshold on an accessibility measure. This takes a real-valued quantity to a binary quantity, eliminating all gradations that do not cross the threshold. This can make the terms "accessible" or "not accessible" sound misleadingly absolute, since the accessibility measure which they summarise is fundamentally probabilistic. α may not be able to access β according to our relation but α might still evolve into β nonetheless as a result of pure chance. In fact, the condition of "not accessible" conflates pairs of genotypes which absolutely cannot mutate into each other (because they share no mutational boundary), and those which are merely unlikely to do so (because their accessibility measure is below our chosen threshold).

Second, the accessibility relation itself is blind to *structural* features of the genotype-phenotype map. This is because the accessibility measure is built by comparing gross areas in genotype space, rather than the shapes that fill those areas. But ignoring the shape of a pre-image means ignoring the mutational connectivity within that pre-image. Figure 3.3 illustrates the problem. Two different genotype-phenotype maps define $B = \phi^{-1}(\beta)$ the same but $A = \phi^{-1}(\alpha)$ differently. In one map, A includes two mutationally disconnected subsets A_1 and A_2. The subset A_2 has no mutational connection to B. If a point a' starts in this A_2 then it cannot mutate into B. It is a *dead end genotype*. However, equation (3.1) for the accessibility measure only considers the total area of these subsets, so the measure is the same for the isolated a' and for the connected a.

The accessibility measure is a useful abstraction. But the fact that it is defined over phenotypes, rather than contiguous genotypic regions, means that it must miss such intra-phenotypic structures. Do such structures matter? In later chapters, we will argue that

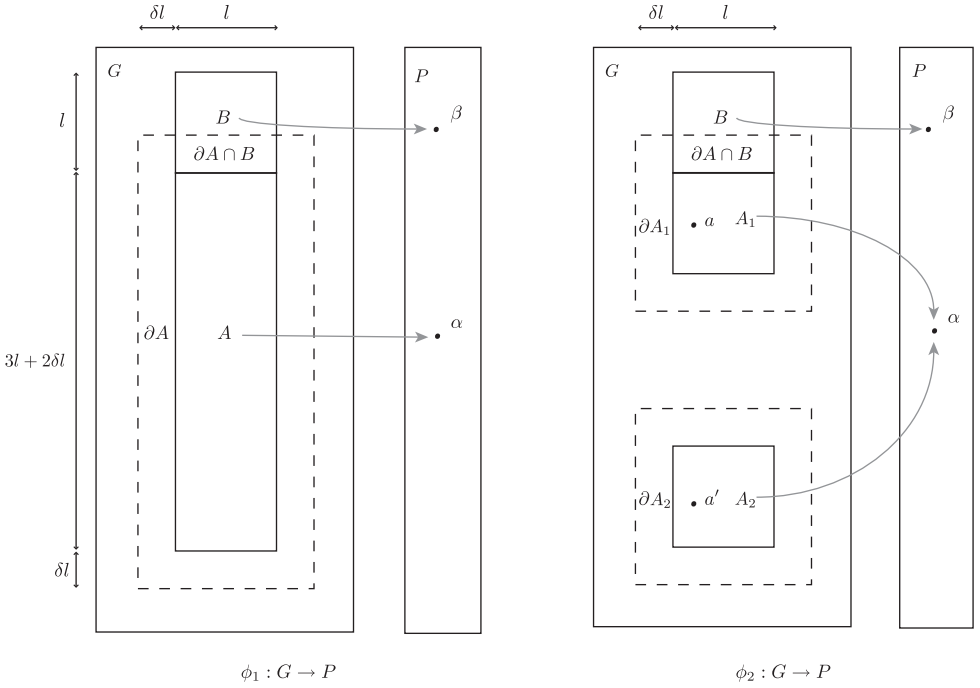


Figure 3.3: The accessibility measure cannot see genotypic dead ends, because it destroy structural information in the genotype-phenotype map. Two different maps, ϕ_1 and ϕ_2, give the same measure, $\kappa_1 = \kappa_2 = \frac{l\delta l}{2((l+2\delta l)^2 - l^2)}$. But since the measure ignores that pre-image $\phi^{-1}(\alpha) = A$ is split into mutationally disconnected components A_1 and A_2, it misses that the genotype a', unlike a, could never mutate into phenotype β. (Figure to scale)

structures like dead end genotypes are the most absolute examples of general properties of genotypic connectivity which determine whether or not a system is evolvable.

3.4 Defining an exact accessibility measure

The last section showed a problem with accessibility measures: they overlook information about the internal structure of pre-images, e.g., the presence of dead end genotypes. Another problem with existing measures is that it is unclear what they are measuring. They are supposed to measure the easiness of a phenotypic transition. This is implicitly a probabilistic argument, but the proposed measures are not obviously probabilities. In this section I will illustrate this by defining the exact, probabilistic accessibility measure, ω, and comparing it to Cupal's measure κ.

Recall that Cupal's measure κ, introduced in equation (3.1), is as follows:

$$\kappa(\alpha \to \beta) \stackrel{\text{def}}{=} \frac{|\partial A \cap B|}{|\partial A|}$$

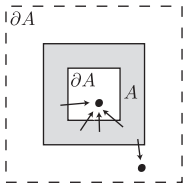


Figure 3.4: Non-neutral mutants of A are not uniformly distributed in ∂A after a mutational event, since some points in ∂A (like the one in the centre) will have a larger number of potential parents in A. κ ignores this by assuming a uniform distribution over mutants.

This is clearly defined but it is hard to know if if it is intended as an approximation. For instance, Cupal et al. (2000) write that κ is the "probability ... that a non-neutral mutant of an x-sequence folds into the shape y", where x and y are phenotypes, and an x-sequence is any genotype yielding phenotype x. But this is easily misunderstood. κ is this probability only if we choose the mutant with uniform probability from the set of all possible non-neutral mutants ∂A. But if we choose the non-neutral mutant from the distribution of non-neutral mutants that would exist *following an actual mutational event*, we get a different probability. That distribution is not likely to be uniform since certain mutants have more potential parents in A than others, as illustrated in figure 3.4. Comparing areas is equivalent to assuming a uniform probability distribution – that is, to assuming that every point in ∂A is equally likely, instead of summing over all points.

It is straightforward to define an alternative accessibility measure, ω, which is the exact probability of a phenotypic transition in the absence of selection. I first calculate this in full generality, then apply some basic approximations, and finally use it to calculate a corrected version κ' of Cupal's measure.

The system has two random processes, each described by a probability distribution. The first random process is the choice of which genotype a in a set A will suffer a mutation. We define a probability distribution for this,

$$d_A(a) \stackrel{\text{def}}{=} \text{probability } a \text{ will mutate,} \qquad (3.5)$$

$$\text{where } A \subset G$$

$$d_A : G \to [0,1]$$

$$\sum_{a \in A} d_A(a) = 1$$

The second process is the choice of where that mutation leads. We define the probability

distribution for this too,

$$\mu(a \to b) \stackrel{\text{def}}{=} \text{probability } a \text{ will mutate into b,} \qquad (3.6)$$
$$\text{where } \mu(a \to b) : G \times G \to [0,1]$$
$$\text{and } \forall a \in A, \sum_{b \in \partial(a)} \mu(a \to b) = 1$$

Having defined these two elementary probability distributions, there is now a well-defined probability for the event $A \to B$, a mutation within A leading to B. This probability is our new accessibility measure ω. We now calculate this value.

We start with the probability that a mutation in a will go to some point in the set B.

$$P(a \to B) = \sum_{b \in B} \mu(a \to b)$$

Then we sum over A weighted by d_A. d_A might be non-uniform if we wanted to account for the genotypic population distribution within a phenotype $\alpha = \phi^{-1}(A)$, or if certain genotypes are intrinsically more labile. Given a mutation in A, the probability that it will go into B, is as follows

$$\begin{aligned} \omega &= P(A \to B) \\ &= \sum_{a \in A} \sum_{b \in B} d_A(a) \mu(a \to b) \end{aligned} \qquad (3.7)$$

This is the probability, given that there is a mutation in A, that the mutant will be in B. If A and B represent the pre-images of two phenotypes α and β, then this is the exact probability that α will transition to β. This measure is exact and complete, but somewhat opaque.

To render it more usable we apply a couple simplifying assumptions concerning the two underlying random processes. First, if we assume a uniform distribution of mutations from a point, so that all of a point's mutants are equally likely, then

$$\mu(a \to b) = \begin{cases} \frac{1}{|\partial a|} & b \in \partial a \\ 0 & b \notin \partial a \end{cases}$$

$|\partial a|$ is just the size of a's mutational boundary. If it is the same for all points, then we can

replace $|\partial a|$ with $|\partial a'|$ where a' is just an arbitrarily chosen point in A. For instance, on a length-n genotype string, made from a symbol alphabet $\mathcal{A}$ of possible nucleotide values, there are n possible locations for a mutation and $|\mathcal{A}| - 1$ alternative values per position, so we can see that are $|\partial a'| = n(|\mathcal{A}| - 1)$.

Second, if we assume a uniform population distribution within the source region, so that every point in the source is equally likely to mutate, then

$$d_A(a) = \begin{cases} \frac{1}{|A|} & a \in A \\ 0 & a \notin A \end{cases}$$

Applying both these assumptions gives us

$$\begin{aligned} \omega|_{\text{approx}} &= \frac{1}{|A|} \sum_{a \in A} \sum_{b \in B \cap \partial a} \frac{1}{|\partial(a')|} \\ &= \frac{1}{|A||\partial(a')|} \sum_{a \in A} |\partial(a) \cap B| \end{aligned} \tag{3.8}$$

Now we can apply this to produce a corrected version of Cupal's accessibility measure κ. Recall that κ is meant to be the probability that A goes into B given a *non-neutral* mutation in A. First we use equation (3.7) to product a corrected version which is true in general:

$$\begin{aligned} \kappa' &= P(A \to B \mid A \to A^c) \\ &= \frac{P((A \to B) \cap (A \to A^c))}{P(A \to A^c)} && \text{(by conditional probability)} \\ &= \frac{P(A \to (B \cap A^c))}{(1 - P(A \to A))} \\ &= \frac{\sum_{a \in A} \sum_{b \in B \cap A^c} d(a)\mu(a \to b)}{1 - \sum_{a,b \in A} d(a)\mu(a \to b)} \end{aligned} \tag{3.9}$$

Applying the assumptions gives us something simpler:

$$\begin{aligned} \kappa'|_{\text{approx}} &= \frac{P(A \to (B \cap A^c))}{(1 - P(A \to A))} \\ &= \frac{\sum_{a \in A} |\partial(a) \cap B \cap A^c|}{|A||\partial(a')| - \sum_{a \in A} |\partial(a) \cap A|} && \text{by eq. (3.8)} \end{aligned}$$

The point is that even under the strong approximations that all elementary probabilities are uniform, this $\kappa'|_{\text{approx}}$ will still only equal Cupal's κ under very special circumstances –

such as when, as described above, every non-neutral mutation in A will generate mutants distributed uniformly over ∂A.

There is no reason to expect this in general. Reconsider figure 3.3, which illustrated the problem of dead end genotypes. In that figure, A_1 and A_2 are at minimum distance of $l + 2\delta l$ from each other. Each is assumed to define its own mutational boundary stretching outward δl in every direction. To suppose that all non-neutral mutants are uniformly distributed in ∂A is equivalent to saying that, in the figure, mutants from A_2 are as likely to end up in ∂A_1 as in ∂A_2. Furthermore, the claim would be that this is the case *no matter how far A_2 is from A_1*. This is quite unlikely.

3.4.1 Comparing accessibility measures

The measures discussed above are collected for comparison in table 3.1. It is also possible to derive a few inequalities relating these measures and others from earlier proposals, although their different sensitivities to the structure of the genotype set A prevents deriving very many results.

Before Cupal, in the course of work on RNA, Fontana and Schuster (1998a) and Fontana and Schuster (1998b, p 495) introduced two other accessibility measures, the *neighbourhood frequency* ν, and the *occurrence frequency* θ. These definitions seem to have been chosen partly to ease calculation given the RNA system under investigation. They are as follows[3]:

$$\begin{aligned} \nu(\alpha \to \beta) &= \frac{|A \cap \partial B|}{|A|} \\ \theta(\alpha \to \beta) &= \frac{|\partial A \cap B|}{(|\mathcal{A}| - 1)n\,|A|} \end{aligned} \tag{3.10}$$

ν is intended to "reflect the likelihood" of finding a phenotype β in the one-mutation neighbourhood of a randomly chosen genotype in $\phi^{-1}(\alpha)$; I do not consider it further. However, θ does have relationships with other measures we have considered.

First we note some basic facts. The factor $(|\mathcal{A}| - 1)n\,|A|$ just counts the total number of

[3] Cupal et al. (2000, p 19) restate these measures in a similar form, clearly defining them using Cupal's mutational *boundary operator* ∂_C, which we have abbreviated in this chapter as ∂. However, a close reading of the original definition of these measures in Fontana and Schuster (1998b, p 495) suggests that they were originally defined using the simpler *mutational operator* ∂_μ, which is introduced in the next chapter. The language is ambiguous. Fortunately this distinction is irrelevant here since the two measures ν and θ give the same value for both operators as long as they are being applied to disjoint sets of genotypes such as phenotype preimages.

	.	given uniform distributions of mutations in A, and from the mutating genotype
	ω	$\omega\|_{\text{approx}}$
given a non-neutral mutation in A	κ'	$\kappa'\|_{\text{approx}}$
given that we pick uniformly from non-neutral mutants of A	κ	

Table 3.1: Accessibility measures of the probability of a mutation from A to B. Different measures result from conditioning on different assumptions about the mutation event. Only ω represents the unconditional probability $P(A \to B)$

genotypes neighbouring A, including points in A itself. From this we can see that

$$|\partial(A)| \leq (|\mathcal{A}| - 1)n\,|A|$$

since ∂A is defined to exclude elements in A. Also we can see that

$$|\partial(A) \cap B| \leq \sum_{a \in A} |\partial(a) \cap B|$$

since the sum might double count certain $b \in B$ that are mutants of more than one $a \in A$.

Using these basic facts, and using our definition (3.10) of θ and definition (3.8) for $\omega|_{\text{approx}}$, we can see the following:

$$|\partial A \cap B| \leq \sum_{a \in A} |\partial(a) \cap B|$$

$$\frac{|\partial A \cap B|}{(|\mathcal{A}| - 1)n\,|A|} \leq \frac{1}{(|\mathcal{A}| - 1)n\,|A|} \sum_{a \in A} |\partial(a) \cap B|$$

$$\theta \leq \omega|_{\text{approx}}$$

This makes sense. It essentially reflects the fact that $\omega|_{\text{approx}}$ is double-counting instances when a possible mutant genotype $b \in B$ can be reached from multiple $a \in A$. This is because it is adding the probabilities of every mutational path from A to B, rather than simply

comparing overlaps: i.e., it is not conditioning on the assumption that there has already been a non-neutral mutation.

Also, it is straightforward to see that

$$
\begin{aligned}
(|\mathcal{A}| - 1)n\,|A| &\geq |\partial(A)| \\
\frac{1}{(|\mathcal{A}| - 1)n\,|A|} &\leq \frac{1}{|\partial(A)|} \\
\frac{|\partial A \cap B|}{(|\mathcal{A}| - 1)n\,|A|} &\leq \frac{|\partial A \cap B|}{|\partial(A)|} \\
\theta &\leq \kappa
\end{aligned}
$$

This also makes sense, since θ's denominator is overcounting mutants.

However, it is not possible to conclude anything general about the relation between $\omega|_{\text{approx}}$ and κ. We might expect $\omega|_{\text{approx}} \leq \kappa$ since, since κ is a conditional probability that already assumes a non-neutral mutation. However, it is not possible to prove this is generally the case, because the actual value of κ will depend on its denominator ∂A, which will depend on the actual structure of the set A.

This can be seen, using definition (3.1) for κ, by noting that

$$
\begin{aligned}
\omega|_{\text{approx}} &\sim^{?} \kappa \\
\frac{\sum_{a \in A} |\partial(a) \cap B|}{|A|n(|\mathcal{A}| - 1)} &\sim^{?} \frac{|\partial A \cap B|}{|\partial A|} \\
\frac{Q}{R} &\sim^{?} \frac{S}{T} \\
\text{where } Q = \sum_{a \in A} |\partial(a) \cap B| &\text{ and } S = |\partial A \cap B| \\
R = |A|n(|\mathcal{A}| - 1) &\text{ and } T = |\partial A|
\end{aligned}
$$

where our inequalities above only give us that $Q \geq S$ and $R \geq T$. So the ultimate relation between $\omega|_{\text{approx}}$ and κ will depend on those terms.

3.5 Pretopologies

The last phase of the topology work suggested using not topologies, but pretopologies (or generalised topological spaces). These pretopologies are also built on accessibility relations,

so they would suffer the same blind spot discussed in section 3.3.3. However, they do seem more promising than topologies. Since pretopologies are not typically used in the analysis of biological systems, the recent papers introducing the idea have taken the liberty of explaining the fundamental axioms used to define them (Stadler et al., 2001; Stadler and Stadler, 2002; Wagner and Stadler, 2003). In this spirit, it may be helpful to provide some background information on this area.

First, a pretopology is an esoteric construct not only within biology but also within pure mathematics. As discussed earlier, one meaning of "topology" is a part of the definition of a topological space. A topological space consists of a set X of elements, and a set T of subsets of X satisfying certain conditions. The set X is the space. The set T is known as the "topology" on X.

It is by weakening the conditions on T required for a topology that we allow another kind of construct, a pretopology. However, while the topological space is fundamental to much other work in mathematics, pretopologies are not widely studied. They are not mentioned at all in some standard reference works (Sugakkai, 1993; Weisstein, 2002) and appear infrequently in the research literature [4]. In fact, the notion is sufficiently esoteric that authors define it incompatibly depending on how they loosen the standard conditions of a topology. What Stadler et al. (2001) calls a pretopology, Belmandt (1993) identifies as a $\mathcal{U}_D$-type pretopological space.[5] Therefore, while it is plausible to hope that the rich

[4] Belmandt (1993) surveys the past work in a paragraph (translation mine):

> The first efforts aiming to construct a "topology of fewer axioms" are due to Fréchet (1928) and date from the end of the 1920s. They were continued by Appert in the middle of the '30s. One finds the same ideas in several works of Portuguese origin, those from A. Monteiro at the end of the '40s, a little later in France from Ky Fan. Some analogous ideas were developed by P.C. Hammer in the United States in the '60s while, at the same time, Čech (Čech, 1966) dedicated a part of his work to "topological" structures freed from the axiom of idempotence.

To give a sense of proportion we may also note that, as of 7 May 2006, a search on the academic reference service MathSciNet, returned 210,098 hits on "topology" and 21,955 hits on "topological space", compared to 60 hits on "pretopology", 64 on "pretopological space", and 28 on "generalized topological space".

[5] Stadler et al. (2001) offer their definition of a pretopological space in section 4.3, page 22. They define a pretopological as the pair $(X, \mathcal{N})$, where X is the base set and $\mathcal{N}$ represents what they call a neighbourhood system. The term neighbourhood system comes from standard topology where it is one of the five equivalent structures used for defining a topological space (as discussed in section 3.7.2). Given that they describe a modification of the standard definition, we might also call their construct a *pseudo-neighbourhood system*. It is the result of dropping the fourth axiom from the standard definition of a neighbourhood system presented in Sugakkai (1993), and quoted in section 3.3.1.

Belmandt (1993, chapter 2) defines a hierarchy of different kinds of pretopological spaces of increasing structure: a plain pretopological space, a pretopological space of type $\mathcal{U}$, of type $\mathcal{U}_D$, and of type $\mathcal{U}_S$, and finally a traditional topological space. His definitions are framed in terms of a *pseudo-closure operator*, although equivalent definitions are discussed. Tables 1 and 2 in Stadler and Stadler (2002) summarise relations between closure axioms and neighbourhood axioms, and a taxonomy of topological structures based on the definitions in Čech (1966). It is not straightforward to determine if any of Belmandt's notions are mathematically equivalent to those in Stadler et al. (2001), since Belmandt defines his closure operator slightly differently.

literature on topological spaces might provide ideas relevant to biological systems, this is unlikely for pretopologies because that literature does not yet exist.

However, while pretopologies have been historically neglected, it is not only in biology that they have recently attracted new attention as a model for empirical data analysis. The most significant work in this area is the book-length treatment, *Manuel de prétopologie*, by the research group publishing under the pseudonym Belmandt (1993). This develops the mathematical fundamentals and starts towards developing applications in social sciences, game theory, graph theory, image recognition, and other areas. Others have followed in these areas (Meziane et al., 1997; Bonnevay and Largeron, 2000; Largeron and Bonnevay, 2002), as well as developing it further as a general technique of applied mathematics (Pagliani, 2004). Citation patterns indicate this literature has not informed any of the biological work. This may indicate a useful direction for future research.

3.6 Conclusion

I have reviewed the efforts to use elementary topology to analyse evolutionary systems. These efforts originally sprang from the realisation that a traditional quantitative genetic model organises phenotypes in a metric space, using their measurable physical traits as the dimensions of the space. A better model would treat physical traits as incidental, and organise phenotypes according to their underlying genetic relationships. This led to the idea of using a topological space, which is based on a more general notion of relationship than a metric space.

However, efforts to use topological spaces have been hampered by a usage of the term inconsistent with the mathematical literature. Furthermore, the only published procedure that produces a valid topological space of phenotypes destroys too much biological information for that space to be a useful representation. For the purposes of evolutionary analysis, topological spaces seem like an interesting dead end.

That being said, the work on topologies also introduced the idea of an accessibility relation, a simpler idea that can be used to represent phenotypes as nodes in a directed graph. While accessibility relations also destroy significant information, such as the existence of dead end genotypes, they are highly intuitive representations of certain biological facts. This augurs well for work which stays "closer" to these relations, by using structures like

graphs or perhaps pretopologies.

It is easy to see why topological spaces *seem* like a good candidate as a new modelling construct. It is true that traditional models organise phenotypes by their physical features and that it would be preferable to do so by their mutational relationships, a more abstract notion. It is also true that traditional models use metric spaces and that topological space are an abstract generalisation of metric spaces. However, this does not imply that topological spaces are the correct abstraction for phenotypes. They turn out not to be. One basic problem, surely, is that many of the distinctions in elementary topology were developed in order to handle the counter-intuitive structures possible in infinite spaces. As biology only allows a finite number of genotypes and phenotypes, these distinctions only add confusion.

The creative challenge is to find mathematical constructs suited to the question they address – in this case, evolvability. In the next chapter, I will present a model demonstrating an instance where the accessibility relations destroys information specifically necessary to distinguish systems of different evolvability. In the course of doing so, I will also develop a new formalism which allows succinct statement, rigorous proofs, and summary diagrams about the structure of genotype space. These structures provide one way to compare and predict the evolvabilities of different systems.

3.7 Appendix

This appendix gathers here for convenience a brief summary of the basic definitions of a metric space, a topological space, and a basis.

It also gives two proofs about building topological spaces from non-redundant bases and neighbourhood systems. These proofs indicate how a past work seems to have run afoul of a crucial subtlety in the difference between a neighbourhood function and a neighbourhood system.

3.7.1 Metric space

A *metric space* is a set X together with a *distance metric*, a real-valued function $d : X \times X \to \Re$ with the following properties (Sugakkai, 1993, "Metric Spaces", p 1014):

- $d(x, y) \geq 0$, with $x = y \iff d(x, y) = 0$ (identity)

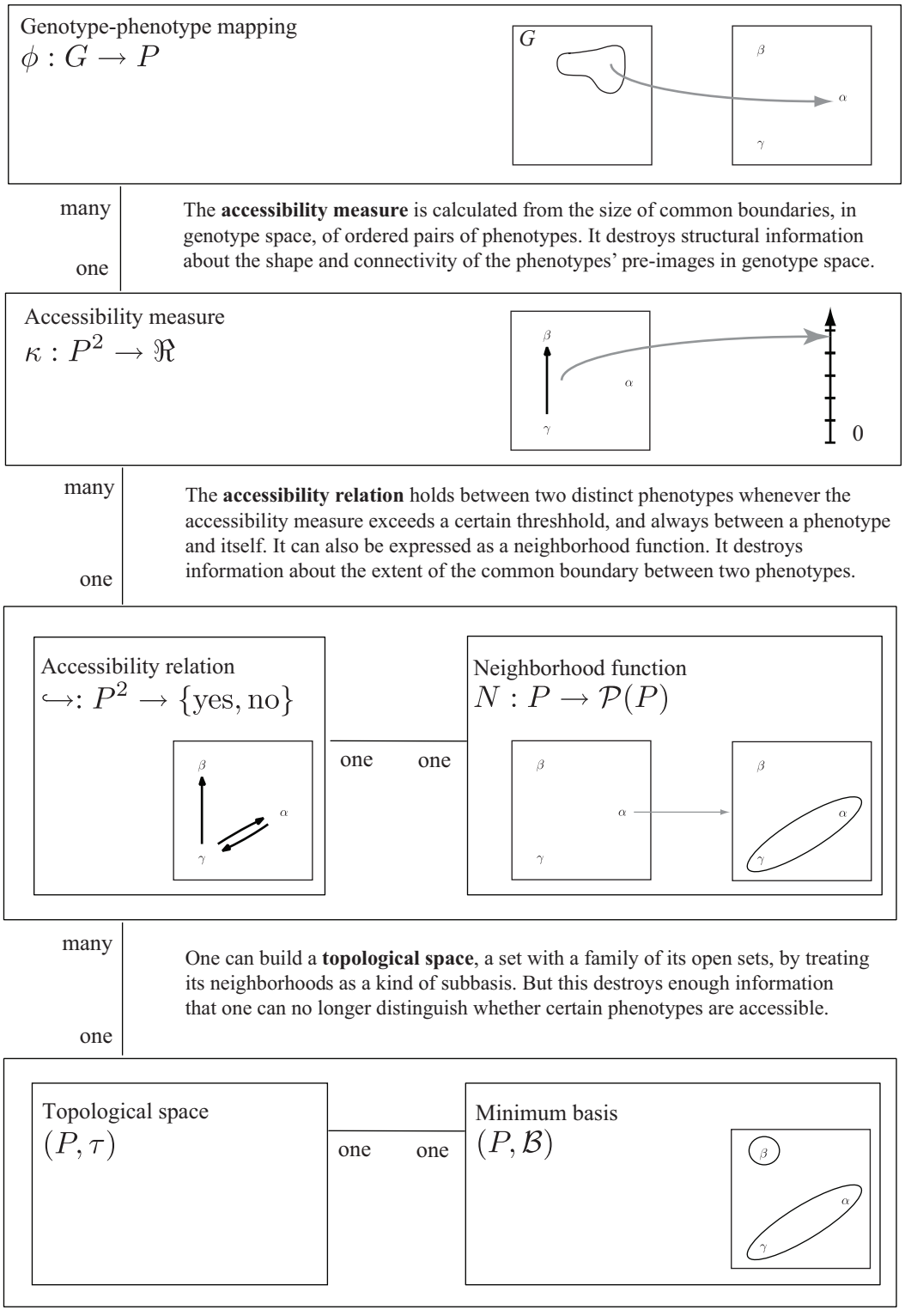


Figure 3.5: Multiplicities between constructs for representing genotype-phenotype mappings. Every many-to-one relationship is a step where representational information is lost. One-to-one relationships mark equivalent formulations.

- $d(x, y) = d(y, x)$ (symmetry)
- $d(x, z) \leq (d, x) + d(y, x)$ (triangle inequality)

The real numbers are an example metric space. If you measure a quantity by assigning it a real number, and compare two measurements by looking at the absolute difference of the two numbers, then you are using a metric space, where $X = \mathfrak{R}$ and $d(x, y) = |y - x|$. Euclidean space is a metric space, where $X = \mathfrak{R}^3$ and $d(v, v') = \sqrt{(\delta x)^2 + (\delta y)^2 + (\delta z)^2}$. Virtually any multivariable quantitative model is built on metric spaces.

Every metric space has an underlying topological space. An *open ball* around a point x is the set $B_\epsilon(x) = \{y \,|\, d(y, x) < \epsilon\}$. This is known as the metric topology, since the set of open balls forms a basis that generates an underlying topological space.

However, not every topological space can be represented by a metric space. Some topological spaces are *non-metrizable*, in the sense that there exists no metric d which yields that topological space (Steen and Seebach, 1978, p 37). It is because of this that a topological space is a strictly more general formalism than a metric space. If one believed topological spaces provided some special flexibility needed for modelling phenotypes, a flexibility that metric spaces lacked, then one would only use non-metrizable topological spaces.

3.7.2 Topological space

The following definitions can be found in any reference work or introductory textbook on general topology. These below paraphrase the terse summaries in Sugakkai (1993, p 1606, "Topological Spaces").

A *topological space* is a set X together with a family $T \subset \mathcal{P}(X)$ of subsets on X such that:

1. $X \in T$ and $\emptyset \in T$
2. the intersection of any two sets in T is itself in T ("closed under finite intersection")
3. the union of any collection of sets in T is itself in T ("closed under arbitrary union")

T is called the *topology* on X. The elements of T are called the *open sets* of X. For instance, the discrete topology, $T = \mathcal{P}(X)$, is the family of every possible subset of X. At the other extreme is the indiscrete topology, consisting of only the empty set and the entire base set, $T = \{\emptyset, X\}$.

The notion of open sets is one of the five equivalent structures used for defining a topological space. Other structures include: defining *closed sets* of the base set, defining a *closure operator* on the base set, defining an *interior operator* on the base set, and defining a *neighbourhood system* on the base set.

Topological basis

Like the topology a *topological basis* $B \subset \mathcal{P}(X)$ is also a family of subsets of X (defined in Sugakkai (1993, p 1608) and Willard (1970, p 38, Theorem 5.3)). A basis is a basis *of* a topology, which is it said to "generate". B is the basis of a unique topological space (X, T) if and only if it has the following properties:

1. $B_i \in T \ \forall i$
2. $\cup_j B_j = X$
3. $B_1, B_2 \in B$ and $x \in B_1 \cap B_2 \implies \exists B_3 \in B$ such that $x \in B_3 \subset B_1 \cap B_2$

B is called the basis of the topological space (X, T) because, when these conditions are met, every open set $t \in T$ can be expressed as the union of elements of B.

A *subbasis* of a topology τ is "a collection $C \subset \tau$ such that the collection of all finite intersections of elements from C forms a base [basis] for τ"(from Willard (1970, p 39), and identically in Sugakkai (1993, p 1608)).

Separation axioms

Topological spaces are taxonomised according to the topological *separation axioms*. Each axiom defines a condition which may hold for a topological space. Many of these have descriptive or historical names (e.g., a Hausdorff space, a completely Hausdorff space, a Tychonoff space), as well as a more compact designation based on abbreviations and a counting convention (e.g. and respectively, T_2, CT_2, $T_{2\frac{1}{2}}$). Considering how one axiom implies another yields a set of relationships among the usual topological spaces, relationships which appear in standard texts(Willard, 1970; Schechter, 1997; Sugakkai, 1993; Steen and Seebach, 1978).

In the case of finite topological spaces, these relationships simplify even further. Also, since it is possible to define a unique non-redundant basis for a finite topological space,

it is convenient to characterise finite spaces in terms of these bases. Cupal et al. (2000) summarises these simplifications. For instance, for finite spaces, a T1 space must be the discrete topology. Since all T2 spaces (Hausdorff spaces) are T1 spaces, if a finite space is Hausdorff then we know it is the discrete topology.

How is this relevant? The separation axioms and the resulting taxonomy of topological spaces constitute a large part of the elaboration of elementary topology. The fact that they are mostly irrelevant to finite topological spaces – that is, to all biological topological spaces – suggests that these refinements will not be useful to biology.

3.7.3 Unique non-redundant basis

Every basis generates a unique topology. But a topology can be generated by multiple different bases. So normally, one cannot check that two topologies are different just by checking that their bases are different, since those two different bases might generate the same topology. However, unlike topological spaces in general, *every finite topological space defines a unique non-redundant basis*.

This is relevant in two ways. First, it means one can compare finite topologies just by comparing their non-redundant bases. One can in effect use the non-redundant basis in lieu of the topology itself in proofs, discussions, and diagrams. This is convenient since a basis is easier to handle than a topology. The technical analysis in this chapter uses the basis in this way – for instance, in figure 3.2, which depicts a topology simply by drawing ellipses to represent the sets in the basis.

Second, this uniqueness is worth understanding closely in order to be clear about what it does not imply. While every finite topological space defines a unique non-redundant basis, that basis is not unique to the accessibility relation used to construct it. That is, the mapping from accessibility relations to bases is many-to-one, even if the mapping from non-redundant bases to topologies is one-to-one.

Proof of uniqueness

Cupal et al. (2000, p 15) proves that every finite topological space defines a unique non-redundant basis. I do not understand his proof, which is rather terse, so I present here a more explicit proof of the uniqueness of the non-redundant basis.[6]

[6]Since Cupal's proof is short, we reproduce it here in its entirety:

The procedure for calculating the non-redundant basis is as follows. We begin with a topological space (X, T), where X is the base set and T is the family of open sets of X (i.e., the topology on X). Now we define

$$T_x = \{t \,|\, x \in t, t \in T\} \tag{3.11}$$

That is, T_x is the family of only those open sets which contain the point x. We further define

$$J(x) = \cap_{t \in T_x} t \tag{3.12}$$

and

$$\mathcal{J} = \{J(x) \,|\, x \in X\} \tag{3.13}$$

First we will now show that $\mathcal{J}$ is a basis of T. We refer to the definition of a basis provided in section 3.7.2.

From the definition of a topological space, we know that the intersection of a finite number of open sets is an open set. Since X is finite, there are only a finite number of open sets, so any intersection of them is itself an open set. Since $J(x)$ is the intersection of open sets, $J(x)$ is an open set. This proves condition 1.

By definition (3.12), we know that $x \in J(x)\ \forall x \in X$. Therefore, $X = \cap_x J(x)$, proving condition 2.

Now we establish condition 3. Consider some point x such that $x \in J_1 \cap J_2$. Since definition (3.12) lets us index $\mathcal{J}$ on X, we may restate it as $x \in J(x_1) \cap J(x_2)$. Knowing this, we must show that there exists a $J_3 \in \mathcal{J}$ such that $x \in J_3 \subset J(x_1) \cap J(x_2)$. Let us try $J_3 = J(x)$. This ensures that $x \in J(x) = J_3$, but we must still show that $J(x) \subset J(x_1) \cap J(x_2)$.

Assume the contrary, that $J(x) \not\subset J(x_1) \cap J(x_2)$. This would mean that there exists $y \in J(x)$ such that $y \notin J(x_1)$, $y \notin J(x_2)$, or both. Without loss of generality, assume the problem is that $y \notin J(x_1)$. If $y \in J(x)$, then from definition (3.12) it is the case that $y \in t\ \forall t \in T_x$.

Lemma 1. The family $\mathcal{B} = \{B(x) \,|\, x \in V\}$, where $B(x)$ is the intersection of all open sets containing x, forms the unique non-redundant basis of that topology τ.

Proof. Since V is finite, $B(x)$ is an open set. it is clear from the definition that for each $z \in B(X) \cap B(y)$ hold $B(z) \subset B(x) \cap B(y)$ [sic], i.e., the intersection of two elements of $\mathcal{B}$ contains another basic element, and the union of elements of $\mathcal{B}$ is open. Thus $\mathcal{B}$ is a basis of τ. Uniqueness can be verified as follows: suppose there is another basis $\mathcal{B}'$. Then, by definition $B(x)$ is the union of elements of $\mathcal{B}'$. Hence either $B(x) \in \mathcal{B}'$, or $B(x)$ is a union of open sets each of which is strictly smaller than $B(x)$. This is impossible since $B(x)$ is the smallest open set containing x. Thus $B(x) \in \mathcal{B}'$ for all x and hence $\mathcal{B}' = \mathcal{B}$.

But our original premise $x \in J(x_1) \cap J(x_2)$ implies that that $x \in J(x_1)$. Since $J(x_1)$ is an open set, this further implies that $J(x_1) \in T_x$. But if $J(x_1) \in T_x$, and $y \in t \; \forall t \in T_x$, then that would then mean that $y \in J(x_1)$. This contradicts our contrary assumption that $J(x) \not\subset J(x_1) \cap J(x_2)$, proving it wrong. Therefore, we have established that $x \in J(x) \subset J(x_1) \cap J(x_2)$, which is what was required for condition 3. Having shown 1, 2, and 3, we have shown that $\mathcal{J}$ is a basis. We know it is a basis of T specifically, since all arguments above use "open set" to mean an element of T, and refer to the same X on which T is defined.

To see that that $\mathcal{J}$ is a *unique* basis of T, it suffices to recognise that our definition of $\mathcal{J}$ in terms of T constitutes a valid functional mapping $f : \mathcal{J} \mapsto T$. That is, it defines only a single $\mathcal{J}$ for any T. We have shown above that $\mathcal{J}$ is a basis of T, and since we already know that every basis defines a unique topology, this suffices to show that $\mathcal{J}$ is generated only by one topology T. This completes the proof.

3.7.4 Converting a neighbourhood system to a topology

Here we show that if $\mathcal{U}$ is a valid neighbourhood system for a finite topological space, then then the unique non-redundant basis of that topological space is given by a minor variant of the construction procedure in Cupal et al. (2000). This result is of interest, because it shows that equation (3.4) in Cupal et al. (2000) would create a topology without destroying information *if* it had been defined using a valid neighbourhood system. This strongly suggests the same misunderstanding of "neighbourhood system" that appears in Fontana and Schuster (1998a,b).

Proof Here we show that if $\mathcal{U}$ is a valid neighbourhood system for a finite topological space, then then the unique non-redundant basis of that topological is space is $\mathcal{B}$ where

$$\mathcal{B} = \{\cap_{U \in \mathcal{U}(x)} U \,|\, x \in X\} \tag{3.14}$$

From the previous section, we already know that for a topology T if

$$T_x = \{t \,|\, x \in t \in T\}$$

then

$$\mathcal{B} = \{\cap_{t \in T_x} t \,|\, x \in X\}$$

is the non-redundant basis of the topological space (X, T). Therefore, it suffices to show that, for all $x \in X$,

$$\cap_{t \in T_x} t = \cap_{U \in \mathcal{U}(x)} U \tag{3.15}$$

Sugakkai (1993, p 1607) describes the neighbourhood system of a topological space in terms of the topology of the space:

> Assume that a system of open sets O ... is given. In this case, each member of O is called an *open set*. A subset U of X is called a *neighborhood* of a point x in X provided that there is an open set O such that $x \in O \subset U$. If $\mathcal{U}(x)$ is the family of all neighborhoods of x, the function $x \to \mathcal{U}(x)$ satisfies [the axioms for a neighborhood system].

If $\mathcal{U}$(x) includes every subset which contains an open set containing x, then it also includes those open sets themselves. In other words,

$$T_x \subset \mathcal{U}(x)$$

So we can write

$$\begin{aligned} B(x) &= \cap_{U \in \mathcal{U}(x)} U \\ &= \left(\bigcap_{\substack{U \in \mathcal{U}(x) \\ U \in T_x}} U \right) \bigcap \left(\bigcap_{\substack{U \in \mathcal{U}(x) \\ U \notin T_x}} U \right) \\ &= \left(\bigcap_{t \in T_x} t \right) \bigcap \left(\bigcap_{\substack{U \in \mathcal{U}(x) \\ U \notin T_x}} U \right) \end{aligned}$$

Now to show the equality (3.15) we need only show that the second term does not affect the value of the expression a whole. The only way it could, is if there were some $U' \in \mathcal{U}(x)$ which was strictly smaller than all the $t \in T_x$. If this U' existed, by the definition of $\mathcal{U}(x)$, it would need to contain some open set. But if it did, this open set would have to belong to T_x by the definition of T_x. This shows that this U' does not exist. This proves our result.

3.7.5 Topology construction example

The following example show how two different accessibility relations (equivalently, neighbourhood functions) yield the same topological basis, following the procedure described in

Cupal et al. (2000), represented by equation (3.4).

We work out the algebra associated with the leftmost and rightmost accessibility relations depicted in figure 3.2. We name their neighbourhood functions M and N.

The problem is simple. We might expect $B(\gamma)$ or $B(\beta)$ to differ between the two different systems M and N because $\gamma \hookrightarrow_M \beta$ while $\gamma \not\hookrightarrow_N \beta$. However, that difference is "canceled out" by the intersection operator which tends to reduce the connectivity to the lowest common denominator:

	M	N
neighbourhood	$M(\alpha) = \{\alpha, \gamma\}$ $M(\beta) = \{\beta\}$ $M(\gamma) = \{\alpha, \beta, \gamma\}$	$N(\alpha) = \{\alpha, \gamma\}$ $N(\beta) = \{\beta\}$ $N(\gamma) = \{\alpha, \gamma\}$
$B(\alpha)$	$B_M(\alpha) = \bigcap_{y\|\alpha\in M(y)} M(y)$ $= \bigcap_{\{\alpha,\gamma\}} M(y)$ $= M(\alpha) \cap M(\gamma)$ $= \{\alpha, \gamma\} \cap \{\alpha, \beta, \gamma\}$ $= \{\alpha, \gamma\}$	$B_N(\alpha) = \bigcap_{y\|\alpha\in N(y)} N(y)$ $= \bigcap_{\{\alpha,\gamma\}} N(y)$ $= N(\alpha) \cap N(\gamma)$ $= \{\alpha, \gamma\} \cap \{\alpha, \gamma\}$ $= \{\alpha, \gamma\}$
$B(\beta)$	$B_M(\beta) = \bigcap_{y\|\beta\in M(y)} M(y)$ $= \bigcap_{\{\beta,\gamma\}} M(y)$ $= M(\beta) \cap M(\gamma)$ $= \{\beta\} \cap \{\alpha, \beta, \gamma\}$ $= \{\beta\}$	$B_N(\beta) = \bigcap_{y\|\beta\in N(y)} N(y)$ $= \bigcap_{\{\beta\}} N(y)$ $= N(\beta)$ $= \{\beta\}$
$B(\gamma)$	$B_M(\gamma) = \bigcap_{y\|\gamma\in M(y)} M(y)$ $= \bigcap_{\{\alpha,\gamma\}} M(y)$ $= M(\alpha) \cap M(\gamma)$ $= \{\alpha, \gamma\} \cap \{\alpha, \beta, \gamma\}$ $= \{\alpha, \gamma\}$	$B_N(\gamma) = \bigcap_{y\|\gamma\in N(y)} N(y)$ $= \bigcap_{\{\alpha,\gamma\}} N(y)$ $= N(\alpha) \cap N(\gamma)$ $= \{\alpha, \gamma\} \cap \{\alpha, \gamma\}$ $= \{\alpha, \gamma\}$
basis	$\mathcal{B}_M = \{\{\alpha, \gamma\}, \{\beta\}\}$	$\mathcal{B}_N = \{\{\alpha, \gamma\}, \{\beta\}\}$

Chapter 4

Evolvable segmentation and the genospace algebra

Chapter 1 explained the idea of organismic evolvability, the capacity of some developmental architectures to translate normal genetic variation into greater phenotypic variation. This capacity is supposed to be due to core components, specific features of the organism which allow relatively few mutations to produce relatively large and coherent phenotypic changes. These features give a sort of levers-versus-machinery architecture to development, where a few regulatory elements exert a kind of abstracted control over the phenotype as a whole.

Chapters 2 reviewed how past efforts to create formal models of evolvability tended to oversimplify it in order to fit existing quantitative methods. Chapter 3 reviewed the ambitious efforts to develop new methods based on elementary topology, efforts which encountered technical problems but generated some valuable ideas. These reviews illustrate the challenge in modelling evolvability, which is to balance a number of explanatory requirements.

A model must be centred not on fitness but on the genotype-phenotype map, since only the map focuses on the essential role of the developmental architecture. A model must describe not just differences between individuals but between broad organismic kinds, like species, since biological contrasts in evolvability are between such kinds. Also, a model must give insight into what causes evolvability. Rather than being a black box like a fitness landscape, it must expose its working parts. One must be able to see how certain formal features of a map are intrinsic indications of evolvability, and also how certain other formal features of a map represent physical features of the particular developmental system under

study.

This chapter presents such a model. This model illustrates one of the most frequently discussed mechanisms supporting evolvability: the capacity for repeated body segments. We compare two genotype-phenotype maps of differing evolvability, ϕ and ϕ_E. These map from the same genotype space G and to the same phenotype space P. But we will see that, in this model, the greater facility for segmentation in ϕ_E makes it more evolvable.

This value of this model is heuristic. It provides examples of

- one kind of abstracted control – the capacity of the evolvable mapping to specify only the *number* of repetitions without describing each repeated segment separately
- the explicit definition of such a mapping
- intrinsic formal indications of a mapping's increased evolvability compared to a naive mapping – the reduction of dead end genotypes; various kinds of increased connectivity between phenotypes
- a new formalism for proving, summarising, and diagramming results about such systems

4.1 Two genotype-phenotype maps

First we define the possible genotypes, how they are connected by mutation, and the possible phenotypes. What are the possible genotypes? Let us use a length-16 binary genome. That is, every genotype consists of 16 loci, and every locus can have an allelic values of 0 or 1. The genotype space G, the set of all genotypes, is simply all 2^{16} of such strings.

$$G = \{\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0},$$
$$\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{1},$$
$$\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{0},$$
$$\vdots$$
$$\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\boxed{1}\}$$

How are these genotypes connected? They are connected by mutation, and we assume mutation is the switching of the allelic value of a single locus. So then $\partial(\gamma)$ is the set of sixteen genotypes formed by changing every one of those sixteen loci of the genotype γ.

What phenotypes are possible? As we are modelling only segmentation, a phenotype is completely defined by one feature – the number of thoracic body segments. Let us allow this to range from zero to three. This gives us our phenotype space P, the set of all possible phenotypes:

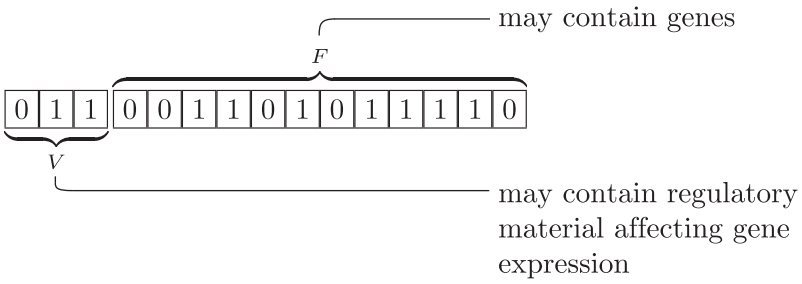
$$P = \{$$ $$\}$$

Having defined the genotype space and the phenotype space, we can now define genotype-phenotype maps. Recall that a genotype-phenotype map is a function which represents a particular developmental system by associating every genotype in G with some phenotype in P.

First we consider the mapping $\phi : G \to P$, which we call the *naive mapping*. This is because of how it directly expresses every gene without any intermediating regulation that augments its evolvability. In this model a "gene" provides all instructions for how to build a body segment. As we presume this to be a nontrivial operation, a gene must be a nontrivial block of genetic information. Let us say the "gene" for describing how to build a thoracic body segment is $\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}$ – that is, three consecutive 1s bounded by 0s, which may be regarded as analogous to start/stop codons.

$\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0} \longrightarrow$

To highlight the contrast between genes and regulation, let us also assume that genes may only appear in a gene coding part of the genome. This is the last twelve loci, which we call the *F-part* (or the fixture). The first three loci make up the *V-part* (or the variant), and they are reserved for the kind of regulatory material which is ignored by the naive mapping:



Given these parameters, we can now formally define the naive mapping ϕ as follows:

$$\phi(\gamma) = \begin{array}{l}\text{phenotype with as many seg-}\\ \text{ments as there are instances in}\\ \gamma_F \text{ of } \boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\end{array} \tag{4.1}$$

So for instance γ has one gene

$$\gamma = \overbrace{\boxed{0}\boxed{0}\boxed{0}}^{V}\overbrace{\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{0}}^{F}$$

and therefore one body segment:

$$\phi(\overbrace{\boxed{0}\boxed{0}\boxed{0}}^{V}\overbrace{\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{0}}^{F}) =$$

We allow two genes to share stop and start markers, so that, for instance,

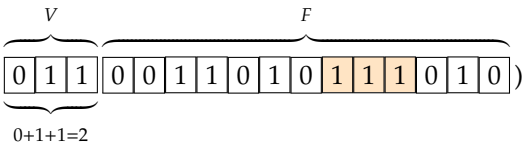
$$\phi(\overbrace{\boxed{0}\boxed{0}\boxed{0}}^{V}\overbrace{\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{0}}^{F}) =$$

To summarise, under the mapping ϕ, every body segment in the phenotype requires a corresponding gene to code for it. This is consistent with imagining a developmental system that merely processes F linearly, transcribing every gene as it was found.

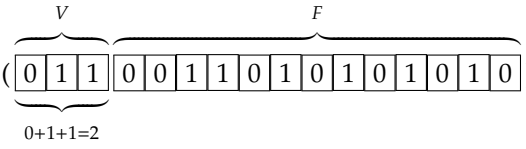
In contrast, we call $\phi_E : G \to P$ the *evolvable mapping*. Rather than expressing every gene directly, regulatory material determines how many times this genetic information is expressed. If there is no gene describing a body segment, then there is no segment in the phenotype. But if there is at least one valid gene, then the number of loci set in the regulatory region V determines how many times that information is expressed and thus how many segments are created. Formally,

$$\phi_E(\gamma) = \begin{cases} 0 \text{ segs} & \text{if no valid gene in } \gamma_F, \\ (\sum_{i \in V} \gamma[i]) \text{ segs} & \text{otherwise.} \end{cases} \tag{4.2}$$

So for instance,

$\phi_E($  $) =$ 

but at the same time

$\phi_E($  $) =$ 

because no gene is present. (Further examples are offered in table 4.4.) This developmental system features a form of abstracted control, a separation of the business of describing how to build a body segment from describing how many times to follow that description.

It is worth pausing to compare ϕ and ϕ_E. They have the same range, P, since there is no phenotype which can be realised only through one mapping. They have the same domain, G, since they both have a defined behaviour for every possible genotype. They differ only in their mapping, and that difference is a stylised representation of a biological mechanism for repeated body segmentation. All their similarities allow an exact, evenhanded comparison of that difference.

In the following sections we will show that the naive mapping ϕ produces a genetic dead end of the type depicted in figure 3.3, while ϕ_E does not. Recall that a dead end is a feature of the genotypic pre-image of a genotype-phenotype mapping. The mapping partitions the sets into phenotypic pre-images $\{\phi^{-1}($  $), \phi^{-1}($  $), \ldots\}$, and there is dead end when one of those pre-images $\phi^{-1}(j)$ includes a subset which does not share the same mutational connections to other phenotypes as the pre-image set $\phi^{-1}(j)$ as a whole. Then the accessibility relation, which treats the set as a whole, will misrepresent the connections of points in that subset.

The following analysis is somewhat technical. It dispenses with phenotype cartoons, treating $\{$  , ,  , $\}$ as $\{0, 1, 2, 3\}$. It also uses algebraic arguments. However, this analysis is much less abstract than it looks. Every algebraic statement can be reduced to a simple picture of how certain sets of points (genotypes) are connected to each other by lines (mutations). The algebra merely expresses an argument that depends on three common

sense "tricks".

1. There is only one positioning that fits three genes into the F-part.

2. Because mutation affects one locus at a time, there is no way genes can move left or right without being destroyed.

   Thus certain two gene genotypes can never mutate into the three gene genotype – these are the dead end genotypes under the naive mapping ϕ.

3. This problem does not appear with the evolvable mapping ϕ_E, because its use of the V-part gives better connectivity between the phenotypes.

The following sections show this rigorously. To do so, this analysis introduces a new tool tailored to these kinds of problems, a mathematical system called *genospace algebra*. This is a propositional calculus based on a graph theory.

Readers wishing to skip all technical detail may jump directly to section 4.3.5. Readers wishing to skip only the technical analysis, but still wanting a gist of the genospace algebra, are advised to read only the next section and then skip forward.

4.2 Genospace algebra

To streamline our argument, we develop an algebraic notation that expresses graph theoretical ideas more compactly. We will then use this to study the pre-images of our mappings, to show that ϕ has a dead end while ϕ_E does not.

If two points a and b are mutational neighbours, we will call them *simply linked* and write $a - b$. We extend this idea to handle linkage between sets as well:

$$a - b \stackrel{\text{def}}{=} \partial(a) \cap b \neq \emptyset$$

$$A - B \stackrel{\text{def}}{=} \exists a \in A \exists b \in B \text{ such that } a - b$$

This effectively restates the idea of Cupal's accessibility measure, as our definitions imply that $A - B \iff \kappa(\alpha \to \beta) \neq 0$ (proof in section 4.6.2). This notion describes when points or sets are linked by single mutations.

But this is not enough. The genotypic dead end is a result of points within a single set differing in their connectivity to an external set. In order to speak about such circumstances,

we need to distinguish the situation where *some* point in a set A has a link to a set B from the situation where *every* point in A has a link to B.

Therefore, if every point in a set A is linked to B, we will say that A is *totally linked* to B, and write $A \to B$.

$$A \to B \stackrel{\text{def}}{=} \forall a \in A, \exists b \in B \text{ such that } a - b$$
$$A \leftrightarrow B \stackrel{\text{def}}{=} A \to B \text{ and } B \to A$$

The last notation is for when two sets are totally linked to each other, which we describe as them being *symmetrically linked*:

These three forms of linkage – simple, total, and symmetric – suffice if we are only interested in single mutations.[1] But we also want to be able to discuss paths of multiple mutations. So we introduce a new kind of association besides linkage: the notion is *connection*.

If a and b are connected by a chain of mutational links within the set Q, we will call a and b *simply connected through* Q and write $a \overset{Q}{\sim} b$. We develop it in strict analogy with the idea of linkage to define *totally connected* and *symmetrically connected*.

$$a \overset{Q}{\sim} b \stackrel{\text{def}}{=} \exists q_1, q_2, \ldots, q_n \in Q \text{ such that } a - q_1 - q_2 \ldots q_n - b$$
$$A \overset{Q}{\sim} B \stackrel{\text{def}}{=} \exists a \in A \exists b \in B \text{ such that } a \overset{Q}{\sim} b$$
$$A \overset{Q}{\rightsquigarrow} B \stackrel{\text{def}}{=} \forall a \in A \exists b \in B \text{ such that } a \overset{Q}{\sim} b$$
$$A \overset{Q}{\leftrightsquigarrow} B \stackrel{\text{def}}{=} A \overset{Q}{\rightsquigarrow} B \text{ and } B \overset{Q}{\rightsquigarrow} A$$

We also define two shorthand forms which will prove convenient, for the frequent case of connectivity through the origin set:

$$A \rightsquigarrow B \stackrel{\text{def}}{=} A \overset{A}{\rightsquigarrow} B$$
$$A \leftrightsquigarrow B \stackrel{\text{def}}{=} A \rightsquigarrow B \text{ and } B \rightsquigarrow A$$

Table 4.2 gathers the key definitions of the relations alongside examples. Table 4.1 summarises how these definitions result from combining the two basic kinds of association

[1] I intend no reference at all to the traditional biological meaning of *genetic linkage*, where particular loci or alleles are inherited jointly.

	linked	connected
simply	–	$\sim$
totally	$\rightarrow$	$\rightsquigarrow$
symmetrically	$\leftrightarrow$	$\leftrightsquigarrow$

Table 4.1: Six forms of connectivity result from combining two kinds of association (direct linkage vs. connection through a path), and three degrees of quantification.

(linkage and connection) with three degrees of quantification (simple, total, and symmetric). With these ideas we can succinctly define a genotypic dead end. We say there is a *dead end genotype* going from α to β when, given $A = \phi^{-1}(\alpha)$ and $B = \phi^{-1}(\beta)$,

$$\exists a, a' \in A \text{ such that } a \overset{A}{\sim} B \text{ but } a' \overset{A}{\not\sim} B \tag{4.3}$$

So if $A \rightsquigarrow B$, then there are no dead ends going from A to B. Without specifying its direction, we can also say that there is a dead end genotype between A and B whenever

$$A - B \text{ but } A \not\leftrightsquigarrow B$$

One more idea proves useful. In many circumstances all points in A can reach B because all points in A can reach each other, allowing them to reach some portal point a_p linked to B. We introduce one last term to express this. If every point in A can reach every other point in A, then we will call A *self-connected* and state this by using the annotation $\overline{A}$:

$$\overline{A} \overset{\text{def}}{=} a_1 \overset{A}{\sim} a_2 \; \forall a_1, a_2 \in A \tag{4.4}$$

From these definitions we can derive the following core lemmas (in which $\square$ is a variable representing any of the connectivity relations):

$$A \rightsquigarrow B \text{ and } B \rightsquigarrow C \implies A \overset{A \cup B}{\rightsquigarrow} C \qquad \text{transitivity} \tag{4.5}$$

$$A \overset{Q}{\rightsquigarrow} B \text{ and } Q \subset R \implies A \overset{R}{\rightsquigarrow} B \qquad \text{passage expansion} \tag{4.6}$$

$$A - B \text{ and } B \subset C \implies A - C \qquad \text{link term expansion} \tag{4.7}$$

$$A - b \text{ and } \overline{A} \implies A \rightsquigarrow b \qquad \text{self-to-total connection} \tag{4.8}$$

$$A \square B \implies (A \times V) \square (B \times V) \qquad \times \text{ distributes over } -, \text{ etc.} \tag{4.9}$$

$$\overline{A} \text{ and } \overline{V} \implies \overline{A \times V} \qquad \overline{A} \text{ closed under } \times \tag{4.10}$$

$$\overline{A}, \overline{B}, A - B \implies \overline{A \cup B} \qquad \overline{A} \text{ closed under linked } \cup \tag{4.11}$$

$$A_1 \overset{Q}{\rightsquigarrow} B \text{ and } A_2 \overset{Q}{\rightsquigarrow} B \iff (A_1 \cup A_2) \overset{Q}{\rightsquigarrow} B \qquad \overset{Q}{\rightsquigarrow} \text{ distributes with } \cup \tag{4.12}$$

Proofs appear in section 4.6.4.

4.3 Analysis of the pre-images

We now consider the phenotype pre-image families under the two mappings ϕ and ϕ_E to show that the naive mapping has dead end genotypes.

Since these mappings differ in how they treat the F-part and V-part of the genotype, we will begin by representing the pre-images of the phenotypes in terms of these parts. V and F are factor spaces of the genotype space $G = V \times F$. Instead of partitioning G into a few large subsets $\{G_0, G_1, \ldots\}$, we will separately partition F into $\{F_0, F_1, F_2, F_3\}$ and V into $\{V_0, V_1, V_2, V_3\}$. We then partition G in terms of the combinations of these partitions of its factors.

We partition F into subsets as follows

$$\begin{aligned} F_i &= \text{points in } F \text{ having } i \text{ genes} \\ &= \phi^{-1}(i)/V \\ &= \{f \mid \phi(v \times f) = i,\ f \in F,\ v \in V\} \end{aligned} \tag{4.13}$$

The pre-image family $\{\phi^{-1}(0), \ldots, \phi^{-1}(3)\}$ of our naive mapping ϕ are now the shaded regions as so:

	V_0	V_1	V_2	V_3	
F_0	$V_0 \times F_0$	$V_1 \times F_0$	$V_2 \times F_0$	$V_3 \times F_0$	$= \phi^{-1}(0)$
F_1	$V_0 \times F_1$	$V_1 \times F_1$	$V_2 \times F_1$	$V_3 \times F_1$	$= \phi^{-1}(1)$
F_2	$V_0 \times F_2$	$V_1 \times F_2$	$V_2 \times F_2$	$V_3 \times F_2$	$= \phi^{-1}(2)$
F_3	$V_0 \times F_3$	$V_1 \times F_3$	$V_2 \times F_3$	$V_3 \times F_3$	$= \phi^{-1}(3)$

Partition V as follows:

$$\begin{aligned} V_i &= \text{points in } V \text{ having } i \text{ 1s} \\ &= \{v \mid \phi_E(v \times f) = i,\ v \in V,\ \phi(f) > 0\} \\ &= \{v \mid \sum v[j] = i\} \end{aligned} \tag{4.14}$$

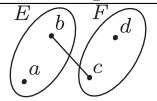
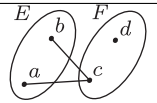
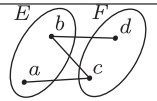
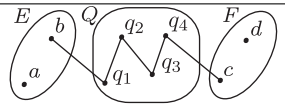
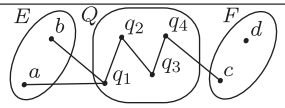
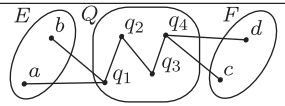
Relation	Definition	Example
simply linked	$x - y \stackrel{\text{def}}{=} x$ can mutate to y and vice versa $X - Y \stackrel{\text{def}}{=} \exists x \in X, y \in Y$ such that $x - y$	$b - c$ $E - F$ 
totally linked	$X \to Y \stackrel{\text{def}}{=} \forall x \in X, \exists y \in Y$ such that $x - y$	$E \to F$ 
symmetrically linked	$X \leftrightarrow Y \stackrel{\text{def}}{=} X \to Y$ and $Y \to X$	$E \leftrightarrow F$ 
simply connected	$x \overset{Q}{\approx} y \stackrel{\text{def}}{=} \exists q_1, \ldots, q_n \in Q$ such that $x - q_1 - \ldots - q_n - y$ $X \overset{Q}{\approx} Y \stackrel{\text{def}}{=} \exists x \in X, y \in Y, x \overset{Q}{\approx} y$	$b \overset{Q}{\approx} c$ $E \overset{Q}{\approx} F$ 
totally connected	$X \overset{Q}{\rightsquigarrow} Y \stackrel{\text{def}}{=} \forall x \in X, \exists y \in Y, x \overset{Q}{\rightsquigarrow} y$ $X \rightsquigarrow Y \stackrel{\text{def}}{=} X \overset{X}{\rightsquigarrow} Y$	$E \overset{Q}{\rightsquigarrow} F$ $Q \rightsquigarrow F$ 
symmetrically connected	$X \overset{Q}{\leftrightsquigarrow} Y \stackrel{\text{def}}{=} X \overset{Q}{\rightsquigarrow} Y$ and $Y \overset{Q}{\rightsquigarrow} X$ $Q \leftrightsquigarrow Y \stackrel{\text{def}}{=} Q \rightsquigarrow Y$ and $Y \rightsquigarrow Q$	$E \overset{Q}{\leftrightsquigarrow} F$ $Q \leftrightsquigarrow F$ 

Table 4.2: Six connectivity relations suffice to describe a genotype space in a way that exposes dead end genotypes and distinguishes direct linkage from connection through a path. Applicable to phenotype pre-images or arbitrary subsets, they can describe the space at any level of details. With the basic lemmas, they can be used to develop proofs of the space's structure.

The pre-image family $\{\phi_E^{-1}(0), \ldots, \phi_E^{-1}(3)\}$ of our evolvable mapping ϕ_E are now the shaded regions as so:

	V_0	V_1	V_2	V_3
F_0	$V_0 \times F_0$	$V_1 \times F_0$	$V_2 \times F_0$	$V_3 \times F_0$
F_1	$V_0 \times F_1$	$V_1 \times F_1$	$V_2 \times F_1$	$V_3 \times F_1$
F_2	$V_0 \times F_2$	$V_1 \times F_2$	$V_2 \times F_2$	$V_3 \times F_2$
F_3	$V_0 \times F_3$	$V_1 \times F_3$	$V_2 \times F_3$	$V_3 \times F_3$
	$\|\|$	$\|\|$	$\|\|$	$\|\|$
	$\phi_E^{-1}(0)$	$\phi_E^{-1}(1)$	$\phi_E^{-1}(2)$	$\phi_E^{-1}(3)$

To compare these two genotype-phenotype mappings, we will compare the internal connections within their respective pre-image families. We make this comparison by first considering the internal connections within the constituent factor space families $\{F_0, , \ldots, F_3\}$ and $\{V_0, \ldots, V_3\}$, and then use those results to understand the pre-image families.

4.3.1 Connectivity of the v-part

We begin by considering the connectivity of the V-part. Its connectivity will be used later in the argument, and it simply illustrates analysis with the genospace algebra.

Before beginning this analysis, however, it is necessary to update our formalisation of mutation. In the previous chapter, ∂A represented Cupal's *mutational boundary operator* $\partial_C A$. This gives the mutants of A, excluding elements of A itself. This misses the effects of mutants within A, and it is harder to handle mathematically. From now on, we will use ∂A to represent the simpler *mutation operator* $\partial_\mu A$, which gives all the mutants of a point or set. In other words,

$$\partial_\mu A \stackrel{\text{def}}{=} \bigcup_{a \in A} \partial_\mu a = \bigcup_{a \in A} \partial_\mu \{\text{mutants of } a\} \tag{4.15}$$

As the two operators produce the same results on single points, and differ only when applied to sets, this distinction only matters now that we are operating on sets.[2]

[2]To clarify: we always use ∂_μ for the *mutation operator* (which returns all possible mutants of a genotype or set of genotypes), and ∂_C for Cupal's *mutational boundary operator* (which returns all possible mutants of a set, minus the set itself). In chapter 3, which was dedicated to Cupal's ideas, we use ∂ as a shorthand for his operator ∂_C. In this chapter and in the remainder of this thesis, we use ∂ as a shorthand for operator ∂_μ. Changing the shorthand seems less confusing than proliferating subscripts.

Consider V_1, the set of points in V having only one non-zero locus. Consider

$$\begin{aligned}
\partial(V_1) &= \partial(\{\boxed{1}\boxed{0}\boxed{0}, \boxed{0}\boxed{1}\boxed{0}, \boxed{0}\boxed{0}\boxed{1}\}) \\
&= \partial(\boxed{1}\boxed{0}\boxed{0}) \cup \partial(\boxed{0}\boxed{1}\boxed{0}) \cup \partial(\boxed{0}\boxed{0}\boxed{1}) \\
&= \left\{ \begin{array}{l} \boxed{\mathbf{0}}\boxed{0}\boxed{0}, \\ \boxed{1}\boxed{\mathbf{1}}\boxed{0}, \\ \boxed{1}\boxed{0}\boxed{\mathbf{1}} \end{array} \right\} \bigcup \left\{ \begin{array}{l} \boxed{\mathbf{1}}\boxed{1}\boxed{0}, \\ \boxed{0}\boxed{\mathbf{0}}\boxed{0}, \\ \boxed{0}\boxed{1}\boxed{\mathbf{1}} \end{array} \right\} \bigcup \left\{ \begin{array}{l} \boxed{\mathbf{1}}\boxed{0}\boxed{0}, \\ \boxed{0}\boxed{\mathbf{1}}\boxed{1}, \\ \boxed{0}\boxed{0}\boxed{\mathbf{0}} \end{array} \right\} \\
&= V_0 \cup V_2 \qquad (4.16)
\end{aligned}$$

The last term implies $\partial(V_1) \cap V_0 \neq \emptyset$ and $\partial(V_1) \cap V_2 \neq \emptyset$. This can be restated as $V_1 - V_0$ and $V_1 - V_2$ or, in one statement, $V_0 - V_1 - V_2$.

Now let us exploit a symmetry to avoid unnecessarily repeating this argument. Swapping the 1's and 0's turns V_0 into V_3, V_2 into V_1, and V_1 into V_2. Recognising this symmetry lets us see that $V_0 - V_1 - V_2$ implies $V_3 - V_2 - V_1$. Putting it together we get

$$V_0 - V_1 - V_2 - V_3$$

This expresses the simple mutational linkage among the V_i. Simple linkage depends on the same notion of overlapping mutational boundaries used in the accessibility measure, so this also establishes that these regions have non-zero accessibility in the language of Cupal.

But accessibility requires only a mutational overlap for *some* point in A. In order to show there are no dead ends, we must see if *every* point v in a given V_i has the same connectivity with some other V_j.This connectivity may depend on every point $v \in V_i$ having a direct mutational link $v - V_j$, or more generally a multi-step connection $v \overset{V_i}{\sim} V_j$. In other words, we want to know if it is also the case that

$$V_0 \leftrightsquigarrow V_1 \leftrightsquigarrow V_2 \leftrightsquigarrow V_3$$

In fact it is so. V_0 contains only the one point $\boxed{0}\boxed{0}\boxed{0}$, any mutation adds a locus, so it is trivial that $v - V_1 \ \forall v \in V_0$, which we write as $V_0 \rightarrow V_1$. The swapping symmetry then implies that $V_3 \rightarrow V_2$. Examining the three terms in equation (4.16), we can also see that every point in V_1 has some intersection with V_0 and V_2. So $V_0 \leftarrow V_1 \rightarrow V_2$. The

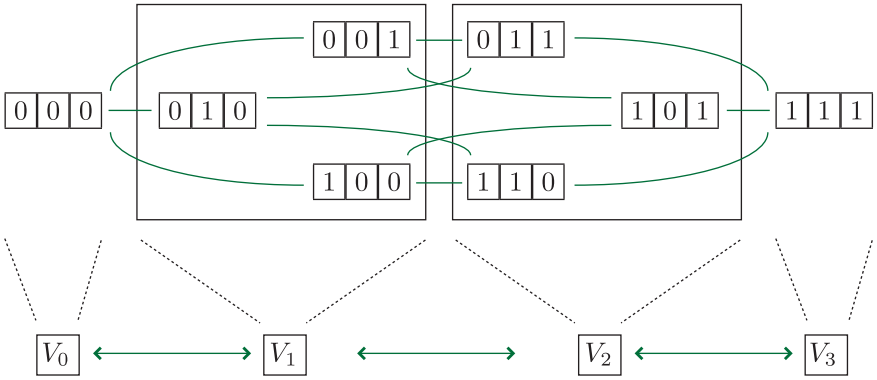


Figure 4.1: Top: a high-resolution view of the simple linkages between all points in V, the set of length 3 binary genotypes, laid out to to distinguish the partition subsets V_0, V_1, V_2, and V_3. Bottom: a low-resolution view of the resulting relations between those subsets – i.e., symmetric connections between certain pairs.

symmetry implies $V_3 \leftarrow V_2 \rightarrow V_1$. Putting it all together, we have a complete analysis of the connectivity of the V-part

$$\left(\begin{array}{lccccccc} & V_0 & \rightarrow & V_1 & & & & \\ \text{and} & & & & & V_2 & \leftarrow & V_3 \\ \text{and} & V_0 & \leftarrow & V_1 & \rightarrow & V_2 & & \\ \text{and} & & & V_1 & \leftarrow & V_2 & \rightarrow & V_3 \end{array}\right)$$

$$\Longrightarrow V_0 \leftrightarrow V_1 \leftrightarrow V_2 \leftrightarrow V_3 \quad (4.17)$$

This states symmetrical linkage. It implies symmetrical connection, $V_0 \leftrightsquigarrow V_1 \leftrightsquigarrow V_2 \leftrightsquigarrow V_3$, a fortiori.

What is the meaning of this contrast? If V_2 and V_3 are symmetrically connected, this means that any v in a V_2 can find a path leading to V_3 without passing through any other sets (and vice versa). But the fact that V_2 and V_3 are also symmetrically linked means that this path need not involve internal movement within V_2. The connection is due to direct linkage. Every point in V_2 can move into V_2 with a single mutation. This is a close form of connection, unlike the situation we will find among the F_i.

The above analysis demonstrates an algebraic approach, using definitions and inference rules to derive true statements. But at bottom, all these statements are about simple visual structures, networks between points and sets of points. The set V is small enough that we can draw that network in full detail, as in figure 4.1. In that diagram, we can see the fact

of total linkage by noting that in any set, every point has direct links to the "neighbouring" sets.

The diagram also makes it obvious that V is a self-connected set: plainly, there is no subset of V which has no links to the rest of V. We can as easily prove that result algebraically:

$$\begin{aligned} & V_0 \leftrightsquigarrow V_1 \leftrightsquigarrow V_2 \leftrightsquigarrow V_3 & \\ \Longrightarrow\ & V_i \overset{V}{\leftrightsquigarrow} V_j \ \forall i,j & \text{by transitivity} \\ \Longrightarrow\ & v \overset{V}{\leftrightsquigarrow} v' \ \forall v, v' \in V & \text{since } \cup V_i = V \\ \Longrightarrow\ & \overline{V} & \end{aligned} \tag{4.18}$$

This completes the analysis of V.

Overall connectivity

Our approach to V provides the template for analysing F, especially as regards what constitutes a final result.

We noted that $V_0 \leftrightarrow V_1 \leftrightarrow V_2 \leftrightarrow V_3$ implies that $V_0 \leftrightsquigarrow V_1 \leftrightsquigarrow V_2 \leftrightsquigarrow V_3$. For that matter it also implies $V_0 - V_1 - V_2 - V_3$. In fact, for the purposes of studying the symmetric associations relevant to dead end genotypes, it implies all true connectivity statements about the V_i. This is why it concludes our analysis of connectivity – it is the inferentially strongest possible true statement about the V_i, so strong that it implies every other true statement.

When this is true of a connectivity statement, then we say that it is the *overall connectivity*. When we know the overall connectivity between all pairs of sets under study, then we have completed our analysis. How do we know when we've found the overall connectivity? The inference rules relating the connectivity relations follow from their definitions, which are discussed in detail in section 4.6.4. Those inference rules imply an ordering of relations, where every relation implies the relations later in the order. This allows an easy guideline for determining the overall connectivity. The overall connectivity between two sets is the first connectivity relation that holds in the following sequence: $\leftrightarrow, \leftrightsquigarrow, \rightarrow, \leftarrow, \rightsquigarrow, \leftsquigarrow, -, \not-$.

4.3.2 Connectivity of the f-part

Now we consider the connectivity of the F_i, a more complicated case. By definition a subset F_i depends only on the presence of the "genes" within the F-part of the genotype. A mutation will take a point from F_i to F_j only if it creates or destroys a gene. But many mutations will do neither. They are neutral mutations, and allow for "internal movement" within an F_i.

F with zero genes We start with F_0. Consider the point $f_0 \in F_0$

$$f_0 = \boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}$$

and its mutants

$$\partial(f_0) = \{\boxed{\mathbf{1}}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0},$$
$$\boxed{0}\boxed{\mathbf{1}}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0},$$
$$\vdots$$
$$\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{\mathbf{1}}\boxed{0}\}$$

Since a gene requires at least three loci set to 1, we can see that no single mutation can create a gene from f_0. Or, $\partial(f_0) \subset F_0$. Furthermore, we can see that any point $f \in F_0$ can mutate to f_0 through a sequence of mutations that replace its 1's with 0's one by one[3]. In other words,

[3]To prove this rigorously is slightly involved. It does not suffice to mutate the 1's to 0's in any order. We must ensure there exists an order of mutations – a mutational path – which does not create a valid gene en route to f_0 by turning a 1 locus into a start stop codon for a valid gene. This occurs for instance, if a mutation strikes position 1 in $\boxed{0}\boxed{\mathbf{1}}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}$ or position 4 in $\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{\mathbf{1}}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}$. Now we prove that a safe path always exists.

Consider a genotype f. Within f, we will call a contiguous sequence of 1's, bounded by 0's, a *run*. More generally, we will call a contiguous sequence of 1's, bounded by 0's or the genotype's endpoints, a *superrun*. The *length* of a run or superrun is the number of 1's in it. Let f be in F_0, so it has no runs of length 3, i.e., no genes. We must show that a sequence of point mutations can take f to f_0, without ever generating a run of length 3.

Summarise our sequence f by listing its superruns and their lengths. Since superruns are strictly bounded from each other, mutating loci within a superrun never changes the length of neighbouring superruns. Therefore, to take f to f_0, it suffices to eliminate each of its superruns separately, setting them to 0. When they have all been replaced with 0's, then we have taken f to f_0. We eliminate every superrun through the following procedure.

If a superrun has a length $l < 3$, then it can be set to zero by l point mutations on its 1's in any order. If a superrun has a length 3, then it must be either at the beginning or at the end point of the genotype, or else it would be a run of length 3. It can be set to zero by three point mutations on its 1's in any order. If a superrun has a length of 4 then we can apply mutations in the following order, which will not create a run of length 3:

$$\boxed{1}\boxed{1}\boxed{1}\boxed{1} - \boxed{1}\boxed{1}\boxed{\mathbf{0}}\boxed{1} - \boxed{1}\boxed{1}\boxed{0}\boxed{\mathbf{0}} - \boxed{1}\boxed{\mathbf{0}}\boxed{0}\boxed{0} - \boxed{\mathbf{0}}\boxed{0}\boxed{0}\boxed{0}$$

If a superrun has a length $4 < l < 13$, then we apply the first mutation to the last locus, reducing it it to a superrun of length $l - 1$, continuing inductively until we have $l = 4$, which we treat as above.

This proves that $f \overset{F_0}{\rightsquigarrow} f_0$ for all $f \in F_0$, where $f_0 = \boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}$.

$F_0 \rightsquigarrow f_0$ or, more verbosely, $x \overset{F_0}{\sim} f_0 \; \forall x \in F_0$. But $\sim$ is transitive: $x \overset{F_0}{\sim} f_0$ and $f_0 \overset{F_0}{\sim} y$ together mean that $x \overset{F_0}{\sim} y \; \forall x, y \in F_0$. In other words, F_0 is self-connected. Or simply

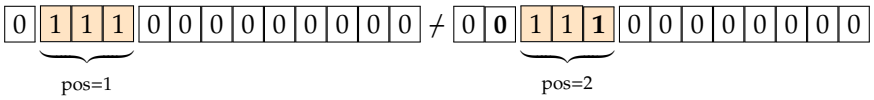
$$\overline{F_0}$$

By being connected to every other point in F_0, f_0 acts as a kind of switching junction tying the entire set together.

F with one gene Now we turn to F_1. What is its connectivity to F_0? Any genotype in F_1 can drop into F_0 by suffering a mutation that destroys its gene, so F_1 is totally linked to F_0: $F_1 \rightarrow F_0$. If the sets are linked and F_0 is self-connected, then through equation (4.8) this means F_0 is totally connected to F_1: $F_0 \rightsquigarrow F_1$. Combining these two results, we have established F_0's overall connectivity with F_1:

$$F_0 \leftrightsquigarrow F_1$$

Is F_1 self-connected like F_0? No. It is partitioned into mutationally disconnected subsets. This is because, although a gene can occupy any position in the F-part of the genome, our definition of mutation does not let genes "move" left or right. Point mutation can only create a gene by adding a missing a locus or destroy a join by changing a correct locus. But for a gene to move left or right would require simultaneous changes, and our mutations occur only one at a time:



Given that the gene cannot move on the genome through mutations within the set F_1, we partition F_1 based on the position of the gene. I.e., $F_1 = \cup F_1^i$, where

$$F_1^i \overset{\text{def}}{=} \{f \,|\, \text{gene at position } i, \text{ and } f \in F_1\}$$

We know that $F_1^i \not\rightsquigarrow F_1^j$, where $i \neq j$. In any F_i, the the noncoding part of the genome behaves like the genome for F_0, which is entirely noncoding. That is, it is self-connected, so we also know that $\overline{F_1^i}$. As with V_1, figure 4.2 summarises how the internal structure of F_1

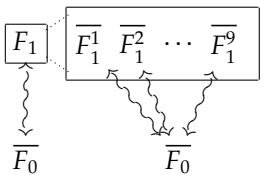


Figure 4.2: Connectivity of F_0 and F_1. Left: F_0 is totally connected to F_1 because F_0 is self-connected. Right: F_1 is not self-connected. It is partitioned into disconnected subsets based on the gene's position. But it still is totally connected to F_0 because each of those subsets is.

determines its connectivity with F_0.

F with two genes Now we turn to F_2. As before, once a first gene is defined, then a second gene can always be created by three correct mutations. Also, if we have two genes, there is always the possibility that one will be destroyed through an unlucky mutation. So we have the overall connectivity

$$F_1 \leftrightsquigarrow F_2$$

Considering the internal structure of F_2, we see it is similar to F_1, except it requires two indices to specify the locations of its two genes. Let

$$F_2^{i,j} = \{f \mid \text{genes at position } i \text{ and } j,\ f \in F_2\}$$

Then $F_2^{i,j} \not\leftrightsquigarrow F_2^{k,l}$ where $i \neq k$ and $j \neq l$, and each subset is self-connected $\overline{F_2^{i,j}}$.

F with three genes Now we turn to F_3, the simplest case because it is a singleton set. Only three genes can fit into the F-part:

$$F_3 = \{\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\}$$

Up until F_3, every pair of F_i's have been symmetrically consider. This will change as we consider F_2 and F_3, and this anomaly show us how how to identify the dead end genotype under the ϕ mapping.

Consider that there is only one arrangement which fits three genes onto a genome. Recall from equation (4.3) that to identify a genotypic dead end, we must identify two phenotypes, an origin α and a destination β. Then we must show that $A = \phi^{-1}(\alpha)$ includes a point a and dead end point a' such that $a \overset{A}{\sim} B$ while $a' \overset{A}{\not\sim} B$. Also recall that F_2, the set of the F-parts coding for two genes, was partitioned into noncommunicating subsets *based on the positions of the two genes*. This leads us to understand where to look for a dead end points. Look for

places where the bar on gene movement prevents a two gene genotype from evolving into a three gene genotype. For instance, it is not the point $f \in F_2^{1,9} \subset F_2$

$$f = \{\,0\,\underbrace{1\,1\,1}_{\text{pos=1}}\,0\,0\,0\,0\,0\,\underbrace{1\,1\,1}_{\text{pos=9}}\,0\,\}$$

since that can mutate into F_3 (that is, that $f \overset{F_2}{\rightsquigarrow} F_3$) through mutations that create the third gene in the centre. But consider a dead end point $f_D \in F_2^{2,9} \subset F_2$, with its first gene positioned differently, such as

$$f_D = \{\,0\,0\,\underbrace{1\,1\,1}_{\text{pos=2}}\,0\,0\,0\,0\,\underbrace{1\,1\,1}_{\text{pos=9}}\,0\,\}$$

Here the transition is impossible (that is, $f_D \overset{F_2}{\not\rightsquigarrow} F_3$). Its existing gene at position 2 cannot move left, so it cannot get out of the way. Generally what we see is that

$$F_2^{i,j} \rightsquigarrow F_3 \text{ iff } i, j \in \{1, 5, 9\} \text{ and } i \neq j$$

meaning the overall connectivity is non-symmetric:

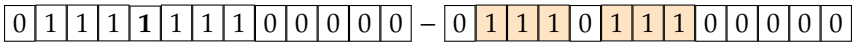
$$F_2 \leftarrow F_3$$

Mutations on stop/start codons This lack of symmetric connection is a telltale sign of a dead end genotype, which is what we aim to show about the naive mapping. But let us take a small detour to complete our analysis of the F-part. We considered if each F_i was self connected, and we analysed the connectivity for the pairs $(F_0, F_1), (F_1, F_2),$ and (F_2, F_3), pairs which correspond to direct gene creation or destruction. But for a complete analysis, we must also calculate the connectivity for the other pairings: $(F_0, F_2), (F_0, F_3),$ and (F_1, F_3). We can see that

$$F_0 \not\leftarrow F_3$$

since no mutation of the one genotype in F_3 will destroy all three of its genes at once. Surprisingly, there are mutations which allow the other two pairings, each of which corresponds to a single mutation creating or destroying two genes simultaneously. This is

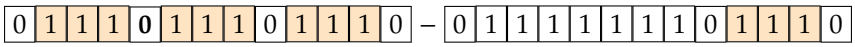
possible when a mutation strikes start/stop codons of neighbouring genes. The link



establishes that $F_0 - F_2$. Since $\overline{F_0}$, this implies the overall connectivity

$$F_0 \rightsquigarrow F_2 \tag{4.19}$$

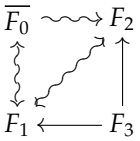
The link



establishes that $F_3 - F_1$. Since F_3 is a singleton, this implies the overall connectivity

$$F_3 \to F_1$$

As both of these relations are nonsymmetric, they only establish more exotic kinds of dead end genotypes within F. But they also let us now present a complete analysis of the connectivity of the F-part:



(4.20)

We indicated that f_D constitutes a kind of dead end point within the F-part of the genome. However, F is only a factor space of our original genotype space $G = V \times F$, and we need to look for dead end genotype within G under the two mappings. We need to take what we have learned about connectivity of the V-part and F-part, combine it with our definitions of the mappings ϕ and ϕ_E, and use it to understand the connectivity of their pre-image families. We treat ϕ first.

4.3.3 Connectivity of ϕ's pre-image family

Consider that using our definition of F_i in equation (4.13), we can express ϕ^{-1} as follows:

$$\phi^{-1}(i) = V \times F_i$$

Since ϕ^{-1} ignores the V-part of the genome entirely, and exactly mirrors the structure of F_i, we can generalise our result from F directly.

Recall result (4.18) that V is self-connected. Using the distributivity equation (4.9) which states that relations are preserved across products, and equation (4.10) which states that the product of self-connected sets is self-connected, we directly calculate the equation system for the $\phi^{-1}(i)$ simply by multiplying $\overline{V}\times$ from the left against all the relations in the equation system (4.20) of the F-part.

$$\overline{V} \times \left(\begin{array}{ccc} \overline{F_0} & \rightsquigarrow & F_2 \\ \updownarrow & \nearrow & \uparrow \\ F_1 & \longleftarrow & F_3 \end{array} \right) = \begin{array}{ccc} \overline{\phi^{-1}(0)} & \rightsquigarrow & \phi^{-1}(2) \\ \updownarrow & \nearrow & \uparrow \\ \phi^{-1}(1) & \longleftarrow & \phi^{-1}(3) \end{array} \tag{4.21}$$

This yields the complete analysis of ϕ's pre-image family.[4] Such a diagrammatic representation of equations summarising all the pre-image connectivities of a genotype-phenotype mapping, we will call a *genospace diagram*.

Having calculated the $\phi^{-1}(i)$ from the F_i, we can also calculate a valid dead end genotype by generating it from f_D in the obvious way. We use an arbitrary $v \in V$:

$$\begin{aligned} g_D &= v \times f_D \\ &= \boxed{0}\boxed{0}\boxed{0} \times \boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0} \\ &= \boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0} \end{aligned} \tag{4.22}$$

To restate: this is a dead end because $g_D \in \phi^{-1}(2)$ but $g_D \overset{\phi^{-1}(2)}{\nsim} F_3$ despite that $g \overset{\phi^{-1}(2)}{\sim} F_3$ for certain other $g \in \phi^{-1}(2)$ such as any $g \in V \times F_2^{1,9}$. Or verbally: this genotype is a dead end

[4] There is one subtlety: $\phi^{-1}(3)$ is self-connected where F_3 was not. This is because F_3 was a singleton, and $\overline{A} \times b \implies \overline{A \times b}$.

because it produces a two segment phenotype but, unlike other similar genotypes, it can never find a mutational path leading to a three segment phenotype.

We have chosen a dead end genotype that is doomed because its leftmost gene is one position away from the correct position 1. But there are many dead end genotypes, corresponding to the many incorrect positionings and the many possible values of the V part. Similarly we have chosen a non-dead end genotype with correct genes in the left and right positions 1 and 9. But there are many non-dead end genotypes, corresponding to the other pairs of correct positions ((1,5) and (1,9)), the neutral mutations within the empty space, and the many possible values of the V part.

In this respect it is interesting to note that there are many more dead end genotypes than non-dead end genotypes, because there are more ways for two genes to be incorrectly positioned than correctly positioned. That is,

$$\left| \bigcup_{(i,j)\notin I} F_2^{i,j} \right| > \left| \bigcup_{(i,j)\in I} F_2^{i,j} \right|$$

$$\text{where } I = \{(1,5),(5,9),(1,9)\}$$

But this fact is missed by the accessibility measure $\kappa(2 \to 3)$, which, because the measure ignores connectivity structure, will misleadingly include all the dead genotypes in assessing the chances of a transition.

4.3.4 Connectivity of ϕ_E's pre-image family

We now consider the evolvable mapping ϕ_E. We will show there are no dead ends. Consider the pre-images expressed in terms of F_i and V_i:

$$\phi_E^{-1}(i) = \begin{cases} E_0 \cup (V \times F_0) & \text{if } i = 0 \\ E_i & \text{if } i = 1,2,3 \end{cases} \tag{4.23}$$

where

$$E_i = (V_i \times (F_1 \cup F_2 \cup F_3)) \tag{4.24}$$

We will establish that connectivities of the pre-images of ϕ_E^{-1} are fully described by the following two relations:

$$\phi_E^{-1}(1) \leftrightarrow \phi_E^{-1}(2) \leftrightarrow \phi_E^{-1}(3) \tag{4.25}$$

$$\overline{\phi_E^{-1}(0)} \nleftrightarrow \phi_E^{-1}(i) \text{ where } i \in 1,2,3 \tag{4.26}$$

Start with

$V_0 \leftrightarrow V_1 \leftrightarrow V_2 \leftrightarrow V_3$	by (4.17)
$(V_0 \leftrightarrow V_1 \leftrightarrow V_2 \leftrightarrow V_3) \times (F_1 \cup F_2 \cup F_3)$	
$E_0 \leftrightarrow E_1 \leftrightarrow E_2 \leftrightarrow E_3$	by (4.24) and (4.9)
$E_0 \leftrightarrow \phi_E^{-1}(1) \leftrightarrow \phi_E^{-1}(2) \leftrightarrow \phi_E^{-1}(3)$	by (4.23)

This establishes equation (4.25). It remains to establish equation (4.26).

First we show the leftward implication, that $\phi_E^{-1}(0) \nleftarrow \phi_E^{-1}(i)$ for $i = 1,2,3$. Consider that equation (4.20), which summarises our findings about the F-part, implies that

$$F_3 \rightsquigarrow F_2 \rightsquigarrow F_1 \rightsquigarrow F_0$$

From there, we work as follows:

$F_3 \overset{F_3}{\rightsquigarrow} F_2 \overset{F_2}{\rightsquigarrow} F_1 \overset{F_1}{\rightsquigarrow} F_0$	by def. of $\rightsquigarrow$
$F_3 \overset{F_1 \cup F_2 \cup F_3}{\rightsquigarrow} F_2 \overset{F_1 \cup F_2 \cup F_3}{\rightsquigarrow} F_1 \overset{F_1 \cup F_2 \cup F_3}{\rightsquigarrow} F_0$	by (4.6), passage expansion
$F_i \overset{F_1 \cup F_2 \cup F_3}{\rightsquigarrow} F_0 \ \forall i \in \{1,2,3\}$	by (4.5), transitivity
$(F_1 \cup F_2 \cup F_3) \overset{F_1 \cup F_2 \cup F_3}{\rightsquigarrow} F_0$	by (4.12), $\cup$ distribution
$(F_1 \cup F_2 \cup F_3) \rightsquigarrow F_0$	by def. of $\rightsquigarrow$
$V_i \times ((F_1 \cup F_2 \cup F_3) \rightsquigarrow F_0)$	by (4.9), $\times$ distribution
$E_i \rightsquigarrow V_i \times F_0$	by def. (4.24)
$E_i \rightsquigarrow \phi_E^{-1}(0)$	by (4.7), since $V_i \times F_0 \subset \phi_E^{-1}(0)$
$\phi_E^{-1}(i) \rightsquigarrow \phi_E^{-1}(0)$ for $i = 1,2,3$	by def. (4.23)

That proves the leftward implication.

Now we show the rightward implication, that $\phi_E^{-1}(0) \rightsquigarrow \phi_E^{-1}(i)$. The leftward implication implies that $\phi_E^{-1}(0) - \phi_E^{-1}(i)$. Given this, using equation (4.8), to prove the rightward implication it suffices to show that $\phi_E^{-1}(0)$ is self-connected. We show this now. First we express $\phi_E^{-1}(0)$ as a union of two terms:

$$\phi_E^{-1}(0) = (V_0 \times F) \cup (V \times F_0)$$

Given equation (4.11), which states that the union of linked self-connected sets is itself self-connected, we need only show that each term is self-connected and that the terms are simply linked. We show each term is self connected by applying equation (4.9) to our past results:

$$\begin{aligned} \overline{V} \text{ and } \overline{F_0} &\implies \overline{V \times F_0} \\ V_0 \text{ is a singleton and } \overline{F} &\implies \overline{V_0 \times F} \end{aligned}$$

We note that, since the terms have a non-zero intersection,

$$(V_0 \times F) \cap (V \times F_0) = V_0 \times F_0$$

they could only fail to be linked as a pathological case. This is not such a case. Consider the null genotype point g_0 and any of its neighbours g_0':

$$\begin{aligned} g_0 - g_0' \text{ and } g_0, g_0' \in V_0 \times F_0 & \\ \implies V_0 \times F_0 - V_0 \times F_0 & \\ \implies V_0 \times F - V \times F_0 & \qquad \text{by (4.7)} \end{aligned}$$

This shows the terms are linked, which shows $\phi_E^{-1}(0)$ is self-connected, which shows the rightward implication $\phi_E^{-1}(0) \rightsquigarrow \phi_E^{-1}(i)$. This establishes equation (4.26), the other relation to be shown besides equation (4.25), thus completing our analysis of the pre-image family of the evolvable mapping. We can put bring both these equations together in the following

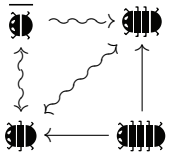
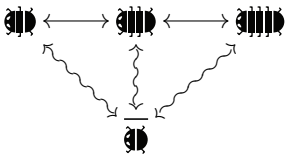
naive	evolvable
	
Every nonsymmetric association indicates dead end genotypes. I.e., two genotypes may have the same phenotype but different access to other phenotypes.	All associations are symmetric since there are no dead end genotypes, suggesting greater evolvability.
	Straight arrows indicate direct linkages between certain phenotypes, enabled by mutations in the regulatory region, suggesting greater evolvability.
Mutations creating or destroying genes yield predictable associations between phenotypes differing by one body segment. Start/stop mutations create counterintuitive associations between phenotypes differing by two body segments.	Destructive mutations in the genecoding region mean every segmented phenotype is connected to the zero segment phenotype.

Table 4.3: Genospace diagrams for naive vs evolvable segmentation

genospace diagram:

$$\phi_E^{-1}(1) \longleftrightarrow \phi_E^{-1}(2) \longleftrightarrow \phi_E^{-1}(3) \qquad (4.27)$$

$$\overline{\phi_E^{-1}(0)}$$

4.3.5 Results of connectivity analysis

Recall that the pre-image of a phenotype is just the set of all genotypes which yield that phenotype. The preceding sections calculated the mutational relationships among all the phenotypic pre-images, both for the naive mapping and for the evolvable mapping. These calculations were done with the genospace algebra, a new formalism which provides an exact way to analyse genotype-phenotype mappings. The analysis can be presented as a genospace diagram, a kind of X-ray picture depicting the logical structure of the map. Structural features in the diagram can indicate evolvability and represent our intuitive

understanding of the map.

The calculation above yielded equation systems (4.21) and (4.27) respectively. Table 4.3 reproduces those equation systems as genospace diagrams, substituting the phenotype cartoons themselves for their pre-images, and summarises the key differences.

Compare the genospace diagrams of the two maps. The first point is that our analysis discovered numerous genotypic dead ends in the naive mapping. In a genospace diagram, every nonsymmetric association indicates such a dead end. In the naive mapping, the total linkage  is easiest to understand: as described, if the two genes are positioned so there is no space for a third gene, then there is no mutational path leading to three segments. In contrast, every genotype with three genes can suffer a mutation leading to two genes. The other nonsymmetric associations are less obvious, resulting from mutations striking stop/start codons, but they all indicate dead end genotypes. In contrast, the evolvable mapping contains many fewer dead end genotypes – in fact, it contains none at all, as is visible in the absence of nonsymmetric associations. This difference is as we would expect. If evolvability is a quality that makes phenotypes transitions, then dead end genotypes are a contra-indicator of that.

But another way to make phenotypic transitions easier is to reduce the number of mutations required for them. We also see that contrast in the diagrams. The evolvable mapping features double-headed straight-backed arrows between all the segmented phenotypes. Those arrows indicate symmetric linkage – that for every genotype, a single mutation suffices to take it to the neighbouring phenotype. In contrast, the wavy arrows of the naive mapping indicate that sometimes a mutational path of multiple steps is required.

The structure of these diagrams is also revealing. The tidy symmetry of the evolvable mapping's diagram is a direct reflection of its simple logic in interpreting the regulatory region (the V-part) of the genome, as opposed to the more involved logic of gene detection in the F-part. These diagrams depict the connectivity structure at the pre-image level. If we want to visualise the connectivity structure at finer levels of resolution – that is, by considering sets of genotypes smaller than the phenotype pre-images – then that structure is fully described in the analysis, and parts of it are also depicted in figures 4.1 and 4.2.

4.4 Generality of this segmentation model

Ungeneralisable assumptions produce only ungeneralisable results. How much do these results depend on assumptions particular to this model?

As regards biological accuracy, it goes without saying that this is a heuristic model, not a representation of current science on mechanisms governing segmentation. As regards general criticisms of the genospace algebra as a tool for understanding genospace structure, I consider these in section 4.6.3. Here I will consider how much the results depend on the specifics of this model, on the exact definitions of the mappings ϕ and ϕ_E.

One criticism is that the F-part having a length of 13 is perfectly tuned to guarantee only one three gene genotype configuration, which will tend to restrict access from the two gene genotypes in F_2. This is true, but it only clarifies why some F_2 points are blocked from reaching F_3; it is not responsible or necessary for blocking them. The fundamental reason is that gene movement is impossible within a pre-image. This is clear when we consider that the analysis demonstrated other dead end genotypes that did not involve singleton sets, for instance the genotypes in F_2 unable to reach F_0 directly, as described in the analysis leading up to equation (4.19).

Another criticism is that the greater connectivity of the evolvable mapping ϕ_E is due to the fact that only it depends on the V-part of the genome, which is completely ignored by ϕ. It is as if ϕ_E maps from a genome of length 16, while ϕ maps from a genome of length 13. Could the absence of dead end genotypes be due to this difference in the size of the effective genospace domains of the mappings, rather than their structure? In fact, this is not so. If we redefined a new naive mapping ϕ', which counted the instance of the genes $\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}$ over the entire length 16, it still can hold a maximum of three genes, giving it the same phenotypic range as ϕ. Instead of being a singleton, the pre-image $\phi'^{-1}(3)$ will be partitioned into subsets based on the positions of those genes, but the bar on gene movement still leads to two gene genotypes that cannot mutate into three gene genotypes. For instance, the two gene genotype $\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{0}\boxed{0}$ cannot accommodate a third valid gene at either end, or in the centre.

A third criticism is that the evolvable mapping from the V-part has a couple unusual properties, which one might suspect to be the true reason for the result showing no dead end genotypes. For instance, once the F-part has at least one gene, the mapping allows

no neutral mutations in V. Also, the mapping ranges over all of P. It is true that ϕ_E is unusual in this way, but the results do not depend on these special properties. They key point for showing the absence of dead end genotypes is that all the V_i are *symmetrically* associated with each other. It is irrelevant whether that association is symmetrical linkage or symmetrical connection. In this mapping, the symmetry is due to the way the V-part escapes the gene movement problem. It is true that there are no neutral mutations in V, and that is why the V_i are all linked rather than connected, but that distinction is not crucial.

However, while these special properties are not crucial for disallowing dead end genotypes, they do let us intuit why such a mapping allows easier transitions between phenotypes. They produce a convenient transparency, which we cannot hope to find in the messiness of experimental systems. Instead of hoping that when evolvability "makes transitions easier" we can always see with our eyes how it does so, we must formalise the notion further to make it exact. Dead end genotypes formalise one kind of obstacle to easy transition. A simple simple improvement on that idea would be to count the number of dead end genotypes between different pre-images, rather than just proving their existence or non-existence. But we should also seek a formalisation that picks up on the other special features that make the evolvability of ϕ_E mapping so obvious.

The next chapter will do this, introducing another analytic measure, the genetic leverage λ. This is unrelated to dead end genotypes. This measure is sensitive to these other features of ϕ_E. It is also more easily applied to untidy and realistic systems. In summary, the first two criticism turn out to be without foundation, whereas the third is responding to something genuinely noteworthy, which will be developed further.

4.5 Discussion of significance

This chapter has compared two genotype-phenotype mappings in order to model a feature commonly proposed to be evolvogenic – the capacity for repetition of body segments. The "evolvable" mapping demonstrates abstracted control of repetition, while the naive mapping requires literal repetition of the genes themselves.

The explicitness of this model gives these longstanding ideas a precise form susceptible to exact analysis. This chapter also introduces the genospace algebra – a formalism based on a graph theoretical treatment of genotype space and designed for this kind of analysis.

While this formalism draws on ideas of the accessibility measure reviewed in chapter 3, it also goes beyond them in its ability to represent dead end genotypes. Dead end genotypes appear only in the naive mapping, a finding consistent with the expectation that they tend to reduce the evolvability.

What is the significance of these results? Obviously, this work does not model the details of current science on segmental development.[5] Instead, it presents two models of segmentation which are stripped down to show just the core logic of the argument about why a certain kind of developmental system is more evolvable. So although this work cannot say much about real segmentation on its own, it can say something about the logical mechanisms that could govern real segmentation.

Within that ambit, this chapter's analysis proves that a certain kind of segmentation-friendly mapping eliminates dead end genotypes. It also proves that in some cases this mapping reduces the number of mutations required to evolve from one phenotype to another. Furthermore, the details of these proofs suggest it is crucial that a genotypic factor subspace (in this case, the V-part) have a kind of magnified importance in the mapping.

But it is not yet clear how this work can be generalised to other models in an exact fashion. What, exactly, is "abstracted control"? Handed a genotype-phenotype mapping at random, how do we detect it? If the claim is that this model of segmentation exemplifies a broader pattern, what is that pattern? Such questions are essentially about what the model means, about how much it implies. Is it just representing a particular claim about segmentation, or is it illustrating a more general dynamic?

The coming chapters will develop another method of assessing evolvability, and a more elaborate model of development. These, it is hoped, will describe the larger idea of which this segmentation model is the simplest possible example.

4.6 Appendix

4.6.1 Example genotypes

Table 4.4 compares the action of the two genotype-phenotype maps on various genotypes.

[5] However, this is not to say it contradicts current knowledge of segmental development. It is true that in *Drosophila melanogaster* different segments are represented by different gene complexes. But in *Strigamia maritima*, the number of segments can vary even within populations, depending partly on the nutritional environment, strongly indicates a deep flexibility in the developmental architecture supporting segment repetition. (Chipman and Akam, 2008)

γ	$\phi(\gamma)$	$\phi_E(\gamma)$
0 0 0 0 1 1 1 0 1 1 1 0 1 1 1 0	3	0
0 0 0 0 1 1 1 0 1 0 1 1 1 0 1 0	2	0
0 0 0 0 0 1 1 0 1 0 1 0 1 0 1 0	0	0
0 1 0 0 0 1 1 0 1 0 1 0 1 0 1 0	0	0
0 1 0 0 1 1 1 0 1 0 1 0 1 0 1 0	1	1
0 1 1 0 1 1 1 0 1 0 1 0 1 0 1 0	1	2
0 1 1 | 0 1 1 1 0 1 0 1 0 1 1 1 0	2	2
V | F		

Table 4.4: Examples of ϕ and ϕ_E

4.6.2 Pre-image linkage is equivalent to nonzero accessibility

Here we show that Cupal's notion of two different phenotypes having non-zero accessibility (introduced in chapter 3) is equivalent to our notion of their two pre-image sets being mutationally linked.

This is essential to our claim that accessibility measures fail to see dead end genotypes, since our demonstration of dead end genotypes depends on showing that certain pre-images pairs are simply connected as opposed to totally connected, and connection is determined by linkage. If accessibility and linkage were not equivalent within the analysis, then our demonstration would show nothing about accessibility.

Formally we wish to show that, for all $\alpha \neq \beta$

$$\phi^{-1}(\alpha) - \phi^{-1}(\beta) \iff \kappa(\alpha \to \beta) > 0 \tag{4.28}$$

Here we take $A \stackrel{\text{def}}{=} \phi^{-1}(\alpha)$ and $B \stackrel{\text{def}}{=} \phi^{-1}(\beta)$ as usual. Recall that $A - B$ if and only if $\partial A \cap B \neq \emptyset$. Recall also that $\kappa(\alpha \to \beta) \stackrel{\text{def}}{=} |\partial_C A \cap B| \,/\, |\partial_C A|$, which implies that $\kappa(\alpha \to \beta) > 0$ if and only if $\partial_C A \cap B \neq \emptyset$. Therefore it suffices for us to show that

$$\partial_C A \cap B \neq \emptyset \iff \partial A \cap B \neq \emptyset \tag{4.29}$$

As is evident, the only subtlety here is that linkage is built from the mutation operator ∂, while accessibility is built from the mutational boundary operator ∂_C. We now show this subtlety presents no obstacles.

Recall that the mutational operator ∂A gives all mutants of A, while the mutational boundary operator $\partial_C A$ gives all mutants of A, minus members of A itself (definitions in equations (3.2) and (4.15)). From this, we may state the relationship between the two operators:

$$\begin{aligned}
\partial A &= \bigcup_{a \in A} \partial a = \bigcup_{a \in A} \{\text{mutants of } a\} \\
\partial_C A &= \{a \,|\, a \in \partial A \text{ and } a \notin A\} \\
\partial A &= A_1 \cup \partial_C A \\
&\quad \text{where } A_1 = \{a \,|\, a \in A \text{ and } a \in \partial A\}
\end{aligned}$$

We see

$$\begin{aligned}
\partial_C A &\subseteq \partial A \\
\partial_C A \cap B &\subseteq \partial A \cap B && \text{by eq. (4.39)} \\
\partial_C A \cap B \neq \emptyset &\implies \partial A \cap B \neq \emptyset && (4.30)
\end{aligned}$$

This establishes the rightward implication of (4.29). Now assume that

$$A \cap B = \emptyset \tag{4.31}$$

We see

$$\begin{aligned}
\partial A \cap B &\neq \emptyset \\
(A_1 \cup \partial_C A) \cap B &\neq \emptyset \\
(A_1 \cap B) \cup (\partial_C A \cap B) &\neq \emptyset && \text{by eq. (4.36)} \\
\emptyset \cup (\partial_C A \cap B) &\neq \emptyset && \text{by assumption (4.31)} \\
\partial_C A \cap B &\neq \emptyset \\
\partial A \cap B \neq \emptyset &\implies \partial_C A \cap B \neq \emptyset && (4.32)
\end{aligned}$$

This establishes the leftward implication of (4.29). So now all we must show is that assumption (4.31) is reasonable.

Since $\phi : G \to \mathcal{P}(P)$ is a function, it is single-valued for a given argument. Thus the family of its pre-images, $\{A \,|\, A = \phi^{-1}(\alpha)\ \forall \alpha \in P\}$, defines a partition on G. I.e., $\phi^{-1}(\alpha) \cap \phi^{-1}(\beta) = \emptyset \iff \alpha \neq \beta$. This means that assumption (4.31) is valid, proving our result.

4.6.3 Limits of the genospace algebra

The genospace algebra is a shorthand for describing the structure of an undirected graph. It does this through statements of the relationships between sets and points within that graph. This is a powerful tool for describing the structure of genotype space, because we can treat the set of all genotypes as a graph where every genotype defines a node in the graph, every mutation defines a link between nodes, and phenotype pre-images define some of the relevant subsets for analysis. Altogether this represents a vocabulary for writing proofs about genospace structures, for diagramming those structures, and even for automated searching of those structures by software systems. Those are its virtues. Here we will consider the main limits of this kind of analysis, in order of ascending significance.

First, it is not due to a limit of the genospace algebra, as such, that we have confined ourselves to imaginary binary genetic systems where loci take only a value of 0 or 1. This is a choice to simplify our models, and we could just as easily use the algebra with the biological nucleotide coding of A,T,C, and G. Furthermore, none of the lemmas we develop in section 4.6.4 depend on assuming a binary genome.

Second, it is similarly not due to a limit of the genospace algebra that we have assumed point mutation. The mutation structure of the space is defined by the mutation operator ∂, and all of the connectivity relations follow from that definition. However, point mutation is a very convenient assumption, and it is crucial for the derivation of the rule for the mutation of products in section 4.6.4. One could calculate a product rule for a mutation operator which allows not only point mutations, but also insertions/deletions, but it would probably not be as analytically tractable. One could still perform computational studies of the connectivity structure of such space, but it might be harder to glean insight from such an investigation.

Third, it is a limitation of the genospace algebra as presented here that it requires that the mutation operator be symmetric – that is, that $a \in \partial b \iff b \in \partial a$, that it is always possible

to mutate to the point you have mutated from. It is this intrinsic symmetry of the mutation operator which guarantees that the simple link relation is also intrinsically symmetric, i.e., $a - b \iff b - a$. (This intrinsic symmetry stems directly from the mutation operator. It is just the standard mathematical notion of the "symmetry" of a binary relation, such as when it describes one of the three properties of an equivalence relation. It should not be confused with our defined terms symmetrical linkage or symmetrical connection.) While it would be possible to develop the algebra starting with a mutation operator that is not symmetric, it would undermine many of the lemmas here and require constantly tracking the ordering of any given relation. This would be so different in style that it might as well qualify as a different algebra.

Last, the most fundamental limitation of the algebra, however, is that it is categorical not ordinate. There are six connectivity statements and each of those statements is either true or false, giving little sense of the strength of connectivity. The statement $a \overset{Q}{\approx} b$ makes no distinction between a and b being connected by a single long and tenuous path through Q, as opposed to an innumerable multitude of short paths. The statement $A \rightsquigarrow B$ tells us that there are dead end genotypes in B, but not whether they constitute a couple points or the great majority. Finally, by starting with the mutation operator ∂, which simply gives the set of possible mutants ∂a from a point a, the algebra loses all distinction of which mutants are probable. Some mutations could be very common, others quite rare.

This is a significant criticism. One could take account of these deficiencies and try to produce an ordinate, high fidelity account of genospace structure. This might start with the refined probabilistic accessibility measure ω which was defined in equation (3.7). Instead of the binary notion of linked or not linked, which led to the unweighted, undirected graph of the genospace algebra, ω could lead to a weighted, directed graph representation of genotype space. This more accurate representation could open the door to more ambitious types of analysis, for instance, to measuring the probability of certain evolutionary trajectories (prior to considerations of selection) by considering probabilities summed over all possible paths of development through this graph. This problem is like, and may reduce to, the canonical algorithmic problem of calculating the flow through a graph network.

However, while it is clear how these refinements would make this analysis more realistic in its details, and while the power to calculate the probabilities of evolutionary trajectories beckons as a kind of seductive holy grail, this thesis will not take any steps in that direction.

The complexity of the analysis would mean that every small increment in the accuracy of the analysis would be purchased at the price of a great increase in its obscurity, complication, and labour. As the basic ideas of evolvability remain so inchoate, I believe there is more value in bringing maximum clarity to the essential logic, rather than maximum accuracy to every peripheral detail.

4.6.4 Mathematics of the genospace algebra

This section offers a mathematical treatment of the genospace algebra's definitions, lemmas, relationship to graph theory, and referenced axioms of set theory.

I call it the genospace algebra because I created it to analyse structure in genospace. However, it is just a propositional calculus based on a graph theory, so it could be used for analysing purely mathematical problems or any subject that could be expressed in graph theoretical terms – e.g., networks of communication, transport, or flow, or else networks of interaction strength between people, between metabolites, etc. (The only important lemmas which depends on specifically biological assumptions are the rule for mutation of products, and the consequent implied rule for how Cartesian products distribute over connectivity relations. These follow from the definition of mutation operator ∂.)

For a discussion of its particular limits as a tool for genospace analysis, see section 4.6.3.

Definitions

We define three forms of *linkage* relations as follows:

$$
\begin{aligned}
a - b &\stackrel{\text{def}}{=} \partial(a) \cap b \neq \emptyset \\
A - B &\stackrel{\text{def}}{=} \exists a \in A, \exists b \in B \text{ such that } a - b && \text{simply} \\
A \to B &\stackrel{\text{def}}{=} \forall a \in A, \exists b \in B \text{ such that } a - b && \text{totally} \\
A \leftrightarrow B &\stackrel{\text{def}}{=} A \to B \text{ and } B \to A && \text{symmetrically}
\end{aligned}
$$

We define three forms of *connection* relations analogously, also including a shorthand for connection through the origin set:

$$
a \overset{Q}{\sim} b \stackrel{\text{def}}{=} \exists q_1, \ldots, q_n \in Q \text{ such that } a - q_1 \ldots q_n - b
$$

$$A \overset{Q}{\approx} B \overset{\text{def}}{=} \exists a \in A, \exists b \in B \text{ such that } a \overset{Q}{\approx} b \qquad \text{simply}$$

$$A \overset{Q}{\rightsquigarrow} B \overset{\text{def}}{=} \forall a \in A, \exists b \in B \text{ such that } a \overset{Q}{\approx} b \qquad \text{totally}$$

$$A \overset{Q}{\leftrightsquigarrow} B \overset{\text{def}}{=} A \overset{Q}{\rightsquigarrow} B \text{ and } B \overset{Q}{\rightsquigarrow} A \qquad \text{symmetrically}$$

$$A \rightsquigarrow B \overset{\text{def}}{=} A \overset{A}{\rightsquigarrow} B$$

$$A \leftrightsquigarrow B \overset{\text{def}}{=} A \rightsquigarrow B \text{ and } B \rightsquigarrow A$$

But to what do these relations apply? The total and symmetric relations are defined between pairs of sets. But the simple relations are defined between both pairs of sets and pairs of points. Borrowing from computer science, one could say that the simple relations are defined *polymorphically*, that they are overloaded on the type of their arguments (Pierce, 2002).

For completeness, one would like to overload the other relations in the same way and to define all the relations on combinations of sets and points. Such a promiscuous mixing of argument types might only sow confusion, were it not for a deeper consistency. Note that any simple association between two points ($a \square b$) applies if and only if the same association applies between two corresponding singleton sets ($\{a\} \square \{b\}$). This guides our mode of extension. We say, for all Q,

$$\forall \square \in \{-, \rightarrow, \leftrightarrow, \overset{Q}{\sim}, \overset{Q}{\rightsquigarrow}, \overset{Q}{\leftrightsquigarrow}, \rightsquigarrow, \leftrightsquigarrow\},$$

$$a \square b \overset{\text{def}}{=} \{a\} \square \{b\}$$

$$A \square b \overset{\text{def}}{=} A \square \{b\}$$

$$a \square B \overset{\text{def}}{=} \{a\} \square B$$

This defines all the relations between any combination of points and sets[6].

Basic lemmas

Since connection consists of one or more linkages, every form of linkage implies its corresponding form of connection. Linkage is stronger than connection. In other words,

[6]To be exact, in the case of the simple associations $-$ and $\overset{Q}{\sim}$, we began by independently *defining* their meaning between points and between sets, and we then deduced from those definitions the implication $a \square b \iff \{a\} \square \{b\}$ for $\square \in \{-, \overset{Q}{\sim}\}$. For the other associations $\{\rightarrow, \leftrightarrow, \overset{Q}{\rightsquigarrow}, \overset{Q}{\leftrightsquigarrow}, \rightsquigarrow, \leftrightsquigarrow\}$, we defined their meaning between points by *stipulating* that the same implication must hold as between sets. It matters not whether a result is defined or deduced, so long as the overall pattern of implications remains unchanged.

$A - B \implies A \sim B, A \to B \implies A \overset{Q}{\rightsquigarrow} B$, and $A \leftrightarrow B \implies A \leftrightsquigarrow B$. Linkage and connection are both kinds of of association. *Symmetrical* association is a shorthand for two statements of total association. *Total* association is a shorthand for a statement of simple association for every element in the source set. Therefore, $A \leftrightarrow B \implies A \to B \implies A - B$ and $A \leftrightsquigarrow B \implies A \rightsquigarrow B \implies A \sim B$. The main immediate implications between the relations are summarised below:

$$\begin{array}{ccccc} A - B & \Longrightarrow & A \overset{Q}{\gtrsim} B \ \forall Q & & \\ \Uparrow & & \Uparrow & & \\ A \to B & \Longrightarrow & A \overset{Q}{\rightsquigarrow} B \ \forall Q & \Longrightarrow & A \rightsquigarrow B \\ \Uparrow & & \Uparrow & & \Uparrow \\ A \leftrightarrow B & \Longrightarrow & A \overset{Q}{\leftrightsquigarrow} B \ \forall Q & \Longrightarrow & A \leftrightsquigarrow B \end{array} \tag{4.33}$$

Other implications follow from the definitions. We present them here. The notation makes these statements brief, exact, provable, and perhaps at first a little cryptic. But these rules all have an intuitive meaning, which is obvious with a little reflection.

Connections are transitive:

$$\begin{aligned} & A \rightsquigarrow B \text{ and } B \rightsquigarrow C \\ & \implies \begin{pmatrix} \forall a \in A, \exists b \in B, \exists a_1, \ldots, a_n \in A \text{ such that } a - a_1 - \ldots - a_n - b \\ \text{and} \\ \forall b \in B, \exists c \in C, \exists b_1, \ldots, b_n \in B \text{ such that } b - b_1 - \ldots - b_n - c \end{pmatrix} \\ & \implies \forall a \in A, a - a_1 - \ldots - a_n - b - b_1 - \ldots - b_n - c \\ & \implies A \overset{A \cup B}{\rightsquigarrow} C \end{aligned}$$

The "passage set" can be freely expanded to a superset:

$$\begin{aligned} & A \overset{Q}{\rightsquigarrow} B \text{ and } Q \subset R \\ & \implies \forall a \in A, \exists b \in B, \exists q_1, \ldots, q_n \in Q \text{ such that } a - q_1 - \ldots - q_n - b \\ & \implies \forall a \in A, \exists b \in B, \exists q_1, \ldots, q_n \in R \text{ such that } a - q_1 - \ldots - q_n - b \\ & \implies A \overset{R}{\rightsquigarrow} B \end{aligned}$$

Terms in a simple linkage can be freely expanded to supersets:

$$
\begin{aligned}
&A - B \text{ and } B \subset C \\
&\qquad \implies \exists a \in A, \exists b \in B \text{ such that } a - b \\
&\qquad \implies \exists a \in A, \exists b \in C \text{ such that } a - b \\
&\qquad\qquad \implies A - C
\end{aligned}
$$

A simply linked set, which is self-connected, is also totally connected:

$$
\begin{aligned}
&\overline{A} \text{ and } A - b \\
&\qquad \implies (\forall a, a' \in A, a \overset{A}{\rightsquigarrow} a') \text{ and } (\exists a_p \in A \text{ such that } a_p - b) \\
&\qquad\qquad \implies \forall a \in A, a \overset{A}{\rightsquigarrow} a_p - b \\
&\qquad\qquad\qquad \implies A \rightsquigarrow b
\end{aligned}
$$

The source in total associations distributes over set unions:

$$
\begin{aligned}
&A_1 \overset{Q}{\rightsquigarrow} B \text{ and } A_2 \overset{Q}{\rightsquigarrow} B \\
&\qquad \iff \begin{pmatrix} a_1 \overset{Q}{\sim} B \ \forall a_1 \in A_1 \\ a_2 \overset{Q}{\sim} B \ \forall a_2 \in A_2 \end{pmatrix} \\
&\qquad \iff a \overset{Q}{\sim} B \ \forall a \in (A_1 \cup A_2) \\
&\qquad\qquad \iff (A_1 \cup A_2) \overset{Q}{\rightsquigarrow} B
\end{aligned}
$$

Similarly,

$$A_1 \to B \text{ and } A_2 \to B \iff (A_1 \cup A_2) \to B$$

Mutation of products

How does the mutation operator act on products of sets? For a point mutation operator acting on Cartesian products, the rule is as follows:

$$\partial(A \times B) = (A \times \partial B) \cup (\partial A \times B) \tag{4.34}$$

We show this below. Consider the point mutation operator $\partial^m : S_m \to \mathcal{P}(S_m)$, where $S_m = \{0,1\}^m$ is the set of bit string genotypes of length m. Formally, the operator is as follows

$$\partial^m(g) = \{s \mid \exists j \in \{1,\ldots,m\}\ \forall i \in \{1,\ldots,m\}\ s[i] = g[i] \iff i \neq j\}$$

Consider a length-m genotype g:

$$\begin{aligned} g &= \underbrace{\boxed{\text{g[0]}}\ldots\boxed{\text{g[k-1]}}}\underbrace{\boxed{\text{g[k]}}\ldots\boxed{\text{g[m-1]}}} \\ &= \quad g[0:k] \quad \times \quad g[k:m] \end{aligned}$$

We call the first k loci its *front*, the last last $m-k$ loci its *rear*. The front may be seen as its own genotype $g[0:k] \in S_k$, as can the rear $g[k:m] \in S_{m-k}$. Now consider the mutants of g. From the definition we know there is only one mutated locus. This locus is located either in g's front or rear. The set of possible mutants of each part is defined by the mutation operator for the factor space of that part's length. So we can build the set of all mutants as a union of these possibilities:

$$\begin{aligned} \partial^m(g_m) &= \{\text{mutants mutated in rear}\} \cup \{\text{mutants mutated in front}\} \\ &= \Big(g[0:k] \times \partial^{m-k}(g[k:m])\Big) \cup \Big(\partial^k(\,g[0:k]\,) \times g[k:m]\Big) \end{aligned} \tag{4.35}$$

Now we verify that this result generalises as we expect to all subsets $A \subset S_k, B \subset S_{m-k}$ where $G = S_m = S_l \times S_{m-k}$. We use the fact that union $\cup$ commutes with ∂ and $\times$. Here we omit the annotation on ∂. Every ∂ should be assumed to be the operator appropriate for its arguments:

$$\begin{aligned} \partial(A \times B) &= \cup_{a\in A, b\in B}\partial(a \times b) && \text{by def. of } \partial \\ &= \cup_{a\in A}\cup_{b\in B}\left((a \times \partial b) \cup (\partial a \times b)\right) && \text{by eq. (4.35)} \\ &= (\cup_{a\in A}\cup_{b\in B}\, a \times \partial b) \cup (\cup_{a\in A}\cup_{b\in B}\,\partial a \times b) \\ &= ((\cup_{a\in A} a) \times (\cup_{b\in B}\partial b)) \cup ((\cup_{a\in}\partial a) \times (\cup_{b\in B} b)) \\ &= (A \times \partial(\cup_{b\in B} b)) \cup ((\partial \cup_{a\in} a) \times B) \\ &= (A \times \partial B) \cup (\partial A \times B) \end{aligned}$$

This establishes our rule for applying the point mutation operator to Cartesian product spaces of binary genome.

Multiplication distributes over connectivity relations

Here we show that Cartesian multiplication by a space distributes over connectivity relations between sets from another space. This is important, since it lets us analyse the connectivity of parts of a genome, and then generalise that to the genome as a whole. We rely on equation (4.34), the rule for the mutation of a product space.

We start by considering the simple linkage relation:

$$
\begin{aligned}
& A - B & & \\
& \partial A \cap B & \neq \emptyset & \text{by def. of } - \\
& (\partial A \cap B) \times V & \neq \emptyset & \text{by eq. (4.40)} \\
& (\partial A \times V) \cap (B \times V) & \neq \emptyset & \text{by eq. (4.38)} \\
& ((A \times \partial V) \cap (B \times V)) \cup ((\partial A \times V) \cap (B \times V)) & \neq \emptyset & \text{by eq. (4.41)} \\
& ((A \times \partial V) \cup (\partial A \times V)) \cap (B \times V) & \neq \emptyset & \text{by eq. (4.36)} \\
& \partial (A \times V) \cap (B \times V) & \neq \emptyset & \text{by eq. (4.34)} \\
& A \times V - B \times V & & \text{by def. of } -
\end{aligned}
$$

All the other connectivity relations are built from the simple linkage relation, and inherit its property of allowing distribution of the Cartesian product. In other words, we find that

$$
\begin{aligned}
& \forall \square \in \{-, \rightarrow, \leftrightarrow, \rightsquigarrow, \leftrightsquigarrow\}, \\
& \qquad A \square B \implies (A \times V) \square (B \times V) \\
& \forall \square \in \{\sim, \rightsquigarrow, \leftrightsquigarrow\}, \\
& \qquad A \overset{Q}{\square} B \implies (A \times V) \overset{Q \times V}{\square} (B \times V)
\end{aligned}
$$

We have shown this for $A - B$. Now we show it for the other relations. We start with $A \overset{Q}{\sim} B$:

$$
\begin{aligned}
& A \overset{Q}{\sim} B \\
& \implies \exists a \in A, b \in B \text{ such that } a \overset{Q}{\sim} b
\end{aligned}
$$

$$\implies \exists a \in A, b \in B \text{ such that } a - q_1 \dots q_n - b$$
$$\implies \exists a \in A, b \in B \forall v \in V \text{ such that } a \times v - q_1 \times v \dots q_n \times v - b \times v$$
$$\implies \exists a \in A, b \in B \forall v \in V \text{ such that } a \times v \overset{Q \times v}{\sim} b \times v$$
$$\implies \exists a \in A, b \in B \forall v \in V \text{ such that } a \times v \overset{Q \times V}{\sim} b \times v$$
$$\implies \exists a \in A, b \in B \exists v \in V \text{ such that } a \times v \overset{Q \times V}{\sim} b \times v$$
$$\implies \exists a \in A, b \in B \text{ such that } a \times V \overset{Q \times V}{\sim} b \times V$$
$$\implies \exists y \in (A \times V), z \in (B \times V) \text{ such that } y \overset{Q \times V}{\sim} z$$
$$\implies (A \times V) \overset{Q \times V}{\sim} (B \times V)$$

Now we show it for $\overset{Q}{\rightsquigarrow}$:

$$\begin{aligned} & A \overset{Q}{\rightsquigarrow} B \\ & \quad \implies \forall a \in A, a \overset{Q}{\sim} B \\ & \quad \implies \forall a \in A, a \times V \overset{Q \times V}{\sim} B \times V && \text{by the result above} \\ & \quad \implies A \times V \overset{Q}{\rightsquigarrow} B \times V \end{aligned}$$

We have shown that $\times$ distributes over $-, \overset{Q}{\sim}$, and $\overset{Q}{\rightsquigarrow}$. Setting $Q = \emptyset$ implies it distributes over $\rightarrow$. Setting Q to the source set implies it distributes over $\rightsquigarrow$. Distribution over $\leftrightarrow, \nleftrightarrow$, and $\overset{Q}{\nleftrightarrow}$ follows directly from distribution over $\rightarrow, \rightsquigarrow$, and $\overset{Q}{\rightsquigarrow}$. This completes our proof that $\times$ distributes over all the connectivity relations.

We have shown that the product distributes over the connectivity relations. Since self-connection is also built from linkage, it is preserved as well:

$$\begin{aligned} & \overline{A} \text{ and } \overline{V} \\ & \qquad \implies (\forall a, a' \in A \; a \overset{A}{\rightsquigarrow} a') \text{ and } (\forall v, v' \in V \; v \overset{V}{\rightsquigarrow} v') \\ & \qquad \implies \exists a_1, \dots, a_n \in A, \exists v_1, \dots, v_n \in V \\ & \qquad \text{such that } a - a_1 \dots a_n - a' \text{ and } v - v_1 \dots v_n - v' \\ & \implies a \times v - a_1 \times v \dots a_n \times v - a' \times v - a' \times v_1 \dots a' \times v_n - a' \times v' \\ & \qquad\qquad \implies \overline{A \times V} \end{aligned}$$

(A related degenerate case is that the product of any self-connected set with a singleton set is self-connected: $\overline{A} \implies \overline{A \times b}$.) If two sets are both self-connected, and linked to each other, then their union is self-connected:

$$\overline{A} \text{ and } \overline{B} \text{ and } A - B$$

$$\implies \left(\begin{array}{l} \forall a \in A,\ a \overset{A}{\sim} a' \\ \text{and } \forall b \in B,\ b \overset{A}{\sim} b' \\ \text{and } \exists a' \in A, b' \in B \text{ such that } a' - b' \end{array} \right)$$

$$\implies \forall a \in A, \forall b \in B\ a \overset{A}{\sim} a' - b' \overset{B}{\sim} b$$

$$\implies \overline{A \cup B}$$

Genospace algebra as graph theory

One can regard the genospace algebra as merely a notation for graph theory. An undirected graph consists of a set of vertexes V and a set of edges E where every edge connects two vertexes. The genospace algebra starts with the idea of the set of genotypes G and a mutation operator ∂. The genotypes are vertexes. The mutation operator defines the simplest relation $a - b$, the idea of *simple linkage* between two points. This defines a notion of an edge between two vertexes.

That relation is the basis for all the other connectivity relations, which are in fact statements representing different combinations of choices between existential or universal quantification ($\exists$ versus $\forall$), and between direct linkage or connection through an intermediate set. For instance, the idea of simple linkage between two sets $A - B$ results from considering linkage $a - b$ and quantifying on existence within A and B. The idea of *total linkage*, $A \to B$ is a result of universal quantification on A. *Symmetric linkage* is total linkage in two directions. The logic is analogous for the case of connection through a set.

Since this scheme is built from a graph theoretical foundation, it is no surprise that it recapitulates standard graph theoretical ideas. The idea of a set in genotype space being self-connected amounts to the statement in graph theoretical terminology that it is a "connected subgraph". The idea of two points being connected through the space as a whole, amounts to the statement that two points are "path connected". The statement $A \overset{Q}{\rightsquigarrow} b$ is the statement that "every point $a \in A$ is path connected to b within the subgraph defined by $a \cup Q \cup b$".

The difference between these descriptions is convenience. Graph theory takes points as the main actors. But there are too many genotype points for the mind to hold them all in close focus. The genospace algebra represents the relationships not only between points, but between points, sets, subsets, and their combinations. This is what makes it possible to summarise a system at different levels of resolution, as in figure 4.1, and succinctly to prove statements about large spaces.

Referenced equalities

The following elementary results are referenced in the preceding arguments.

Set theory axioms We rely on these basic axioms of set theory and distributivity properties:

$$(A \cup B) \cap V = (A \cap V) \cup (B \cap V) \tag{4.36}$$

$$(A \cup B) \times V = (A \times V) \cup (B \times V) \tag{4.37}$$

$$(A \cap B) \times V = (A \times V) \cap (B \times V) \tag{4.38}$$

$$A \subset B \implies A \cap V \subset B \cap V \tag{4.39}$$

$$A \neq \emptyset \implies A \times V \neq \emptyset \tag{4.40}$$

$$A \neq \emptyset \implies A \cup V \neq \emptyset \tag{4.41}$$

Cartesian product The Cartesian product is defined as

$$A \times B \stackrel{\text{def}}{=} \{(a, b) \,|\, a \in A, b \in B\}$$

and with

$$a \times b \stackrel{\text{def}}{=} (a, b) \tag{4.42}$$

$$A \times b \stackrel{\text{def}}{=} \{(a, b) \,|\, a \in A\} \tag{4.43}$$

These are natural because they give

$$A \times B = \cup_{a \in A, b \in B} a \times b$$

Chapter 5

Genetic leverage

Chapter 4 introduced the *genospace algebra*, a formalism for analysing the mutational relationships between phenotypes by representing them as network structures in genotype space. These structures can be depicted in a *genospace diagram*, a picture showing which phenotypes can mutate into each other and giving some information about how easily they can do so. This picture lets you compare the modes of evolvability in different systems.

However, there are two drawbacks to this approach. First, it provides qualitative as opposed to quantitative information, as the diagrams reveal structure but do not summarise the effect of that structure. Second, it can require laborious analysis to capture so much detail. At the most basic level, evolvability comes down to how many mutations it takes to evolve from one phenotype to another. Focusing only on this basic issue yields a simpler, less nuanced, but more convenient measure of evolvability. This measure is λ, the genetic leverage. It is defined partly in terms of an associated construct, the genetic fulcrum f, which helps to identify the causes of evolvability.

This chapter introduces and defines these two constructs, which are summarised in figure 5.2. Together they provide an approximate but exactly defined method for measuring the evolvability of a genotype-phenotype mapping and for identifying which aspects of a mapping contribute to its evolvability.

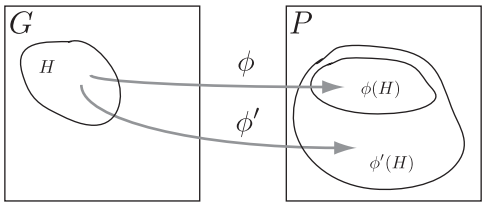


Figure 5.1: Over the genotype subset H, the range of ϕ' is greater than the range of ϕ. Mutations within this subset will tend to have more effect under ϕ'. The size of a range divided by its underlying genotype set gives the genetic leverage λ, a crude measure of evolvability.

5.1 Definition of genetic leverage

Genetic leverage ignores the detailed structure of phenotypic relations and provides a single aggregate measure. It is the ratio of the size of two sets – a set of genotypes H, and the set of their associated phenotypes. This phenotype subset, known mathematically as the *range of ϕ over H*, is written $\phi(H) \stackrel{\text{def}}{=} \{\phi(g) \,|\, g \in H\}$. The genetic leverage is then as follows:

$$\lambda_\phi(H) \stackrel{\text{def}}{=} \frac{\left|\phi(H)\right|}{|H|}$$

It is crucial to choose a subset H so that this ratio measures something biologically meaningful. This is a subtle matter and to do this I will in the next section introduce the concept of the *genetic fulcrum*, which is a way of defining the subset H. But before doing that, I would like to provide two examples to illustrate the basic idea behind defining leverage as this kind of ratio.

5.1.1 Motivation

The basic idea is that leverage measures (very roughly) the evolutionary challenge posed by redundancy in the genotype-phenotype mapping. This can be seen by examining two familiar cases.

First, consider the leverage over a special subset, the entire genotype space G. Every phenotype is produced by some genotype, so the range of ϕ over G is the entire phenotype space P. The leverage is then the ratio of their sizes:

$$\lambda = \frac{|P|}{|G|} = \frac{\text{number of phenotypes}}{\text{number of genotypes}} \ll 1$$

This measures the redundancy of the mapping ϕ over the whole space. The tininess of this ratio reflects one of the familiar, fundamental difficulties for evolutionary search – the difficulty of finding new phenotypes when the search space of genotypes is so much larger.

The smaller the ratio, the more difficult the search.

Next consider the leverage over another special subset, the mutational neighbourhood of an individual g'. In this case the range of ϕ over the mutational neighbourhood $\partial(g')$ is the set of phenotypes one mutation away from g'. The leverage is as follows:

$$\lambda = \frac{\left|\{\phi(g) \mid g \in \partial(g')\}\right|}{\left|\partial(g')\right|}$$

For example, we might apply this to the naive segmentation mapping, which simply counts genes. Consider the all-zero genotype. There are sixteen mutations available to it, one for each position. None of them produces a valid gene, so they all yield the same zero-segment phenotype – that is, $\partial g_{000} = \{$  $\}$. This gives us

$$\begin{aligned} g_{000} &= \boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0} \\ \lambda &= \left|\phi(\partial(g_{000}))\right| / \left|\partial(g_{000})\right| \\ &= \left|\{\text{}\}\right| / \left|\partial(g_{000})\right| \\ &= 1/16 \end{aligned}$$

Compare this leverage to that of a genotype that has one complete gene and is one mutation away from a complete second gene. It can suffer mutations which have no effects on its existing gene (at positions 1 to 3 or 13 to 16), mutations which add a second gene and thus a second segment (at position 10), or mutations which destroy its existing gene and segment (at positions 4 to 8). This creates a phenotypic mutational neighbourhood of three phenotypes, yielding a leverage as follows:

$$\begin{aligned} g_{100} &= \boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{0}\boxed{1}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0} \\ \lambda &= \left|\phi(\partial(g_{100}))\right| / \left|\partial(g_{100})\right| \\ &= \left|\{\text{  }, \text{  }, \text{}\}\right| / \left|\partial(g_{100})\right| \\ &= 3/16 \end{aligned}$$

The second genotype has more linkages to new phenotypes. And this is reflected in the higher genetic leverage, which provides a crude measure of the difficulty of phenotype change around g_{100}. In other words, $\lambda(\partial(g_{100})) > \lambda(\partial(g_{000}))$ because g_{100} is more evolvable

than g_{000}.

5.1.2 Genetic leverage and the developmental architecture

As discussed in section 2.3.2, some past work has defined evolvability using variants of the above calculation. Such an approach treats evolvability as a property of an individual, defined by the individual's mutational neighbourhood (Turney, 1999; Burch and Chao, 2000; Smith et al., 2002; Dichtel et al., 2004). In the terminology of chapter 2 one would say that g_{100} has higher individual mutational evolvability.

However, as discussed in that section, despite the convenience of this approach for simple experiments, there is a serious conceptual problem with treating evolvability as an individual property. The evolvability we seek to understand in the biological world is not a property of individuals but of their developmental architectures. For instance, Yang (2001) does not study the evolvability of a single beetle, but compares the evolvability of hemi- and holometabolous insect taxa, trying to understand if their two different architectures lead to different patterns of diversification.

So what we need is a measure of evolvability that measures – that applies to – a developmental architecture. But if the measure is formal, then this will require a new formal construct that describes the developmental architecture itself. In this case, what good is genetic leverage, a mathematical construct that applies to sets of genotype? This brings us back to the original question of how to define the special subset H, the subset which will be used to calculate the genetic leverage of a mapping.

The solution I propose is that, for the purposes of formalisation, we can specify a developmental architecture using a set of genotypes. Instead of describing a particular developmental architecture in some new formal language, we simply point to the particular set of individual genotypes individuals that possess that developmental architecture. We measure the leverage of the developmental architecture by measuring the leverage of that set of genotypes.

For example, consider again Yang's study comparing the evolvability of hemi- and holometabolous insect taxa. These two taxa possess different developmental architectures. These two taxa are also represented by different sets of individuals, the individuals which share those architectures. How would we measure the leverages of these different developmental architectures? The hard route would be to find some formal construct that directly

describes those architectures, and then define an evolvability measure that applied to those constructs. The easier route is just to calculate the leverage of the two sets of individuals, since those individuals already implicitly represent those architectures. In short: while we could define a developmental architecture by formalising a description of it, we can also define it by simply enumerating all the individuals and only the individuals that possess it.

Enumerating those individuals sidesteps a number of hard questions that would come up if we tried to describe the architecture directly with a new formal construct. These are questions such as: How do you describe the physical, mechanistic aspect of an architecture? (Would that be in some generalised formal language of physical mechanism – a holy grail which it is easier to imagine than to find?) How do you relate the architecture's physical aspect to its genetic underpinning? How do you give your description of an architecture clear boundaries when it is shared by different individuals and it makes up only a part of them physically?

Enumerating the individuals which possess a particular developmental architecture sidesteps these questions because they are all answered, at least implicitly, by the existing constructs of the genotype-phenotype mapping. The physical aspect of an architecture is implied by the definition of P, which already describes possible phenotypes. The genetic underpinnings – the genetic architecture – is already implied by ϕ, which encapsulates all facts about the dependence of the physical mechanism on the genotype. The boundaries of the developmental architecture are defined by our choice of which individuals possess the architecture. Instead of drawing a line saying this is an the organism's developmental system and this is not, we have the easier job of drawing a line saying these individuals count as sharing the "same" developmental system and these do not. This is a familiar problem of taxonomy, instead of an unfamiliar problem of systems theory.

It is true that this implicit way of formalising the developmental architecture does not provide new insights into the questions it sidesteps, such as how to formally describe an architecture. But at least it is clear and stated in an existing formalism. Perhaps a more descriptive, systems theoretic formalisation of a developmental architecture will be easier to imagine, building from this clarity.

This step – defining H as the individuals sharing a given developmental architecture – is crucial both in the mathematical development of the formalism of genetic leverage and in the interpretation of what this formalism means, of what genetic leverage is actually

measuring. Section 5.2, which follows, develops the formalism with an example. Section 5.3 will return to address the question of interpretation, of what it means in our model to say that a set of genotypes "represents" a developmental architecture.

5.2 Definition of genetic fulcrum

Which subset of individual genotypes represents an evolvable developmental architecture? How do we define this subset? Let us work toward an answer by revisiting the segmentation model from the preceding chapter.

5.2.1 Finding an intuitive fulcrum for the evolvable mapping

The key feature of the segmentation model was that the evolvable mapping ϕ_E provided greater mutational connectivity via the small regulatory region of the genome (the V-part) than what the naive mapping ϕ provided via the larger gene-coding region (the F-part). In both mappings, variation in those respective regions of the genome sufficed to produce all possible phenotypes. But locus for locus, the regulatory region in the evolvable mapping packed more evolutionary "punch".

Purely regulatory variation under the evolvable mapping ϕ_E, given a valid gene, corresponds to a subset of genotypes H_E. This is the subset produced by allowing any locus value in the regulatory region of the genome, while also fixing the gene-coding region to a value f that contains a valid gene:

$$H_E = V \times f_E$$

where V = {all length-3 strings}

$$f_E = \boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{0}\boxed{1}\boxed{0}$$

In other words the structure of the evolvable mapping ϕ_E, with its clear functional division between the regulatory and gene-coding parts of the genome, strongly suggests this subset H_E as a natural candidate for calculating the leverage.

But does this choice of f_E work? It seems to work well. With this particular choice, what is exceptional about the evolvable mapping now shows itself in two exceptional mathematical features. First, the range of ϕ_E over the small set H_E encompasses the entire

phenotype space – that is, the range is *complete*. Second, because of the relative smallness of H_E, the genetic leverage is *high*. In other words,

$$\begin{aligned}\phi_E(H_E) &= \{\text{}\} = P \\ \lambda_{\phi_E}(H_E) &= \left|\phi_E(H_E)\right| / |H_E| \\ &= |P| \,/\, |V| \\ &= 4/(2^3) = 1/2\end{aligned}$$

These two mathematical features confirm H_E as a good choice for calculating the leverage of the evolvable mapping.

Generalising toward definitions

Let us recapitulate the above logic in order to generalise it.

In this calculation the genome subsequence f_E is a special kind of constant, a fixed element around which the rest of the genotype varies, and which magnifies the phenotypic effect of that variation. For this reason, let us call this fixed subsequence f_E the *genetic fulcrum*, after the fixed pivot in a mechanical lever. In general, then, we have the following definition: given a genome based on strings of symbols, a genetic fulcrum f is an arrangement of fixed values for certain loci on the genome.

We used the fulcrum f_E to define the subset H_E. Specifying a fulcrum always implies the set of genotypes which contain that fulcrum, so specifying a fulcrum is simply a shorthand for defining a subset of genotypes. In general, this subset which is defined by a fulcrum f, let us call it the fulcrum's *swing* and let us write it $\tilde{f}$.

We started with the mapping, ϕ_E, which was already known to be evolvable because of verbal arguments based on the idea of a core component and because of mathematical features observed during analysis with the genospace algebra. We then found that this mapping also had two notable mathematical features when considered in terms of leverage and fulcrum. The range of the fulcrum's swing encompasses all possible phenotypes – i.e., it was complete. And the leverage was relatively high. These features seem like two new, quantitative indications of evolvability.

If we can now take these features as general indications for evolvability, this suggests that the question "Is ϕ evolvable?" can now be reduced to the following formal question:

> Does there exist a suitable natural fulcrum f – a fulcrum where the range $\phi(\tilde{f})$ is the complete phenotype space and where the genetic leverage $\lambda_\phi(\tilde{f})$ is high?

Does this generalised rule work? To test it I will now apply it to the naive mapping to see if it indicates that the naive mapping is, as we would expect, less evolvable than the evolvable mapping.

5.2.2 Finding the natural fulcrum of the naive mapping

Unlike the evolvable mapping the structure of the naive mapping ϕ does not intuitively suggest an obvious natural fulcrum, so we must experiment with alternatives.

The naive mapping does not install a clear separation between a gene-coding region and a regulatory region that acts indirectly on it. In fact, the regulatory region is ignored. Given this, our first choice might be to calculate the leverage of the gene-coding region alone. This effectively treats the inert regulatory part as the fulcrum, and is the exact complement of the choice made for the evolvable mapping.

$$
\begin{aligned}
H_1 &= \overbrace{\boxed{0}\boxed{0}\boxed{0}}^{f_1} \times \{\text{all length-13 strings}\} \\
\phi(H_1) &= \{\text{}\} = P \\
\lambda_\phi(H_1) &= \left|\phi(H_1)\right| / |H_1| \\
&= 4/(2^{13}) = 1/2^{11}
\end{aligned}
$$

The calculation shows the naive mapping has a lower leverage than the evolvable mapping, as expected.

Identifying the *natural* fulcrum

But perhaps this comparison is too biased by our choice of the fulcrum, which echoes the tidy organisation of the evolvable mapping? To provide an evenhanded comparison between the two mappings, we must calculate each mapping's genetic leverage by using whichever fulcrum is natural for that mapping. But this leads to an important question: what makes a fulcrum "natural" for a mapping, when the mapping itself does not intuitively suggest one choice strongly?

For now, since we are interested in the mapping's maximum possible evolvability, let us stipulate the following: a fulcrum is *natural* for a mapping when its range completely covers the phenotype space, and when it maximises the genetic leverage compared to other possible fulcra. That is, we are taking the two features we observed in the fulcrum f_E – high leverage, complete range – and defining the natural fulcrum by analogous properties – maximum leverage, out of all possible fulcra with a complete range. The qualifier "natural" indicates not that this is a total characterisation of the mapping, but that this is a measure of a mapping's leverage which has been defined in an automatic and unbiased way, without reference to some informal assessment of the mapping.[1]

In the case of the naive mapping, there are fulcra superior to our first choice f_1. For instance, a second choice might be to increase the leverage by choosing a fulcrum that is extended to contain the coding for one gene:

$$H_2 = \overbrace{\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{1}}^{f_2} \times \{ \text{ length-9 strings } \}$$

Given that the leftmost locus of the variable region is a stop/start codon, and will disrupt the gene in the fulcrum, this leaves enough flexibility that the range over H_2 still encompasses all phenotypes. So that gives

$$\lambda_\phi(H_2) = |P| \,/\, |H_2| = 4/(2^9) = 1/(2^7)$$

Compared to the first choice this second choice yields a higher genetic leverage but at the cost of obscurity. Instead of defining a region where 1s are counted (as with the evolvable mapping), or a region where genes are counted (as with the original choice), this choice exposes a more recondite pattern of behaviour in the free region, a behaviour based in the right half on forming valid genes and in the left half on disrupting a gene that is already in place but out of view. If an analyst were unaware of the entire mapping, only privy to the sequence values of the free region, then it would be hard to guess the more consistent logic that defines the mapping as a whole.

We can trade away more clarity for even more leverage. If one accepts a noncontiguous

[1] It might also be interesting to consider other measures of the leverage of a mapping – for instance, the expected leverage. But it is unclear what this would mean, since it would require averaging over all possible fulcra, where different fulcra represent different developmental architectures.

sequence as a fulcrum, it is possible to find a fulcrum for the naive mapping with a much higher leverage. We take a hint from the genospace algebra analysis of the naive mapping, which exposed the counterintuitive phenotypic connectivities resulting from mutations on the start/stop codons. In this third choice, the two loci with asterisks are allowed to vary freely and the rightmost locus can serve as a shared start/stop codon:

$$H_3 = \{\overbrace{\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{1}\boxed{1}\boxed{*}\boxed{0}\boxed{1}\boxed{1}\boxed{1}\boxed{*}\boxed{1}\boxed{1}\boxed{1}\boxed{0}}^{f_3^{(non-contiguous)}} \mid \text{asterisks variable}\}$$

From this we calculate the range and leverage:

$$\begin{aligned}\phi(H_3) &= P \\ \lambda_\phi(H_3) &= |P|/|H_3| \\ &= 4/(2^2) = 1\end{aligned}$$

In fact, this leverage is even higher than that of the evolvable mapping!

Defects of the segmentation model

This is a surprising result and it undermines the argument that the "evolvable mapping" is more evolvable than the naive mapping. Of course it might be mitigated by other differences between the mappings – the structure of connectivity between phenotypes or the size of the phenotypic pre-images, which will affect actual transition probabilities – but at least by the measure of leverage the naive mapping turns out to be more evolvable. What does this show? Does it show that leverage is a poor measure of evolvability? That the basic idea of a core component is flawed? Or that the segmentation model is an imperfect formalisation of that idea?

I do not think it undermines the logic motivating the definition of genetic leverage. The leverage is doing what it is supposed to do: revealing that (contrary to casual appearances) there are fixed loci configurations that render even this very literal mapping highly responsive to a few mutations. In fact the leverage is showing that specific features of the segmentation model make it a poor formalisation of the idea of a core component.

In the calculation, the high leverage of the naive mapping is a result of the fact that

start/stop loci can be shared between adjacent genes. Because of this, a single well-placed mutation in the rightmost free locus can disrupt two genes simultaneously. But the basic idea of a core component is that these kinds of coordinated, multi-gene effects are supposed to be reserved for the regulatory region and a result of the distinctive patterns of epistatic dependency upon the regulatory region. If we defined the genes of the F-part so that they were self-contained, then this sort of effect would be impossible, and the naive mapping would not have a higher leverage.

More generally, another problem is the excessive simplicity of the phenotype space in this model. The idea of a core component is that it enables new kinds of phenotypic variability. Mutations outside of the component should cause it to be expressed in a variety of ways. Mutations within the core component, it is assumed, would be more likely simply to disable that component. The segmentation model is too simple to distinguish between simple, disabling phenotypic effects (from mutations in the core component) and a more varied set of coordinated phenotypic effects (from regulatory mutations affecting how the core component is expressed). In the segmentation model, the only effect any variation can produce is varying a number from 0 to 4. By limiting the variety of phenotypes, this setup limits the supposed advantages of the evolvable mapping and inflates the leverage of the naive mapping in comparison.

Chapter 6 introduces a more elaborate model, the evolvable modularity model, which allows variety in the type and ordering of segments. This makes the model more difficult to handle mathematically, but it brings it much closer to the idea of a core component as it applies to biological organisms. With this change, the evolvable mappings have, as one would expect, a higher genetic leverage.

5.3 Interpretation of this formalism

5.3.1 Heuristic uses of the leverage and fulcrum

While the preceding example highlights a weakness of the segmentation model, it also illustrates the strength of genetic leverage and the fulcrum for aiding insight. Leverage provides an objective way to pick out features of a mapping that are not intuitively obvious, and the fulcrum provides a rigorous way to explore our intuitions.

h	$\phi_E(h \times f_E)$	$\phi_E(h \text{ with } f'_E)$
0 0 0	0	0
0 0 1	1	1
0 1 0	1	0
0 1 1	2	2
1 0 0	1	0
1 0 1	2	2
1 1 1	3	3

Table 5.1: Under the mapping ϕ_E, the two fulcra f_E and f'_E give the same leverage but different positions for the remaining free loci. However, fulcrum f_E yields more insight since it exposes the mapping's underlying logic of simply counting the 1s in the regulatory region.

This is not to say that one should credit these concepts because they gibe with intuition (although, in fact, these concepts were developed by formalising intuitive ideas). Rather, it is that the precision of these concepts can extend our intuition into difficult cases.

By definition the natural fulcrum must, first, give the maximum genetic leverage and, second, encompass the whole phenotype space in its range. We can hope that the natural fulcrum will also reflect some structure in the mapping that makes intuitive sense, like the division between the gene-coding and regulatory region in the evolvable mapping. But this last condition must remain only a hope since it may conflict with the two definitional requirements. In a mapping plucked from a biological system, not a system designed for illustrative clarity, the fulcrum which maximises leverage might be cryptic and bizarre instead of obvious and intuitive. This is the *objectifying value* of genetic leverage: it provides an objective way to identify the natural fulcrum via calculation, in obscure cases where our intuition is misleading.

However, even once we find a natural fulcrum which maximises genetic leverage it is still worth looking for an intuitive fulcrum, since this is more likely to shed light on the underlying biology. Dependency patterns in the genotype-phenotype map are not mathematical curios. They are representations of concrete biological constraints rooted in the dynamics of development and gene expression. The purpose of the mathematical apparatus is to shed light back onto this concrete system by distilling its essential features.

A high genetic leverage under some fulcrum is evidence of evolvability; but real insight into that evolvability, a real understanding of it, depends on comprehending what concrete biological system that particular fulcrum represents. An intuitive fulcrum can help us understand, rather than merely measure, the evolvability of the system. This is the *framing value* of the genetic fulcrum: it provides a well-defined construct for framing hypotheses about the structure in the genotype-phenotype mapping that are intuitively comprehensible and easily related to the concrete biology.

This contrast between calculation and insight is illustrated, for instance, by noting how the evolvable mapping ϕ_E has a cryptic natural fulcrum in addition to the intuitive natural fulcrum. The fulcrum f_E clearly corresponded to a gene that coded for a body segment, making it easy to recognise that the free region regulated repetitive expression of that gene. This cryptic fulcrum f'_E is also natural because it also gives the maximum genetic leverage. However, it is noncontiguous. Instead of fixing the gene-coding region and freeing the regulatory region, it works by fixing one locus of the regulatory region and freeing one locus of the gene-coding region:

$$H'_E = \{\overbrace{\boxed{*}\boxed{*}\boxed{1}\boxed{0}\boxed{1}\boxed{1}\boxed{*}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}}^{f'_E} \mid \text{asterisks variable}\}$$

This gives the same leverage of $\lambda_{\phi_E}(H'_E) = 1/2$. But if one looked only at the free loci allowed by this fulcrum, compared to the intuitive fulcrum, it would be harder to guess the larger pattern that governs the entire mapping – the simple pattern of counting 1s in the regulatory region. (Table 5.1 makes this comparison.) So although these fulcra both give the same leverage, they do not equally hint at the underlying developmental logic of abstracted repetition.

In sum, choosing a fulcrum is inevitably complex since it reflects the complexities of the underlying problem – the multiple mathematical ways to choose a fulcrum, the hard-to-formalise fact that some choices aid intuition, and the meaning of a choice in terms of the underlying biology.

Biological idea	Formalisation
Developmental architecture	$\phi : G \to P$, the genotype-phenotype map
Is this developmental architecture evolvable?	Is ϕ evolvable? • Are there evolvability signatures in its genospace diagram? • Is its leverage λ_ϕ high?
How is the developmental architecture evolvable?	What are the particular structures in the genospace diagram of ϕ?
What makes this developmental architecture evolvable?	What is ϕ's natural fulcrum f?
What is the developmental architecture's chief core component?	What phenotypic features are produced by f?

Table 5.2: The formal tools of the genospace diagram, genetic leverage, and the genetic fulcrum all represent concrete biological ideas.

5.3.2 Fulcrum as the core component

The segmentation models were introduced as abstractions, already fully-formed, rather than worked up as approximations of a known empirical system. This may obscure what it means to say that constructs like the fulcrum should "correspond" to concrete biological realities. To flesh that out I would like to review the relationships between the informal ideas of chapter 1 – the developmental architecture and the core component – and the formalisations introduced in this chapter and earlier – the genotype-phenotype map, the genetic fulcrum, and the genetic leverage.

As discussed in section 1.2.2, the term developmental architecture refers to the developmental system of an organism, but conceived of down to its genetic underpinnings rather than as a sequence of physical transformations. Shared developmental architectures tend to define meaningful groupings of individuals, such as body plans. A given developmental architecture might be evolvable, or not. Throughout this dissertation, I have formalised the developmental architecture with the construct of the genotype-phenotype mapping function, $\phi : G \to P$, a rule associating genotypes with phenotypes. A given map might be evolvable, or not.

One particular way a developmental architecture might be evolvable, which I have

focused on in depth, is for it to possess a core component that increases evolvability. As described by Kirschner and Gerhart, the theory of core components assumes that certain regulatory, metabolic, and developmental mechanisms are relatively constant over time, that they establish a context shaping future evolutionary change, and that they are sufficiently logically distinct to warrant analysis in their own right. These mechanisms are the actual, *physical* core components of the developmental architecture of the organism. These stable physical mechanisms will leave their signature in corresponding stable features of the genotype-phenotype mapping function. In the mapping, these features are the genetic, *formal* representation of the core components of the developmental architecture.

This is what is captured by the fulcrum. The fulcrum is the fixed part of the genotype which corresponds to – which codes for – the core components.

Of course merely calling something a core component does not establish that it increases evolvability. Any bundle of features might be spuriously regarded as a core component, just as any subsequence might be tried as a fulcrum. But the defining requirement of a core component is that it should create a fertile context for evolutionary changes outside that component. This requirement is what is captured by the measure of genetic leverage, which measures how much a fulcrum amplifies the effect of variation outside the fulcrum.

This also clarifies a subtle question which may have troubled alert readers: what did it mean to say a genotype-phenotype mapping was evolvable when we are now saying a particular fulcrum is evolvable? A genotype-phenotype map is evolvable because it supports a high leverage fulcrum. That is, a developmental architecture is evolvable because it includes a core component that produces a high evolvability.

5.3.3 Fulcrum synthesises two schools of research

It is also worth pointing out how the leverage-and-fulcrum scheme synthesises past work in evolvability, which may be regarded as falling into two broad schools of research, separated by incompatible methodological approaches.

On one hand one school includes research on molecular evolution in RNA, computational evolution, and various detailed mathematical models (described in chapter 2). The methodologies of these areas offer total knowledge of the genotype at the per locus level. This leads, very naturally, to an approach that explains evolvability in terms of *microscopic variation in the genotype* – specifically, how the genotype-phenotype map translates genotypic

variation into phenotypic variation.

This approach is correct. Evolvability is fundamentally about variability. This dissertation follows in that tradition, as the genetic leverage is essentially a measure of the amplification of variability.

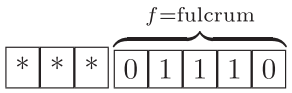
But focusing so narrowly on the microscopic structure of variation blinds one to its causes. This is the weakness of this approach: it leaves evolvability only as a brute fact, an irreducible property of certain genotype-phenotype mappings and not of others. In addition, these fields all describe relatively esoteric kinds of phenotypes, so the physical structure of the phenotype itself cannot offer much insight into the reason for these effects on evolvability. This approach *documents* the evolvability, but it does not *explain* what causes it.

On the other hand is the school that includes research on anatomy, development, and macroevolutionary patterns (described in chapter 1). This work studies the proliferation of actual biological species, the variation in their body plans, and the likely role of developmental systems in shaping that variation. Real organisms are complex systems and it is not easy to understand them internally at the microscopic, per-locus level. But from the outside one can survey their variety and deduce their history, seeing the big picture of their commonalities and evolution. This leads, very naturally, to an approach that explains evolvability in terms of *macroscopic constancy in the phenotype* – the parts of the organism, or the patterns in the body plan, which remain unchanged across taxa and over time.

This approach is also correct. According the theory of core components, these constant parts of the organism create the context that enables what is inconstant and variable. These constant parts of the organism correspond to the constant parts of the genotype – formalised here as the fulcrum.

Evolvability studies have been broadly separated by microscopic approaches which focus on variability in the genotype and observational approaches which focus on constancy in the phenotype. What is distinctive about the idea of the leverage-and-fulcrum scheme is that it simultaneously describes the constant parts of the organism which *cause* evolvability and the patterns of variability in the organism which *constitute* evolvability.

1. Choose a fulcrum (f), a fixed arrangement of loci values on the genome

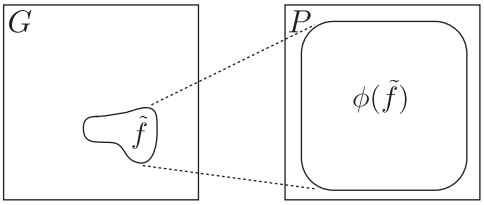


2. Every fulcrum implicitly defines a swing ($\tilde{f}$), the set of genotypes with the fixed loci values of the fulcrum.

The more loci are fixed in the fulcrum, the smaller the number of genotypes in the swing. The smaller the swing, the fewer mutations are required to evolve from one genotype to another.

$$\tilde{f} = \{0\,0\,0\,0\,1\,1\,1\,0,\ 0\,0\,1\ \vdots\ ,\ 0\,1\,0\ \vdots\ ,\ \vdots\ \vdots\ ,\ 1\,1\,1\ \vdots\ \}$$

3. One important case is the natural fulcrum, the largest fulcrum such that the range of the genotype-phenotype mapping (Φ) over the swing covers every possible phenotype.



This is the fulcrum which most amplifies the phenotypic effect of genotypic mutation, without limiting the possible results of mutation.

4. The degree of amplification can be measured by the genetic leverage (λ), the size ratio of the range over the swing to the swing itself. This is a crude measure of evolvability.

$$\text{leverage} = \lambda_\phi(\tilde{f}) = \frac{\left|\phi(\tilde{f})\right|}{\left|\tilde{f}\right|}$$

Figure 5.2: The genetic leverage of the natural fulcrum gives a crude measure of the evolvability of a mapping.

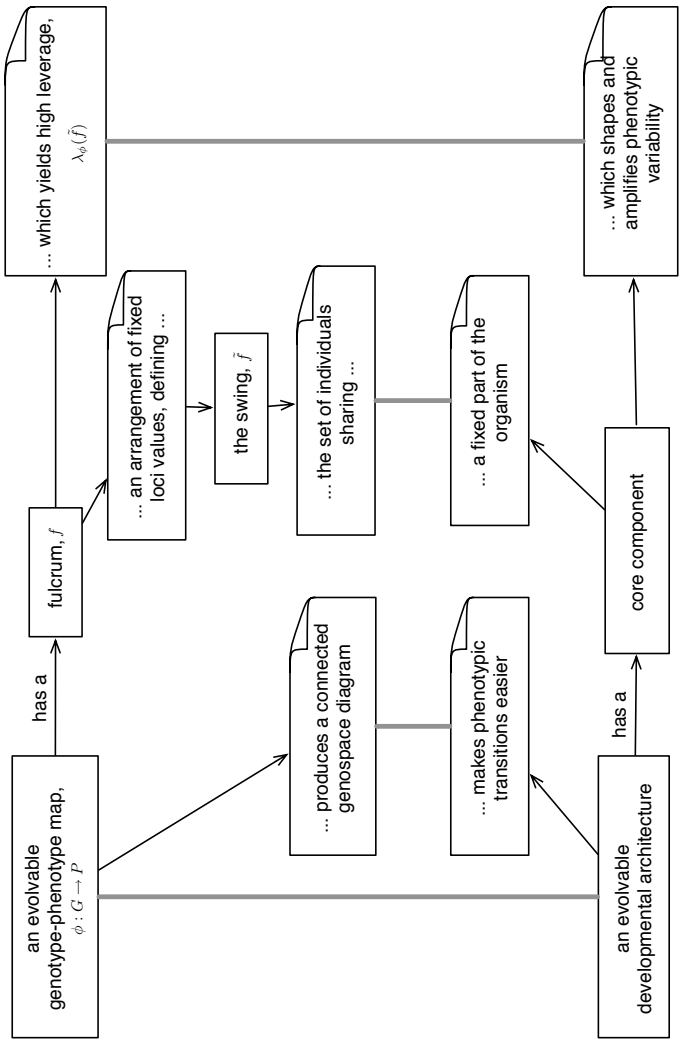


Figure 5.3: All mathematical constructs (above) formalise biological ideas (below).

Chapter 6

Evolvable modularity

One tantalising promise is that evolvability might help to explain the evolution of modularity and other forms of biological complexity. And one excellent reason to doubt this is that it seems to explain the unknown in terms of the incomprehensible. Both evolvability and modularity remain rather vague terms, which is why this dissertation is trying to formalise them.

To that end, chapter 4 analysed a model of repeated segmentation. As the repetition of identical body segments presents the simplest imaginable instance of phenotypic modularity, it has appeared in the very earliest discussions of the connection between evolvability and modularity (Dawkins, 1989). The simplicity of segmentation makes it easy to illustrate basic ideas about evolvability – for example, the idea that evolvability is grounded in the structure of mutational relations, and that it results from the genotype exerting a kind of abstracted control over the phenotype.

However, this simplicity also makes it hard to see how these ideas apply to modularity in general. For example, what does "abstracted control" mean, besides abstracting description from repetition? Repeated segmentation alone is too simple to represent the fuller, intuitive, biologically inspired notion of modularity.

To capture a richer notion of phenotypic modularity, this chapter introduces a slightly more elaborate model. This model allows for two types of modules, segments with wings and segments with legs. Implementing this modest change in an explicit way has significant ramifications. It requires new features in the genotype-phenotype mapping, features which enable new modes of abstraction and thus further illustrate the idea of abstracted control.

This model also provides another example of analysing a system's evolvability using genetic leverage and genetic fulcra.

All of this, it is hoped, will make more explicit the argument for how the evolvability of certain developmental architectures might facilitate the evolution of modularity.

6.1 Two genotype-phenotype maps

First we define the possible genotypes. As with the segmentation model, we assume a binary genome, where every locus has an allelic value of 0 or 1. We assume every genome has a length of 24 loci. This length is chosen to be long enough to support a more complicated mapping function while also small enough to allow convenient analysis (not least, for instance, to allow the entire genome to be printed on a single page). The genotype space G is thus as follows:

$$G = \{[0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0],$$
$$[0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1],$$
$$\vdots$$
$$[1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1]\}$$

How are these genotypes connected? As with the segmentation model, we assume mutation always changes the allelic value of a single locus.

Now we define the phenotype space. Introducing different kinds of segments also makes it possible to distinguish the order of those segments in the body, which increases the number of phenotypes. As with the segmentation model, let us assume that there may be between zero and three segments. As many as three possible positions in a body, combined with two possible types of segment in each position, gives us fifteen possible phenotypes ($2^0 + 2^1 + 2^2 + 2^3 = 15$). So our phenotype space is as follows:

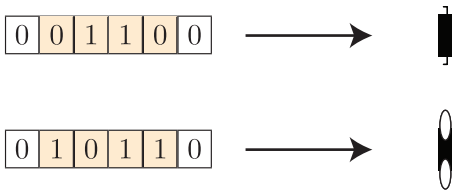
$P = \{$  ,
 ,
 ,

Since phenotypes are distinguishable by segment ordering, it will be necessary to define the mapping so that it specifies segment ordering.

(This is the first new mapping feature that follows from stipulating multiple segment types: *module ordering*.)

Having defined the genotype and phenotype spaces, it now remains to describe the mappings. Similarly to the segmentation model, we divide the genome into two parts: the first six loci will make up the regulatory region (V part), and the remaining eighteen loci will make up the gene-coding region (F part). The mappings depend on genes and regulatory elements, so before describing the mappings we will describe the structure of these elements.

Gene structure The segmentation model allowed only one kind of gene and one kind of body segment. Since there are now two kinds of body segments – wings and legs – let us assume there are two corresponding kinds of genes, as shown in the highlighted regions:



The sequences themselves are arbitrary. What matters is their fixed association with the body segments.

(This variety of genes is the second complication that follows from stipulating multiple segment types: *gene discrimination*.)

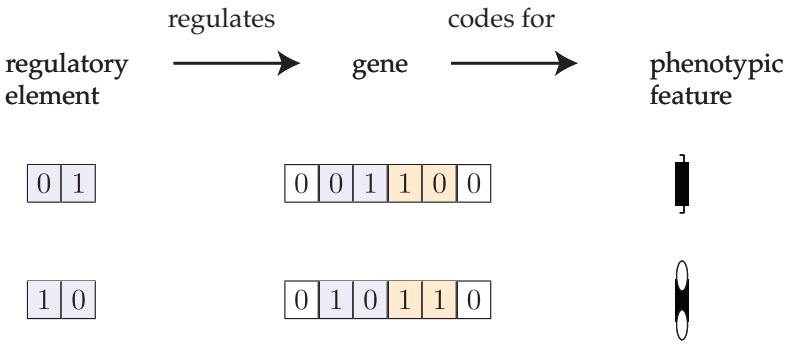
As previously, we suppose that the genes must be bounded by 0's. For the sake of consistent usage I will continue to call these start/stop codons, although that is no longer quite accurate since the wing gene contains a 0 that does not stop it from being read.

How are genes read? For the sake of simplicity let us posit that the gene reading process works from left to right, reading only one gene at a time, so that if two genes share any loci or start/stop codons then only the leftmost gene is expressed. This is a slight change from the reading process in the segmentation model, which allowed shared start/stop codons.[1]

[1]Why adopt this new restriction? It is to keep the model tractable. The segmentation model used a simpler, more inclusive reading process: it assumed that every instance of the subsequence defining a properly bounded

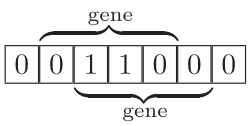
Regulatory elements Allowing two kinds of body segments complicates gene regulation as well as the genes themselves. In the segmentation model, when the regulatory region (the V-part) was processed by the evolvable mapping, it only needed to produce a repetition counter between zero and three. This indicated how many times the segment gene would be expressed. Since the model presumed only one kind of gene, there was no need for the regulatory information to specify which gene the repetition counter referred to. But since we now allow two kinds of phenotypic traits, we need not only different kinds of genes for the different kinds of tissue but also different regulatory elements that specify which gene they regulate.

How should regulatory elements specify a given gene? Let us keep it simple. Let us stipulate that a regulatory element consists of two loci, and that it regulates any gene which begins with those same values for its first two loci. I will call the first two loci of a gene its *signature* and highlight it differently as follows:



gene would be recognised as a valid gene. Since it is possible for subsequences to overlap each other (for instance, the sequence 1 1 1 contains two instances of the subsequence 1 1), this makes it theoretically possible for two genes to overlap by sharing loci. But since a gene in the segmentation model was only a row of three 1's (1 1 1), the only kind of overlap that was possible in practise was of the bounding start/stop codons.

The best thing would be to build the modularity model keeping everything but the mappings in common with the segmentation model: the same reading process, the same use of a binary genome, the same genome structure, etc. Unfortunately this is not practical. Because the genes are more complex, the same reading process could cause the genes themselves to overlap. For instance, if we defined two genes 1 1 0 0 and 0 0 1 1, then consider the following sequence:



This is analogous to overlapping open reading frames, which play a role in viral DNA and also sometimes in mammals (Kozak, 2001). However, these interesting effects also greatly complicate the analysis of the ordering and the number of genes that can fit in the F region, so it is simpler to eliminate them. I do not believe this significantly compromises this model. Overlapping reading frames are an interesting exception, not the rule, in biological systems.

Another way to avoid these effects would be to create a third allelic value reserved for start/stop codons. But switching from a 2-valued genome to a 3-valued genome would do even more to spoil even-handed comparison with the segmentation model, so I choose the simpler solution of disallowing these effects by stipulation.

This defines the basic relationship between regulatory elements, genes, and phenotypic features.

(Establishing this relationship between regulatory elements and genes is the third new mapping feature that follows from stipulating multiple segment types: *regulatory addressing*.)

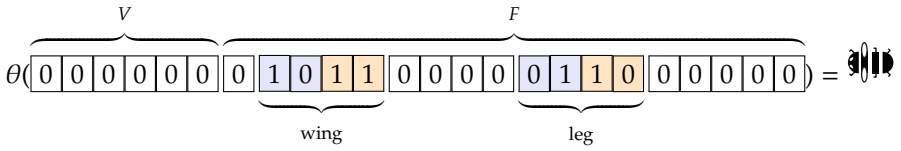
The regulatory elements lack start/stop codons. As such, they are not able to float freely within the regulatory region. Therefore, we suppose that they may appear only at three fixed positions or slots within the six loci gene regulatory region: the two loci blocks starting at position 1, position 3, and position 5 on the genome.

In contrast, the genes do posses start/stop codons, so we suppose they may appear anywhere in the fifteen loci gene-coding region.

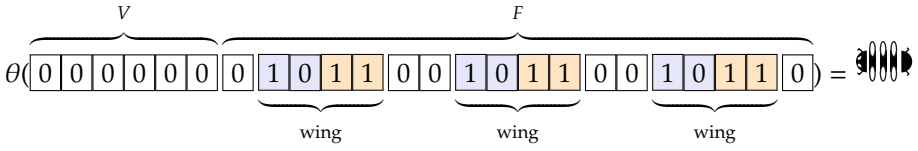
Genotype-phenotype mappings Having defined the structure and allowed positions of genes and regulatory elements, we can now define the genotype-phenotype mappings. First we define the naive mapping $\theta : G \to P$. This mapping ignores the regulatory region entirely and simply express the genes as it reads them left to right, disallowing overlapping reading frames. In other words,

$$\theta(\gamma) = \begin{array}{l}\text{phenotype with segments in or-}\\ \text{der of corresponding genes in } \gamma\end{array}$$

For instance, the following genome has a wing gene then a leg gene:



The following has three wings:

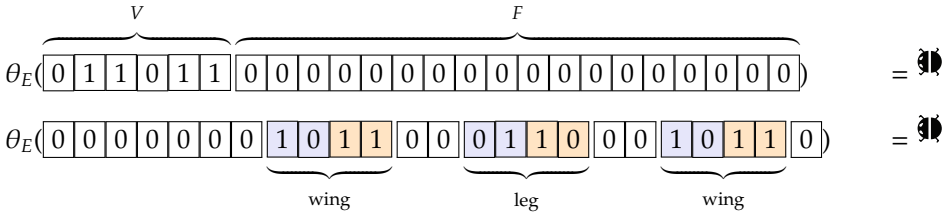


How do we define the evolvable mapping $\theta_E : G \to P$? Let us aim for the simplest possible solution. As with the segmentation model, we suppose that a regulatory unit can only

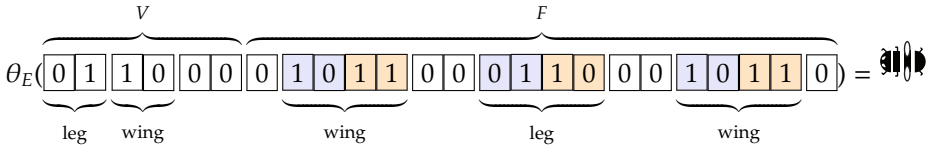
affect a gene if that gene exists somewhere in the gene-coding region. Let us also suppose that the regulatory region, like the gene-coding region, is directly processed left to right. The three possible positions for regulatory elements then correspond to the three possible phenotypic features. In other words,

$$\theta_E(\gamma) = \begin{array}{l}\text{phenotype with segments in order of corre-}\\ \text{sponding regulatory elements, provided the}\\ \text{regulated genes are present}\end{array}$$

So for instance, if regulatory elements are present but genes are not, or if genes are present but unregulated, then no features are expressed:



But if genes are presents and are regulated, then the regulatory elements determines the order and number of segments in the phenotype:



6.2 Analysis of the mappings

Informally, we would expect the regulatory mechanisms of the evolvable map θ_E to make it more evolvable than the naive map θ. Let us now analyse these two maps in detail to see if that is the case, using the formal tools of the genetic leverage developed in chapter 5.

6.2.1 Comparison by genetic leverage

First we calculate the genetic leverage of the evolvable mapping θ_E. To calculate a leverage, we must find a fulcrum. Based on the structure of the mapping, an intuitive choice of fulcrum would seem to be any value of the gene-coding region which possessed both genes

intact. Let us try that first. As usual, the fulcrum implies the swing.

$$f_E = \boxed{0}\underbrace{\boxed{1}\boxed{0}\boxed{1}\boxed{1}}_{\text{wing}}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\underbrace{\boxed{0}\boxed{1}\boxed{1}\boxed{0}}_{\text{leg}}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}$$

$$\tilde{f_E} = V \times f$$

where V = {all length-6 strings}

Now we verify the range of θ_E is complete and calculate the genetic leverage:

$$\begin{aligned}
\theta_E(\tilde{f_E}) &= P \\
\lambda_{\theta_E}(\tilde{f_E}) &= \left|\theta_E(\tilde{f_E})\right| / \left|\tilde{f_E}\right| \\
&= |P| \,/\, |V| \\
&= 15/(2^6) = 15/64 \approx 2.3 * 10^{-1}
\end{aligned}$$

Can we increase the leverage by making the fulcrum a little bigger, by finding some locus in the regulatory region whose value we can fix without affecting the range? In fact, no. Because the two valid values for regulatory elements $\boxed{0}\boxed{1}$ and $\boxed{1}\boxed{0}$ differ from each other in both their first and second positions, and because all three slots in the regulatory region are needed to encode the three segment phenotypes, there is no better fulcrum. f_E is in fact the natural fulcrum for θ_E, the fulcrum which gives the mapping's maximum genetic leverage.

How does it compare with the genetic leverage for the naive mapping θ? To calculate its genetic leverage, we must find its natural genetic fulcrum. One obvious starting point is to begin by fixing the regulatory region, which is irrelevant to the naive mapping θ.

$$\begin{aligned}
f_1 &= \boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0} \\
H_1 = \tilde{f_1} &= \boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0} \times \{\text{all length-18 strings}\} \\
\theta(H_1) &= P \\
\lambda_\theta(H_1) &= |\theta(H_1)|/|H_1| \\
&= 15/(2^{18}) \approx 5.7 * 10^{-5}
\end{aligned}$$

This implies a difference of four orders of magnitude between the genetic leverage of θ and θ_E. But recall that in the segmentation model, a little investigation revealed additional fulcra that improved the leverage substantially. Are superior fulcra present here as well, or is this the natural fulcrum?

In order to encode for all the three segments phenotypes, the fulcrum must allow three valid genes to be in place. Because of packing limits, this fixes the positions of those genes. Therefore we know the natural fulcrum cannot fix loci that would interfere with three valid genes. However, it can fix loci that are compatible with three valid genes, allowing variation within the genes to disrupt them.

Sticking to a contiguous fulcrum, one obvious improvement is to fix the first start/stop locus, the one adjacent to the regulatory region:

$$
\begin{aligned}
f_2 &= \boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0} \\
H_2 &= \tilde{f}_2 = \boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0} \times \{\text{all length-17 strings}\} \\
\theta(H_2) &= P \\
\lambda_\theta(H_2) &= |\theta(H_2)|/|H_2| \\
&= 15/(2^{17}) \approx 1.1 * 10^{-4}
\end{aligned}
$$

Allowing for noncontiguous fulcra, we can further imagine fixing all the start/stop codons:

$$
\begin{aligned}
H_3 &= \overbrace{\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}}^{V}\overbrace{\boxed{0}\boxed{*}\boxed{*}\boxed{*}\boxed{*}\boxed{0}\boxed{0}\boxed{*}\boxed{*}\boxed{*}\boxed{*}\boxed{0}\boxed{0}\boxed{*}\boxed{*}\boxed{*}\boxed{*}\boxed{0}}^{F} \\
\theta(H_3) &= P \\
\lambda_\theta(H_3) &= |\theta(H_3)|/|H_3| \\
&= 15/(2^{12}) \approx 3.7 * 10^{-3}
\end{aligned}
$$

This is lower, but still leaves two orders of magnitude difference in the leverage. Can we fix more loci? Can we fix loci within the gene slots and still maintain the variability necessary to produce all genotypes? In fact we can, since we can fix the locus which the two valid genes have in common – the third locus. This will cause the mutations in the three other

loci to determine whether the gene is a wing gene, a leg gene, or an invalid sequence.

$$H_4 = \overbrace{\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}\boxed{0}}^{V}\overbrace{\boxed{0}\boxed{*}\boxed{*}\boxed{1}\boxed{*}\boxed{0}\boxed{0}\boxed{*}\boxed{*}\boxed{1}\boxed{*}\boxed{0}\boxed{0}\boxed{*}\boxed{*}\boxed{1}\boxed{*}\boxed{0}}^{F}$$

$$\theta(H_4) = P$$

$$\lambda_\theta(H_4) = |\theta(H_4)|/|H_4|$$

$$= 15/(2^9) \approx 3.0 * 10^{-2}$$

This appear to be the maximum genetic leverage of θ. It still much lower than the leverage of the evolvable mapping θ_E by a factor of eight:

$$\frac{\lambda_\theta}{\lambda_{\theta_E}} = \frac{15/(2^9)}{15/(2^6)} = 2^{-3} = 1/8$$

This confirms the intuitive expectation, based on the theory of core components, that the regulatory mechanisms increase evolvability by reducing the amount of mutational variation necessary to achieve complete phenotypic variability.

6.3 Abstracted control

What is the significance of this model? For one, it provides an example where genetic leverage confirms that core components increase evolvability.

But in addition, by providing a richer example of modularity, it also illustrates how evolvability might promote modularity and complexity through developmental mechanisms of abstracted control. The deeper motivation for the models in this dissertation is not just to provide examples of evolvability, nor just to understand segmentation or modularity as such. It is to present exact models that can test this much grander idea that evolvability promotes complexity.

Indeed, the idea is so grand it might be called meaningless. Evolvability, complexity, abstracted control – these are all vague terms. I have tried to pin down "evolvability" as much as possible by formalising the idea. However, in the case of "complexity", such a formalisation would be a lie. Despite much research no one has produced a definition of the term that is both exact and an accurate representation of what people are talking about

when they talk about complexity.[2] I will simply assume that complexity, in organisms or mechanisms, is an advanced form of modularity that includes functional integration. That is, we call organisms complex because they are divided into many parts which work together in subtle way. This modest definition at least does not pretend to a specious precision.

Chapter 4 introduced the term "abstracted control". This idea is so tied up with the nature of complexity that it inherits the same frustrating resistance to formalisation, but formal models let us spell out specific instances of it. The evolvable segmentation mapping ϕ_E featured the *abstraction of repetition*, i.e., the abstraction of the number of body segments from the description of how to build a body segment. The evolvable modularity mapping θ_E displays other modes of abstraction, further illustrating the idea.

Aspects of modularity in the phenotype

To describe these new modes of abstraction in θ_E it is helpful first to consider the new phenotypes. P clearly defines a set of modular phenotypes but what does "modular" really mean? The word gets at a cluster of ideas, or aspects of modularity.[3] I will point out two.

First, the mere fact that it is possible to speak individually of a "wing segment"or a "leg segment" implies that these body segments exist as "more or less self-contained units", to borrow the words of the dictionary definition of a module (OED, 2002b,a). I will refer to this quasi-independence as a module's *integrality*. Second, the variety of viable orderings of the wing and leg segments shows that they are compatible with each other. Although their insides are different, their outsides are similar enough that they can join up. They are like Lego blocks of different colours that can nevertheless all snap together. I will refer to this aspect by saying that these modules have *compatible interfaces*.

[2] Adami (2002) provides a good review of the limitations of past computational and information theoretic definitions of complexity, arguing for shared Kolmogorov complexity as an alternative candidate.

[3] Historically the scope of the word "modularity" has been closer to machines than organisms. The earliest usage of the word implies only the sense of measure or extent ("According to the module of my slender skill" 1583), although an architectural usage, indicating a plan or design, soon hints at the modern connotations of construction ("Send unto us a plat forme or module of the situation of the said mylne upon the river." 1589). Architecture also provides the first use of the more specific modern sense, suggesting parts designed to function together ("The unit used in this kind of modular design may be defined as the 'multiple module'. Modularity of this kind has much greater potentialities as an integrating factor for building as a whole." 1937). This appears in the early 20th century, the same time as the introduction into factories of modern repetitive flow production methods requiring interchangeable parts (i.e., assembly line mass production). This suggests how technological advance has shaped our current perspective on living things.

Modes of abstraction in the mapping

Here is a key point: these two aspects of modularity in the phenotypes P are the prerequisites for the modes of abstraction in the mapping θ_E. What are these modes?

First and most easily overlooked, the fact that modules are integral – that they exist as quasi-independent entities – is what makes it possible for parts of the mapping to *address* them as individual, quasi-independent entities. For instance, in θ_E, every regulatory element determines the tissue type of a specific body segment. If a phenotype could not be reasonably subdivided into body segments, and instead were nothing but a vast tangle of uniformly intertwined mechanisms, then a genotype-phenotype mapping could not produce coherent effects on specific segments because there would be no real segments to specify. And even with a modular phenotype, a mapping could still fail to reflect that modularity, with every single mutation scrambling the internals of multiple modules willy-nilly. In other words, just regulating specific phenotypic modules, qua modules, is a powerful mode of abstraction. I will refer to this as the *abstraction of addressing*.

Second, the fact that body segments have compatible interfaces is what makes it possible for the mapping θ_E to control the ordering of body segments – that is, to control the ordering of modules independently of the ordering of the genes describing them. This easy regulatory control of ordering I will call the *abstraction of ordering*.

So these two modes of abstraction are further examples of abstracted control. And looking back, we can also see how the abstraction of repetition in the segmentation model also depended on compatible interfaces.[4]

6.3.1 Abstraction control and complexity

So what? What does this array of terminology actually explain? The value of these terms, I think, is that they clarify the claim that evolvability promotes modularity and complexity. It is not just that evolvability facilitates evolution, that complexity is difficult to evolve, and so evolvability is especially handy for the evolution of complexity. There is a more specific story about how they are connected. Here is how it goes.

[4]One could draw a finer distinction between being able to affect *some* module as a single thing (what we might call the abstraction of singularity), and being able to specify a *particular* module to affect it as a single thing (the abstraction of addressing, which presumes singularity). Under this distinction the abstraction of repetition requires the abstraction of singularity, but not of addressing. I will neglect this subtlety here.

Aspects of modularity have design value

The first thing to notice is that the modularity of phenotypes – their possessing various aspects of modularity – is a positive fact in its own right. One could imagine a world of organisms that were not modular, that were just tangles of uniform interdependency. It is not logically inevitable that we see everywhere more or less modular creatures. It demands explanation.

Why are organisms as modular as they are? It is a speculation but surely reasonable to imagine that it is partly because this is simply the way it makes sense for complex things to be built. For instance, consider the integrality of modules, their quasi-independence. One result of this is that if a module is damaged or fails, the effects are mainly confined to that module. Of course if a module is essential to the organism as a whole, like the heart, its failure is still catastrophic, but integrality guarantees that many other forms of damage are relatively limited. This is a valuable property and it is why we recognise integrality in many products of human engineering as well.

Integrality is not the only such property. There are other aspects of modularity that are generic to complex, integrated, modular systems, whether they are organisms or products of human engineering: compatible interfaces, specialisation of function, the use of hierarchical structures of dependence, certain forms of redundancy, etc.

In other words, the hypothesis is that the various aspects of modularity are not arbitrary properties which are handy for defining complexity. They are intrinsically valuable because of universal constraints on the engineering of complex systems.

Modes of abstraction facilitate modular design

The second thing to notice is that not only do aspects of modularity define ways it makes sense for *things to be built*, they also seem to entail ways that it makes sense *to build things*.

These are the various modes of abstraction. They not only depend on the aspects of modularity. They also make it easier to describe succinctly the complex structures possessing those aspects, to describe how to build them and modify them. This is because they are essentially abstracting essential regularities in the structures themselves.

For instance, if a system features repeated components, then the abstraction of repetition allows you to describe it just by describing the types of components and specifying how

many times they should be repeated, rather than repeating the description for every instance of it. This was the model of chapter 4.

In this chapter the abstraction of ordering obviously facilitates new modular design by re-ordering modules of old designs. Re-ordering them via regulation is easier than re-ordering the underlying genes.

More fundamentally, what is going on in is that the abstraction of addressing is enabling the manipulation of modules as atomic blocks, instead of modifying their internal. This has evolutionary value because it tends to preserve the integrity of modules. When regulatory elements address a module as a whole, regulatory mutations are unable to disrupt a module's internal mechanisms. Or to put it differently, this abstraction facilitates evolutionary changes that are integrity-preserving. It might be good or bad to add an extra pair of eyes, or to shift them from the anterior to the posterior, but it will usually be a mistake to destroy the integrity of the eye as a module. This is nothing more than the expectation that it is better to work creatively with the modules you have than to start from scratch.

Interplay of evolvability and complexity

This argument suggests a deep interplay between evolvability and complexity. While it might be the case that evolvability merely makes all sort of evolution easier, including complexity, there are also these theoretical or speculative reasons to believe that the sorts of abstraction that increase evolvability also specifically facilitate complexity.

That is, a language of abstracted control is especially suited for building complex, integrated, multi-component systems, and for changing them easily and non-destructively. This is because the regularities in a system that make an abstracted description possible are the same regularities that are characteristic of complex, integrated, functioning systems. These regularities make changes easier because abstracted description is a compressed, shortened description, so it requires fewer mutations to effect changes. These regularities facilitate changes that are non-destructive, because such a language can respect the functional constraints of the system.

So the relationship between evolvable developmental architecture and phenotypic complexity is a two way street. An evolvable developmental architecture facilitates certain kinds of phenotypic modularity; and phenotypic modularity enables certain kinds of evolvable developmental architectures. Or to put it another way, when a phenotype is built from

certain kinds of building blocks, then it can be described in a *certain kind* of genetic language which facilitates positive evolutionary change. This is the distinction I have earlier described between modularity of the phenotype and modularity of the mapping. Riedl describes similar ideas in terms of the "imitative epigenotype" (Riedl, 1978).

Abstraction premium vs. size premium

Last, this line of thinking provides a new way to understand the sources of evolvability in core components-type theory. As is shown by calculations of genetic leverage, the evolvability is higher the smaller one can make the regulatory region. A key part of the core component model is that the fixed region, the fulcrum, coding for the core components, is just much larger than the size of the free region of regulatory elements. So to a certain extent, the degree of increased evolvability depends simply on how small the regulatory elements are. The smaller they are, the smaller the swing, the higher the genetic leverage. One might call this contribution to the increased evolvability under an evolvable mapping the *size premium*.

But this is not the only dynamic at work. This is not just because regulatory and coding elements have different sizes. It is because, as the model of the phenotype becomes richer, more complex modes of abstraction become possible in the genotype-phenotype mapping, and a small amount of regulatory material is able to parametrise an increasingly large number of phenotypes because of the greater abstraction and greater compression with which the regulatory material is describing those phenotypes. One might call this contribution to the increased evolvability the *abstraction premium*.

Chapter 7

Conclusion

This chapter summarises the principle findings of this dissertation. These findings fit into two broad categories – a critical review of the literature on evolvability to clarify the basic concept and the argument for why it matters, and the development of new formal models and methods for studying evolvability.

While the term "evolvability" may be uncommon the topic is central and important. Perhaps the main point to come out of this review is that evolvability merely addresses, from a modern Darwinian perspective, very old questions about the origin and the unique complexity of living things – i.e., the question of why a rich biological world exists at all, instead of a world of mere physics. These venerable and profound questions deserve direct attention, especially now that recent experimental and conceptual work has created exciting new avenues for progress.

7.1 Critical review

This dissertation reviews, for the first time, all the literature on evolvability. It does not discuss every single published use of the term (this would be impossible), but it does touch on every distinct line of research and it discusses the significant research in depth. The literature on evolvability spans multiple disciplines and is rather inchoate. This makes such a review difficult but also valuable. The broad perspective allows us to distinguish different senses attached to the term, to identify the underlying key ideas which are easily obscured by different research approaches, and to recognise how these approaches can

complement each other in the future.

7.1.1 Kinds of evolvability

This review exposes five distinct concepts of evolvability which I designate as follows: trait evolvability, individual fitness evolvability, individual mutational evolvability, organismic evolvability, and substrate evolvability. These concepts are most easily categorised based on the *locus* of evolvability, or what kind of thing evolvability is a property of – a particular phenotypic trait, an individual organism, an organism as a kind like a species, or the underlying logical substrate that defines the organism's world.

Chapter 1 discusses the concept of organismic evolvability, the ability of a kind of organism, like a species, to evolve more easily, especially as regards the evolution of complexity. This ability is rooted in the organism's propensity to generate useful phenotypic variation from random genotypic variation. Since it is the developmental system which translates genotypes to phenotypes, the study of evolvability is essentially the study of the developmental system.

The idea that evolvability is real and important is suggested by a few lines of evidence. Differences in evolvability may explain observed differences in diversification and in complexity between lineages in the macroevolutionary record. Some organisms seem better at developing variants than others. Also, given that an organism is a kind of a machine, common sense suggests that some organisms, like some machines, can be more easily changed and reconfigured than others – that is, that they are more evolvable. Last, it is possible to imagine ways that this same reconfigurability would also lead to many of the features that are loosely captured under the name of complexity.

Reviewing the literature on organismic evolvability highlights two main points. First, these are all very old ideas. They date back to the earliest work on taxonomy and development (albeit appearing in a different form before Darwin) and discussion of them has continued to the modern day. Evolvability is just a modern expression of these thoughts. Second, these ideas have always been dogged by a certain vagueness, which appears increasingly awkward when other parts of biological science have become more concrete. For instance, Kirschner and Gerhart's have recently proposed the valuable idea of a "core component", a part of the organism that facilitates change of other parts of the organism. Yet despite offering numerous examples they do not quite pin down this idea in a way that

makes it as clear as, say, the structure of DNA or the process of natural selection. This vagueness has fed a nagging, perennial concern about the coherence of evolvability as a scientific idea.

Most of the other concepts of evolvability have come from more recent efforts to formalise organismic evolvability, efforts which have oversimplified it and thus instead produced definitions of new but narrower concepts. These are the concepts of trait evolvability, individual fitness evolvability, and individual mutational evolvability, which are due to research in population genetics, molecular evolution, microbial evolution, and evolutionary computation. Trait evolvability is a population genetic measure of the responsiveness of a particular trait to selection. Individual fitness evolvability is a measure of statistical properties of the distribution of fitness in an individual's neighbourhood in genotype space. Individual mutational evolvability is a characterisation of the mutational structure of an individual's neighbourhood in genotype space. The review discusses all these in chapter 2.

Last, there is also the concept of substrate evolvability, the capacity of a physical or logical rule system – like the rules of chemistry, or of a simulated world – to support evolution of entities that are defined by those rules. This idea has connections with organismic evolvability but it mainly relates to foundational issues in evolutionary theory that appear most clearly in astrobiology and artificial life. In these fields this concept has mostly been implicit. This review is able to draw it out more, spelling out its conceptual foundations and relating it to existing ideas in computer science and the philosophy of science.

7.1.2 Key ideas and methodologies

Because the original concept of organismic evolvability has the deepest research history, this concept exposes the key ideas that pervade all discussions of evolvability – both the basic structure of the concept, and the typical criticisms of it. This review identifies the five key ideas in the evolvability literature. These five key ideas are the idea (1) that evolvability is a property of the structure of phenotypic variability, (2) that this is determined by an organism's developmental architecture, (3) that evolvability leaves specific traces in patterns of diversification in the macroevolutionary record, (4) that evolvability may play an important role in explaining the evolution of phenotypic complexity, and (5) that there may be no way that Darwinian selection can promote the evolution of evolvability. These ideas are first sketched out in a discussion of Dawkins's seminal biomorphs model, and

then discussed more fully in the context of the wider biological literature. This discussion appears in chapter 1.

The work which spawned the newer concepts of evolvability started from disciplines outside of traditional evolutionary theory and it introduced methods from these other discipline. These disciplines better support quantitative methods and they have tended to redefine evolvability to fit those methods. This broad perspective of this review makes it possible to show how these various redefinitions of evolvability and different research approaches can clarify these five key ideas. In this way, this review realises conceptual gains from connecting up the parts of a rather scattered discussion.

For one, research using the concept of individual fitness evolvability has proved that the evolution of evolvability is compatible with Darwinian selection. This research also sheds light on the key idea of variability, by providing the simplest models for demonstrating this idea. The concept of substrate evolvability shows intimate connections between artificial life and work in astrobiology, and it suggests informative connection between evolvability in organisms and deeper questions about the fundamental conditions necessary for evolution. But most importantly, research on individual mutational evolvability in the molecular evolution of RNA shapes has stimulated the development of new formal methods for analysing evolvability as a property of the genotype-phenotype map.

The work on RNA has noted the pervasive role of many-to-one associations, or degeneracy, in actual genotype-phenotype maps. This has led to pioneering statistical studies on the large-scale pattern of such associations, as well as other studies and theoretical work on the importance of mutationally connected sets of phenotypically equivalent genotypes, or so-called neutral networks. In general, this work has focused new attention on how underlying mutational relationships between genotypes determine the true structure of relationships between phenotypes. Although not all efforts to use this new awareness have been successful (work attempting to use topological spaces went somewhat astray), the basic concept is profound and it has generated valuable ideas such as the notion of an accessibility measure, a measure of the mutational connection between two phenotypes.

7.2 New formal methods

The work above provides the inspiration for the second half of this dissertation, which presents new formal methods for modelling and analysing evolvability. This dissertation builds on the quantitative methods developed to articulate these newer, narrower notions of individual evolvability. It then modifies and extends these methods to treat evolvability as a property of the organism, in order to restate the longstanding ideas of organismic evolvability in an exact fashion.

Mathematics can be irksome. But this formal approach is crucial for making progress on evolvability. A big problem with theories of evolvability has been that, in the end, they seem vague. This is hard to avoid when these are essentially theories about genetics and development, areas which for decades have been black boxes, studied only externally by their effects, and inaccessible to direct experimentation. In addition, many theories relate evolvability to complexity, another phenomenon which tantalises the intuition while defying exact formalisation.

My formal approach builds on one foundation stone, the genotype-phenotype map. While the idea of a genotype-phenotype map is a commonplace in mathematical discussions of evolution the idea is usually a casual conceit. This is natural since we do not understand any organism well enough to write down its genotype-phenotype mapping in full detail. But what if we did know the entire mapping in detail for some organism? RNA shapes almost present a case of this. Then how would we study the map? What tools would we use, for instance, to predict its evolvability? The fact is, those tools do not exist yet. The fact that detailed genotype-phenotype maps have not been available, from experiment, has concealed a failure in biology to conceptualise how one would analyse them, in theory. But even before detailed knowledge of maps becomes available it still clarifies our ideas to consider how we would analyse them. This is one way to understand the technical work of this dissertation: it is as an effort to develop these conceptual tools, to take the idea of a genotype-phenotype map seriously and push it as far as it will go.

7.2.1 Technical contribution

The technical contribution of this thesis can be divided into four parts: a critical analysis of work on topologies, two methodological innovations (genospace algebra and fulcrum-and-

leverage analysis), and a simple heuristic model of evolvability due to segment repetition and simple modularity.

The first part is the analysis of errors and important innovations in the literature on topological spaces. This is presented in chapter 3. An ambitious body of work tried to use the mathematical structure of a topological space to analyse genotype-phenotype maps in new ways, such as providing a formalisation for the idea of continuity in evolutionary change. This work uses genotype-phenotype maps to construct these topological spaces. Unfortunately, under closer examination, it turns out this construction procedure destroys so much information that these spaces are quite unsuitable for modelling purposes. However, this work did bring forward many useful ideas: the idea of a *phenotypic pre-image*, which is the set of genotypes which map onto a particular phenotype; the *accessibility measure*, a measure of the strength of mutational linkage between two phenotypes; and the *accessibility relation*, a cruder binary relation derived from the accessibility measure.

The second part technical part of this dissertation is an original methodological contribution, the invention of the *genospace algebra*. This is a formal language for describing genotype-phenotype maps. It is a propositional calculus based on graph theory, and it provides a system for making calculations, proofs, and diagrams about mutational structures in genotype space. This formalism is flexible enough to describe these structures at any degree of resolution. As a result, it can be used to represent mutational structures between phenotypes, between chosen interesting classes of phenotypes, between neutral networks, or between arbitrary subsets of genotypes. This is all presented in chapter 4.

The essential challenge in analysing a genotype-phenotype map is that it is a massive and complicated structure which is quite opaque to the traditional quantitative methods of examining variation one trait at a time or one gene at a time. It is necessary to come to grips with the overall structure of the mapping, the overall structure of degeneracy, the overall structure of epistatic dependency across different loci over the entire genome. The genospace algebra is a solution to this problem. It reveals broad logical structures in the mapping which correspond to recognisable concrete features of the developmental architecture.

The third part is also a methodological contribution. This is the invention of the measure of *genetic leverage* and the *genetic fulcrum*. While the genospace algebra allows exact qualitative descriptions of the structure of genotype space, genetic leverage ignores structure

and provides a single crude numerical measure of evolvability. It goes along with the idea of the genetic fulcrum, which provides a formalisation of the idea of a "core component", a basic notion in the concept of organismic evolvability. This is presented in chapter 5. This leverage-and-fulcrum scheme formalises the traditional arguments for organismic evolvability in a way that makes them susceptible to exact analysis.

Fourth, and last, these two analytical tools are developed in the context of two simple heuristic models. In general, evolvability is a property of an organism's developmental system that facilitates its long-term evolution. But evolvability in general always reduce to evolvability in particular, some concrete set of features of the developmental architecture. The simplest and perhaps most popular example of a developmental feature supporting evolvability is the capacity for repeated segmentation. If there were something like a "repeat loop" in an organism's development architecture, that would make certain kinds of changes much easier than if the organism had to re-evolve the genetics for every new segment. Repetition is the basic case of more complex examples of physical modularity. Chapters 4, 5, and 6 develop an exact model of this idea and extent it to describe more elaborate cases of modularity. This model provides an object for testing these new measures of evolvability, as well as an example of the kind of models that can be built to study the numerous other mechanisms which have been proposed as supportive of evolvability.

Bibliography

OED Online, chapter "organism". Oxford University Press, June 2008. URL http://dictionary.oed.com/cgi/entry/00333659.

C. Adami. What is complexity? *Bioessays*, 24(12):1085–1094, December 2002. ISSN 0265-9247. doi: 10.1002/bies.10192. URL http://dx.doi.org/10.1002/bies.10192.

C. Adami, C. Ofria, and T. C. Collier. Evolution of biological complexity. *Proceedings-National Academy of Sciences USA*, 97(9):4463–4468, 2000. 0027-8424 Fortnightly: 15-30 issues per year National Academy of Sciences; 1999 USA English.

Christopher Adami and C.T. Brown. Evolutionary learning in the 2d artificial life system "avida". In R Brooks and P Maes, editors, *Artificial Life IV*, page 377. MIT Press, July 1994.

Lee Altenberg. *Advances in Genetic Programming*, chapter The Evolution of Evolvability in Genetic Programming, pages 47–74. The M.I.T. Press, 1994.

Lee Altenberg. *Foundations of Genetic Algorithms*, chapter The schema theorem and Price's theorem, pages 23–49. MIT Press, 1995a.

Lee Altenberg. *Evolution and Biocomputation: Computational Models of Evolution*, chapter Genome growth and the evolution of the genotype-phenotype map, pages 205 – 259. Springer, 1995b.

Lauren W. Ancel and Walter Fontana. Plasticity, evolvability, and modularity in RNA. *Journal of Experimental Zoology*, 288(3):242–283, 2000. ISSN 1552-5015. doi: 10.1002/1097-010X(20001015)288:3⟨242::AID-JEZ5⟩3.0.CO;2-O. URL http://dx.doi.org/10.1002/1097-010X(20001015)288:3<242::AID-JEZ5>3.0.CO;2-O.

William Bains. Many chemistries could be used to build living systems. *Astrobiology*, 4(2): 137 – 167, 2004.

Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004. doi: doi:10.1038/nrg1272.

David Baulcombe. Dna events: An RNA microcosm. *Science*, 297(5589):2002 – 2003, September 2002. doi: 10.1126/science.1077906. URL http://www.sciencemag.org/cgi/reprint/297/5589/2002.pdf.

M. A. Bedau, E. Snyder, and N. H. Packard. A classification of long-term evolutionary dynamics. In C. Adami, editor, *Artificial life*, pages 228–237, Los Angeles; CA, June 1998. Cambridge Mass. ISBN 0262510995. International conference; 6th Theme title: Life and computation: the boundaries are changing. Also known as ALIFE VI ARTIFICIAL LIFE -PROCEEDINGS- 1998 CONF 6 MIT Press; 1998 English Papers.

Mark Bedau, Emile Snyder, and Titus Brown. A comparison of evolutionary activity in artificial evolving systems and in the biosphere. In P Husnabds and I Harvey, editors, *Proceedings of the European Conference on Artificial Life*, pages 125–134, 1997.

Mark A. Bedau and C. Titus Brown. Visualizing evolutionary activity of genotypes. *Artificial Life*, 5(1):17–36, 1999. CAMBRIDGE MA 1064-5462 Quarterly: 4 issues per year MASSACHUSETTS INSTITUTE OF TECHNOLOGY Great Britain English.

Mark A. Bedau and Norman H. Packard. Evolution of evolvability via adaptation of mutation rates. *Biosystems*, 69(2-3):143–162, May 2003. URL http://www.sciencedirect.com/science/article/B6T2K-483BXCF-2/2/5b44748dfb7130153f7dc1db5f58a852. TY - JOUR.

Z Belmandt. *Manuel de prétopologie et ses applications*. Hermes Sciences Publications, 1993. URL http://www.amazon.fr/exec/obidos/ASIN/2866013794/qid193958-8169748.

J. H. Bennet, editor. *Natural selection, heredity, and eugenics*. Oxford University Press, 1990. URL http://digital.library.adelaide.edu.au/coll/special//fisher/natsel/.

S. Bonnevay and C Largeron. Data analysis based on minimal closed subsets. In H.A.L. Kiers

et al., editors, *Data Analysis Classification and Related Methods*, pages 303–308. Springer, Berlin, 2000.

E. Bornberg-Bauer and H. S. Chan. Modeling evolutionary landscapes: mutational stability, topology, and superfunnels in sequence space. *Proc Natl Acad Sci U S A*, 96(19):10689–10694, September 1999. ISSN 0027-8424. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve\&db=pubmed\&dopt=Abstract\&list_uids=10485887.

Christine Bounama, Werner von Bloh, and Siegfried Franck. How rare is complex life in the milky way? *Astrobiology*, 7(5), 2007.

J. F. Brookfield. Evolution: the evolvability enigma. *Current Biology*, 11(3):R106–R108, February 2001. ISSN 0960-9822. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve\&db=pubmed\&dopt=Abstract\&list_uids=11231170.

David J Buller. *Function, Selection, and Design*, chapter Etiological Theories of Function, pages 281–307. State University of New York Press: Albany, 1999.

Christina L Burch and Lin Chao. Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature*, 406:625–628, August 2000.

John Cairns, Julie Overbaugh, and Stephan Miller. The origin of mutants. *Nature*, 335:142, September 1988.

Peter Calow. Constraints. In *The encyclopedia of ecology & environmental management*, pages xv, 805, [7] of plates. Blackwell Science, Oxford, 1997. ISBN 0865428387 0865428387 (pbk.). URL http://www.xreferplus.com/entry/1378869. editor-in-chief, Peter Calow. ill. (some col.); 29 cm.

E Čech. *Topological Spaces*. John Wiley, 1966.

A. Channon. Passing the alife test: Activity statistics classify evolution in geb as unbounded. In J. Kelemen and P. Sosik, editors, *European Conference on Artificial Life; Advances in artificial life*, pages 417–426, Prague, Czech Republic, September 2001. Springer. ISBN 3540425675. 6th Also known as ECAL 2001; Includes bibliographical references and index LECTURE NOTES IN COMPUTER SCIENCE 2001 VOL 2159 English Papers.

Ariel D. Chipman and Michael Akam. The segmentation cascade in the centipede strigamia maritima: Involvement of the notch pathway and pair-rule gene homologues. *Developmental Biology*, 319(1):160 – 169, 2008. ISSN 0012-1606. doi: DOI: 10.1016/j.ydbio.2008.02.038. URL http://www.sciencedirect.com/science/article/B6WDG-4S0208V-1/2/6b1ca753e4874475194cf064bd1f65b7.

James A Coffman and Eric H Davidson. *Encyclopedia of Life Sciences*, chapter Genetic Networks. John Wiley & Sons, Ltd: Chichester, April 2001.

M Conrad. Natural selection and the evolution of neutralism. *Biosystems*, 15(1):83–85, 1982. ISSN 0303-2647 (Print).

M Conrad. The geometry of evolution. *Biosystems*, 24(1):61–81, 1990.

Michael Conrad. Bootstrapping on the adaptive landscape. *Biosystems*, 11:167 – 182, 1979.

Jack B Copeland. The church-turing thesis. In Edward N Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, 2002. URL http://plato.stanford.edu/archives/fall2002/entries/church-turing.

Lindsay G Cowell, Hye Jung Kim, Tiina Humaljoki, Claudia Berek, and Kepler-Thomas-B. Enhanced evolvability in immunoglobulin v genes under somatic hypermutation. *Journal of Molecular Evolution*, 49 (1):23–26, 1999.

Francis Crick. *What Mad Pursuit*. Basic Books, 1988.

James P. Crutchfield and K Young. Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105 – 109, 1989.

J Cupal, S Kopp, and P. F. Stadler. RNA shape space topology. *Artificial Life*, 6(1):3–23, 2000.

Lewis Dartnell. *Life in the Universe: A Beginner's Guide*. Oneworld Publications, 2007.

Charles Darwin. *On the Origin of the Species*. 1859.

Richard Dawkins. *The Extended Phenotype*. Oxford University Press, 1982.

Richard Dawkins. *The Blind Watchmaker*. Longman Scientific & Technical, Harlow, 1986.

Richard Dawkins. Evolution of evolvability. In C Langton, editor, *Artificial Life*, pages 201–220. Addison-Wesley, 1989.

Richard Dawkins. *Climbing Mount Improbable*. W. W. Norton, 1996.

AM Denli, BB TOps, RH Plasterk, RF Ketting, and GJ Hannon. Processing of primary microRNAs by the microprocessor complex. *Nature*, 432(7014):231 – 235, 2004.

Daniel Dennett. *Encyclopedia of Evolution*, pages E83–E92. Oxford University Press, 2002.

David Deutsch. Quantum theory, the church-turing principle and the universal quantum computer. In *Proceedings of the Royal Society of London*, volume A400, pages 97 – 117, 1985.

Danjoy Dichtel, Marie Laure, and Marie-Anne Felix. Phenotypic neighborhood and micro-evolvability. *Trends in Genetics*, 20(5):268–276, 2004.

Peter Dittrich, Jens Ziegler, and Wolfgang Banzhaf. Artificial chemistries: A review. *Artificial Life*, 7(3):225–275, [print] Summer, 2001 2001. Article- English.

Keith Douglas. Super-turing computation: A case study analysis. Master's thesis, Carnegie Mellon University, 2003.

SR Eddy. Computational genomics of noncoding RNA genes. *Cell*, 109(2):137–140, April 2002.

Brian D. Farrell. "inordinate fondness" explained: Why are there so many beetles? *Science*, 281(5376):555–559, 1998. doi: 10.1126/science.281.5376.555. URL http://www.sciencemag.org/cgi/content/abstract/281/5376/555.

R. A Fisher. Average excess and average effect of a gene substitution. *Annals of Eugenics*, 11, 1941.

W. Fontana. Modelling 'evo-devo' with RNA. *Bioessays*, 24(12):1164–1177, December 2002. ISSN 0265-9247. doi: 10.1002/bies.10190. URL http://www3.interscience.wiley.com/cgi-bin/abstract/101019953/ABSTRACT.

W. Fontana, D. A. Konings, P. F. Stadler, and P. Schuster. Statistics of RNA secondary structures. *Biopolymers*, 33(9):1389–1404, September 1993. doi: 10.1002/bip.360330909. URL http://dx.doi.org/10.1002/bip.360330909.

Walter Fontana and Peter Schuster. Continuity in evolution: One the nature of transitions. *Science*, 280(536):1451–1455, May 1998a. Inst. Theor. Chem., Univ. Wien, Waehringerstr. 17, A-1090 Wien, Austria May 29.

Walter Fontana and Peter Schuster. Shaping space: the possible and the attainable in RNA genotype-phenotype mapping. *Journal of Theoretical Biology*, 194(4):491–515, October 1998b. URL http://www.sciencedirect.com/science/article/B6WMD-45KNDB5-3/2/2ae6631ab8217739abca410979a729b1. TY - JOUR.

Sidney W Fox. Simulated natural experiments in spontaneous organization of morphological units from protenoid. In Sidney W Fox, editor, *The Origins of Prebiological Systems and Their Molecular Matrices*. Academic Press, 1965.

M Fréchet. *Espaces abstraits*. Hermann, Paris, 1928.

G Fusco and A Minelli. Measuring morphological complexity of segmented animals: Centipedes as model systems. *Journal of Evolutionary Biology*, 13 (1):38–46., 2000.

Sergey Gavrilets. *Fitness Landscapes and the Origin of Species*. Monographs in population biology. Princeton University Press, 2004.

M Gell-Mann. What is complexity? *Complexity*, 1:16–19, 1995.

Murray Gell-Mann and Seth Lloyd. Information measures, effective complexity, and total information. *Complexity*, 2(1):44–52, 1996.

John Gerhart and Marc Kirschner. *Cells, Embryos And Evolution: Toward a Cellular & Developmental Understanding of Phenotypic Variation & Evolutionary Adaptability*. Blackwell Publishing, 1997.

John Gerhart and Marc Kirschner. Colloquium papers: The theory of facilitated variation. *PNAS*, 104(1):8582–8589, 2007. doi: 10.1073/pnas.0701035104. URL http://www.pnas.org/cgi/content/abstract/104/suppl_1/8582.

Ulrike Göbel. *Neutral Networks of Minimum Free Energy RNA Secondary Structures*. Phd, University of Vienna, Vienna, 2000. URL http://www.tbi.univie.ac.at/papers/Abstracts/goebel_diss.ps.gz.

Stephen J Gould and Richard C Lewontin. The spandrels of san marco and the panglossian paradigm: a critique of the adaptationist programme. In *Proceeds of the Royal Society of London*, 1979.

Stephen Jay Gould. *The Structure of Evolutionary Theory*. Belknap Press, March 2002.

P Grassberger. Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys*, 25:907–928, 1986.

W Gruner, R Giegerich, D Strothmann, C Reidys, J Weber, Ivo L Hofacker, Peter F Stadler, and Peter Schuster. Analysis of RNA sequence structure maps by exhaustive enumeration.1. neutral networks. *Monatshefte fur Chemie*, 127(4):355–374, April 1996a. ISSN 0026-9247.

W Gruner, R Giegerich, D Strothmann, C Reidys, J Weber, Ivo L Hofacker, Peter F Stadler, and Peter Schuster. Analysis of RNA sequence structure maps by exhaustive enumeration.2. structures of neutral networks and shape space covering. *Monatshefte fur Chemie*, 127(4):375–389, April 1996b. ISSN 0026-9247.

T. F. Hansen and Günter P. Wagner. Modeling genetic architecture: a multilinear theory of gene interaction. *Theor Popul Biol*, 59(1):61–86, February 2001. ISSN 0040-5809. doi: 10.1006/tpbi.2000.1508. URL http://dx.doi.org/10.1006/tpbi.2000.1508.

Thomas F. Hansen. Is modularity necessary for evolvability? remarks on the relationship between pleiotropy and evolvability. *Biosystems*, 69(2-3):83–94, May 2003. URL http://www.sciencedirect.com/science/article/B6T2K-483BXCF-3/2/f4cad930efbc63f1253f051d412e8e7d. TY - JOUR.

L. He, R Kierzek, J SantaLucia, A E Walter, and D. H. Turner. Nearest-neighbor parameters for g-u mismatches. *Biochemistry*, 30:11124, 1991.

Rolf Herken. *The Universal Turing machine: a half-century survey*. Oxford University Press, 1988.

Ivo Hofacker. *Encyclopedia of Life Sciences*, chapter RNA Secondary Structure Prediction. John Wiley & Sons, Ltd: Chichester, September 2005.

D. Houle. Comparing evolvability and variability of quantitative traits. *Genetics*, 130(1): 195–204, 1992. URL http://www.genetics.org/cgi/content/abstract/130/1/195.

Paul Humphreys. Emergence. In *The Encyclopedia of Philosophy*. MacMillan, 2006.

Martijn A. Huynen, Peter F. Stadler, and Walter Fontana. Smoothness within ruggedness: The role of neutrality in adaptation. *Proceedings of the National Academy of Sciences of*

the United States of America, 93(1):397–401, 1996. Inst. Theoretische Chemie, Wahringer Strasse 17, A-1090 Vienna, Austria.

J. A. Jaeger, D. H. Turner, and M. Zuker. Improved predictions of secundary structures for RNA. *Proc Natl Acad Sci USA*, 86:7706–7710, 1989.

Stuart Kauffman. *The origins of order: self-organization and selection in evolution*. Oxford University Press, 1993.

Marc Kirschner and John Gerhart. Evolvability. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15):8420–8427, 1998. Dep. Cell Biol., Harvard Med. Sch., Boston, MA 02115 July 21.

Marc Kirschner and John Gerhart. *The Plausibility of Life*. Yale University Press, 2005.

M Kozak. Extensively overlapping reading frames in a second mammalian gene. *EMBO Rep*, 2(9):768–769, 2001. ISSN 1469-221X (Print). doi: 10.1093/embo-reports/kve183.

K Kuratowski. Sur le problème des courbes gauches en topologie. *Fund. Math*, 15:271–283, 1930.

Christian R. Landry, Bernardo Lemos, Scott A. Rifkin, W. J. Dickinson, and Daniel L. Hartl. Genetic properties influencing the evolvability of gene expression. *Science*, page 1140247, 2007. doi: 10.1126/science.1140247. URL http://www.sciencemag.org/cgi/content/abstract/1140247v1.

Christine Largeron and Stephane Bonnevay. A pretopological approach for structural analysis. *Information Sciences*, 144(1-4):169–185, July 2002. doi: doi:10.1016/S0020-0255(02)00189-5. URL http://www.sciencedirect.com/science/article/B6V0C-45R7TW7-2/2/6b5b3ef613c7163d8fca9f780984aa81.

R. E. Lenski, C. Ofria, R. T. Pennock, and C. Adami. The evolutionary origin of complex features. *Nature*, 423(6936):139–144, May 2003. ISSN 0028-0836. doi: 10.1038/nature01568. URL http://myxo.css.msu.edu/papers/nature2003/Nature03_Complex.pdf.

Richard E. Lenski and John E. Mittler. The directed mutation controversy and neo-darwinism. *Science*, 259(5092):188–194, 1993. ISSN 00368075. URL http://www.jstor.org/stable/2880562.

R. C. Lewontin. Adaptation. *Scientific American*, 239(3):212–222, September 1978.

Thomas Lufkin. *Encyclopedia of Life Sciences*, chapter Hox Genes: Embryonic Development. John Wiley & Sons, Ltd: Chichester, September 2005.

Michael Lynch, M. Pfrender, K Spitze, N Lehman, J Hicks, D Allen, L Latta, M Ottene, F Bogue, and J Colbourne. The quantitative and molecular genetic architecture of a subdivided species. *Evolution*, 53(1):100–110, February 1999. URL http://tinyurl.com/25vneh.

T. F. Mackay. The genetic architecture of quantitative traits. *Annu Rev Genet*, 35:303–339, 2001. doi: 10.1146/annurev.genet.35.102401.090633. URL http://dx.doi.org/10.1146/annurev.genet.35.102401.090633.

James B Maclaurin. *Encyclopedia of Life Sciences*, chapter Fitness: Philosophical Problems. John Wiley & Sons, Ltd: Chichester, March 2001. doi: doi:10.1038/npg.els.0003443.

Tomoaki Matsuura, Muneyoshi Yamaguchi, Ko-Mitamura-Elizabeth-P, Yasufumi Shima, Itaru Urabe, and Tetsuya Yomo. Importance of compartment formation for a self-encoding system. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 99 (11), pages 7514–7517., 2002.

John Maynard-Smith and Eors Szathmary. *The Major Transitions in Evolution*. W.H. Freeman and Company Limited, 1995.

John Maynard-Smith, R. Burian, S. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, and L. Wolpert. Developmental constraints and evolution: A perspective from the mountain lake conference on development and evolution. *The Quarterly Review of Biology*, 60(3):265–287, 1985.

Mark A McPeek and Jonathan M Brown. Clade age and not diversification rate explains species richness among animal taxa. *American Naturalist*, 169(4), 2007.

Daniel W. McShea. Metazoan complexity and evolution: is there a trend? *Evolution*, 50: 477–492, 1996.

A. Meziane, Tahar Iftene, and Nazha Selmaoui. Satellite image segmentation by mathematical pretopology and automatic classification. In Jack Desachy and Sharam Tajbakhsh,

editors, *Image Processing, Signal Processing, and Synthetic Aperture Radar for Remote Sensing*, volume 3217, pages 232–236, London, United Kingdom, 1997. SPIE. URL http://link.aip.org/link/?PSI/3217/232/1.

S. L. Miller. Production of amino acids under possible primitive earth conditions. *Science*, 117(3046):528 – 529, May 1953.

Melani Mitchell and Charles Taylor. Evolutionary computation: an overview. *Annu. Rev. Ecol. Syst.*, 30:593–616, 1999.

Ernst Nagel. *The Structure of Science*. Harcourt, Brace and Company: Burlingame, California, 1961.

R Nussinov and A. B. Jacobson. Fast algorithms for predicting the secondary structure of single-stranded RNA. *Proc Natl Acad Sci USA*, 77:6309–6313, 1980.

OED. *OED Online*, chapter "module, *n.*". Oxford University Press, September 2002a. URL http://dictionary.oed.com/cgi/entry/00313178.

OED. *OED Online*, chapter "modularity, *n.*". Oxford University Press, September 2002b. URL http://dictionary.oed.com/cgi/entry/00313178.

C. Ofria, C. Adami, and T. C. Collier. Design of evolvable computer languages. *Ieee Transactions on Evolutionary Computation*, 6:420–424, Numb 4 2002. 1089-778X IEEE USA English.

Charles Ofria and Claus O. Wilke. Avida: a software platform for research in computational evolutionary biology. *Artif. Life*, 10(2):191–229, 2004. ISSN 1064-5462. doi: http://dx.doi.org/10.1162/106454604773563612.

Charles Ofria, C.T. Brown, and C Adami. *Introduction to Artificial Life*, chapter The avida user's manual. Springer, 1988.

Piero Pagliani. Pretopologies and dynamic spaces. *Fundam. Inf.*, 59(2-3):221–239, 2004. ISSN 0169-2968. URL http://iospress.metapress.com/openurl.asp?genre=article&issn=0169-2968&volume=59&issue=2&spage=221.

C. Pál. Yeast prions and evolvability. *Trends Genet*, 17(4):167–169, April 2001. ISSN 0168-9525. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve\&db=pubmed\&dopt=Abstract\&list_uids=11275304.

A. N. Pargellis. Digital life in the amoeba world. In *Artificial Life*, volume 7, pages 63–75. MIT Press, March 2001.

Andrew N. Pargellis. The spontaneous generation of digital "life". *Phys. D*, 91(1-2):86–96, 1996a. ISSN 0167-2789. doi: http://dx.doi.org/10.1016/0167-2789(95)00268-5.

Andrew N. Pargellis. The evolution of self-replicating computer organisms. *Phys. D*, 98(1): 111–127, 1996b. ISSN 0167-2789. doi: http://dx.doi.org/10.1016/0167-2789(96)00089-9.

John W. Pepper. The evolution of evolvability in genetic linkage patterns. *Biosystems*, 69(2-3):115–126, May 2003. URL http://www.sciencedirect.com/science/article/B6T2K-4851DSN-2/2/a58800ea5126eab6a38c4b80b2ec7556. TY - JOUR.

Benjamin C Pierce. *Types and Programming Languages*. MIT Press, 2002.

Mircea Podar, Steven H. D. Haddock, Mitchell L. Sogin, and G. Richard Harbison. A molecular phylogenetic framework for the phylum ctenophora using 18s rRNA genes. *Molecular Phylogenetics and Evolution*, 21(2):218–230, 2001.

Anthony M. Poole, Matthew J. Phillips, and David Penny. Prokaryote and eukaryote evolvability. *Biosystems*, 69(2-3):163–185, May 2003. URL http://www.sciencedirect.com/science/article/B6T2K-48413XY-4/2/cfaad58a7b9de5b3c95a4246b556d82d. TY - JOUR.

W Provine. *Sewall Wright and evolutionary biology*. University of Chicago Press, 1986.

S. Rasmussen. Aspects of Information, Life, Reality, and Physics. *Artificial Life II*, pages 767–773, 1992.

S Rasmussen, C Knudsen, R Feldberg, and M Hindsholm. The coreworld – emergence and evolution of cooperative structures in a computational chemistry. *Physica D*, 42(1-3):111, June 1990.

Thomas Ray. An approach to the synthesis of life. In C.G. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 371–408. Addison-Wesley, 1992. URL http://www.talkorigins.org/faqs/tierra.html.

I Rechenberg. *Evolutionstrategie*. Frommann-Holzboog, 1973.

Mark Rendel. *The evolutionary dynamics of neutral networks: Lessons from RNA*. PhD thesis, University of Oxford, 2008.

Jean-Philippe Rennard. *Recent Developments in Biologically Inspired Computing*, chapter Chapter XII: Perspectives for Strong Artificial Life, pages 301–320. Idea Group Publishing, August 2004.

Rupert Riedl. *Order in Living Organisms*. Wiley, 1978.

L. H. Rieseberg, J. Whitton, and K. Gardner. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, 152(2):713–727, Jun 1999.

E Rivas and SR Eddy. Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics*, 16(7):583–605, 2000.

Alexander Rosenberg and Frederic Bouchard. Fitness. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, Winter 2002. URL http://plato.stanford.edu/archives/win2002/entries/fitness/.

David Rowe, John Leaney, and David Lowe. Defining systems evolvability – a taxonomy of change. In *Proceeding of the IEEE Conference on Computer Based Systems*, pages 45–52. IEEE Press, 1998.

Gary Ruvkun. Molecular biology: Glimpses of a tiny RNA world. *Science*, 294(5543):797 – 799, October 2001. doi: 10.1126/science.1066315. URL http://www.sciencemag.org/cgi/reprint/294/5543/797.pdf.

Eric Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, 1997. URL http://www.math.vanderbilt.edu/~schectex/ccc/.

P Schuster and W Fontana. Chance and necessity in evolution: lessons from RNA. *Physica D-Nonlinear Phenomena*, 133(1-4):427–452, September 1999. ISSN 0167-2789.

P Schuster, W Fontana, P. F. Stadler, and Ivo Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc Roy Soc (London)B*, 255:279–284, 1994.

G. G. Simpson. *Tempo and Mode in Evolution*. Columbia University Press, New York, 1944.

Tom Smith, Phil Husbands, Paul Layzell, and Shea Michael O. Fitness landscapes and evolvability. *Evolutionary Computation*, 10(1):1–34, 2002. 1063-6560 Quarterly: 4 issues per year MASSACHUSETTS INSTITUTE OF TECHNOLOGY USA English.

Tom Smith, Phil Husbands, and Michael O'Shea. Local evolvability of statistically neutral gasnet robot controllers. *Biosystems*, 69(2-3):223–243, May 2003. URL http://www.sciencedirect.com/science/article/B6T2K-48413XY-2/2/d77e4a85314fc5a796da572c30918cba. TY - JOUR.

Elliott Sober. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. MIT Press, 1984.

Lee Spector and Alan Robinson. Genetic programming and autoconstructive evolution with the push programming language. *Genetic Programming and Evolvable Machines*, 3(1): 7–40, 2002. ISSN 1389-2576. doi: http://dx.doi.org/10.1023/A:1014538503543.

Barbel M. R. Stadler and Peter F. Stadler. Generalized topological spaces in evolutionary theory and combinatorial chemistry. *J. Chem. Inf. Comput. Sci.*, 42:577–585, 2002. [ASAP Article 10.1021/ci0100898 S0095-2338(01)00089-0 Web Release Date: January 29, 2002].

Barbel M. R. Stadler, Peter F. Stadler, Günter P. Wagner, and Walter Fontana. The topology of the possible: Formal spaces underlying patterns of evolutionary change. *Journal of Theoretical Biology*, 213(2):241–274, November 2001. URL http://www.sciencedirect.com/science/article/B6WMD-457VCYN-1H/2/cb9848580ebc9065719152a8b87cb230. TY - JOUR.

Alison A Staton and Antonio J Giraldez. *Encyclopedia of Life Sciences*, chapter MicroRNAs in Development and Disease. John Wiley & Sons, Ltd: Chichester, July 2008.

Lynn Arthur Steen and J. Arthur Seebach Seebach. *Counterexamples in Topology*. Springer-Verlag: New York, 1978.

Bing Su and Rui Zhang. *Encyclopedia of Life Sciences*, chapter MicroRNA Evolution in the Human Genome. John Wiley & Sons, Ltd: Chichester, March 2008.

Nihon Sugakkai. *Encyclopedic Dictionary of Mathematics*. MIT Press, 2nd edition, 1993. ISBN 0262590204.

H Suzuki, N Ono, and K Yuta. Several necessary conditions for the evolution of complex forms of life in an artificial environment. In *Artificial Life*, volume 9, pages 153–174. MIT Press, Spring 2003.

Naoyuki Takahata. *Encyclopedia of Life Sciences*, chapter Molecular Evolutioni: Neutral Theory. John Wiley & Sons, Ltd: Chichester, December 2007.

Ella Tour and William McGinnis. *Encyclopedia of Life Sciences*, chapter Hox Genes and Body Plan: Evolution. John Wiley & Sons, Ltd: Chichester, July 2006. doi: doi:10.1038/npg.els.0005113.

Heather L True and Susan L Lindquist. Computation: evolutionary, neural, molecular. *Nature*, 407(6803):477–483, September 2000. ISSN 0028-0836. URL http://dx.doi.org/10.1038/35035005.

Alan Turing. On computable numbers, with an application to the entscheidungsproblem. In *Proceedings of the London Mathematical Society*, pages 230 – 265, 1936.

D.H. Turner, N. Sugimoto, and S. Freier. RNA structure prediction. *Ann Rev Biophys Biophyc Chem*, 17:167–192, 1988.

Peter D. Turney. Increasing evolvability considered as a large-scale trend in evolution. In *Proceedings of the 1999 Genetic and Evolutionary Computation Conference (GECCO-99) Workshop Program*, pages 43–46, July 1999.

Mark C Ungerer and Loren H Rieseberg. *Encyclopedia of Life Sciences*, chapter Speciation: Genetics. John Wiley & Sons, Ltd: Chichester, 2005.

E. van Nimwegen, J. P. Crutchfield, and M. Huynen. Neutral evolution of mutational robustness. *Proc Natl Acad Sci U S A*, 96(17):9716–9720, August 1999. ISSN 0027-8424. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve\&db=pubmed\&dopt=Abstract\&list_uids=10449760.

George von Dassow and Ed Munro. Modularity in animal development and evolution: Elements of a conceptual framework for evodevo. *The Journal of Experimental Zoology*,

285(4):307–325, December 1999. doi: DOI:10.1002/(SICI)1097-010X(19991215)285:4⟨307::AID-JEZ2⟩3.0.CO;2-V.

Andreas Wagner. Robustness, evolvability, and neutrality. *FEBS Letters*, 579(8):1772–1778, 2005.

Günter P. Wagner. Feedback selection and the evolution of modifiers. *Acta Biotheoretica (Historical Archive)*, 30(2):79–102, June 1981. doi: DOI:10.1007/BF00047674. URL http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1007/BF00047674.

Günter P. Wagner and Lee Altenberg. Complex adaptations and the evolution of evolvability. *Evolution*, 50(3):967–976, 1996. Literature-Review Cent. Computational Ecol., Dep. Biol., Yale Univ., New Haven, CT 06511.

Günter P. Wagner and Manfred D. Laubichler. Rupert Riedl and the re-synthesis of evolutionary and developmental biology: Body plans and evolvability. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 302B(1):92–102, 2004. URL http://dx.doi.org/10.1002/jez.b.20005.

Günter P. Wagner and Peter F. Stadler. Quasi-independence, homology and the unity of type: A topological theory of characters. *Journal of Theoretical Biology*, 220(4):505–527, Feburary 2003. URL http://www.sciencedirect.com/science/article/B6WMD-481MJN8-5/2/a59daca32badf8ad672d089bf724b025. TY - JOUR.

Chip Walter. Digital darwin. *MIT Technology Review*, 1:1, November 2003. URL http://www.technologyreview.com/articles/03/11/innovation31103.asp.

P. D. Ward and D Brownlee. *Rare Earth*. Copernicus, 1999.

M. S. Waterman and T. F. Smith. RNA secondary structure: A complete mathematical analysis. *Mathematical Biosciences*, 42:257–266, 1978.

Richard Watson. *Compositional Evolution*. MIT Press, 2006.

J Weber. *Dynamics of Neutral Evolution – A case study on RNA secondary structures*. PhD thesis, Friedrich Schiller Universitat, Jena, Germany, Jena, Germany, 1997.

Kenneth M Weiss. *Encyclopedia of Life Sciences*, chapter Allele Spectrum of Human Genetic Disease. John Wiley & Sons, Ltd: Chichester, January 2006. doi: DOI:10.1038/npg.els.0006033.

Eric W Weisstein. *CRC Concise Encyclopedia of Mathematics*. Chapman & Hall/CRC, 2nd edition, 2002.

M J West-Eberhard. Evolution in the light of developmental and cell biology, and vice versa. *Proc Natl Acad Sci U S A*, 95(15):8417–8419, 1998. ISSN 0027-8424 (Print).

C Wilke. Selection for fitness vs. selection for robustness in RNA secondary structure folding. *Evolution*, 55:2412–2420, 2001. doi: DOI:10.1043/0014-3820(2001)055⟨2412:SFFVSF⟩2.0.CO;2.

Claus O. Wilke and Christoph Adami. The biology of digital organisms. *Trends in Ecology & Evolution*, 17(11):528–532, November 2002. URL http://www.sciencedirect.com/science/article/B6VJ1-46NX8Y4-1/2/c8e790dec9151ac529d95f825a6a48cc. TY - JOUR.

Stephen Willard. *General Topology*. Addison-Wesley, 1970.

George C Williams. *Natural Selection: Domains, Levels, and Challenges*. Oxford Series in Ecology and Evolution. Oxford University Press, 1992.

James Maurice Wilson. *A lecture on some properties and peculiarities of water: a chapter in natural theology*. Whittaker, 1881.

Stephen Wolfram. *A New Kind of Science*. Wolfram Media Inc., 2002.

Troy E Wood and Loren H Rieseberg. *Encyclopedia of Life Sciences*, chapter Speciation: Introduction. John Wiley & Sons, Ltd: Chichester, 2005.

David S Woodruff. Declines of biomes and biotas and the future of evolution. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 98 (10), pages 5471–5476., 2001.

S Wright. Evolution in mendelian populations. *Genetics*, 16:97–159, 1931.

Asao Yamauchi, Toshihiro Nakashima, Nobuhiko Tokuriki, Masato Hosokawa, Hideki Nogami, Shingo Arioka, Itaru Urabe, and Tetsuya Yomo. Evolvability of random polypeptides through functional selection within a small library. *Protein Engineering Design & Selection*, 15(7):619–626, 2002. URL http://peds.oxfordjournals.org/cgi/content/abstract/15/7/619.

Andrew S. Yang. Modularity, evolvability, and adaptive radiations: a comparison of the hemi- and holometabolous insects. *Evolution & Development*, 3(2):59–72, 2001. URL http://www.blackwell-synergy.com/doi/abs/10.1046/j.1525-142x.2001.003002059.x.

ZHAO-BANG Zeng, CHEN-HUNG Kao, and CHRISTOPHER J. Basten. Estimating the genetic architecture of quantitative traits. *Genetical Research*, 74:279–289, 1999. doi: doi: 10.1017/S0016672399004255.

M Zuker and D Sankoff. RNA secondary structures and their prediction. *Bull Math Biol*, 46: 591–621, 1984.

M Zuker and P Stiegler. Optimal computer folding of larger RNA sequences using thermodynamics and auxilliary information. *Nucleic Acids Res*, 9:133–148, 1981.