

UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

ELECTRONICS AND COMPUTER SCIENCE

Volume 1 of 1

**On The Interaction Of Function, Constraint And Complexity In
Evolutionary Systems**

by

Adam Paul Davies

Thesis for the degree of Doctor of Philosophy

April 2014

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Doctor of Philosophy

ON THE INTERACTION OF FUNCTION, CONSTRAINT AND COMPLEXITY IN EVOLUTIONARY SYSTEMS

by Adam Paul Davies

Biological evolution contains a general trend of increasing complexity of the most complex organisms. But artificial evolution experiments based on the mechanisms described in the current theory generally fail to reproduce this trend; instead, they commonly show systematic trends of complexity minimisation. In this dissertation we seek evolutionary mechanisms that can explain these apparently conflicting observations. To achieve this we use a reverse engineering approach by building computational simulations of evolution. One highlighted problem is that even if complexity is beneficial, evolutionary simulations struggle with apparent roadblocks that prevent them from scaling to complexity. Another is that even without roadblocks, it is not clear what drives evolution to become more complex at all. With respect to the former, a key roadblock is how to evolve 'irreducibly complex' or 'non-decomposable' functions. Evidence from biological evolution suggests a common way to achieve this is by combining existing functions – termed 'tinkering' or 'building block evolution'. But in simulation this approach generally fails to scale across multiple levels of organisation in a recursive manner. We provide a model that identifies the problem hindering recursive evolution as increasing 'burden' in the form of 'internal selection' as joined functions become more complex. We show how having an ontological development process that occurs by local growth, as present in most complex biological organisms, resolves this problem, enabling evolution to occur recursively. Meanwhile, to understand what drives complexity in evolution we

provide a model showing that under certain conditions a well-studied concept from the computational study of algorithms – complexity lower bounds – applies in evolution. The model shows how the ‘difference’ between the conditions required by an organism’s replicator and its external environment results in a minimum complexity floor that varies as the external environment changes. We find that selection in such a system produces a system-wide, overall trend of increasing complexity of the most complex organisms (as environments are colonised), coupled with local trends of complexity minimisation in individual environments (as evolution seeks to minimise its cost of resources) –thereby resolving the tension between biological observations and theoretical outcomes. Our simulations and analytic results demonstrate (a) how evolution can, when complexity is beneficial, scale to complexity over multiple organisational levels, and (b) the conditions in which complexity is beneficial in evolution. These models describe a set of phenotypic, ontogenetic and environmental conditions that are generally present in biological evolution, in which evolution consistently generates an overall trend of increasing complexity of the most complex organisms.

Table of Contents

<u>ABSTRACT</u>	i
Table of Contents	i
List of tables	v
List of figures	vii
DECLARATION OF AUTHORSHIP	xi
Acknowledgements	xiii
Chapter 1: Introduction and literature review	1
1.1 Motivation	1
1.2 Theoretical perspective.....	5
1.3 Approach and previous work	7
1.3.1 Complexity roadblocks.....	9
1.3.2 Complexity drivers	18
1.3.3 Complexity lower bounds.....	21
1.3.4 The origins of complexity lower bounds in evolution	23
1.3.5 Homeogenesis.....	26
1.4 Summary of major topics of this dissertation	31
1.5 Implications.....	32
1.6 Models and approach	34
1.6.1 Modelling approach.....	34
1.6.2 Model 1: Combinatorial Exaptation – exaptation by joining functions	35
1.6.3 Model 2: Metabolic evolution by homeogenesis	36
1.6.4 Model 3: Complexity trends of evolution in a system of complexity lower bounds	38
1.6.5 Model 4: Complexity lower bounds in NAND circuit calculations	39

1.7	Claims	40
1.8	Contributions	41
1.9	Scope	42
1.10	Structure of this dissertation	43
Chapter 2:	Combinatorial Exaptation	45
2.1	Introduction	45
2.2	Previous work	47
2.2.1	Exaptation	47
2.2.2	Tinkering	48
2.2.3	Building Block Mechanisms	48
2.2.4	Recursive evolution and encapsulation	51
2.3	Theoretical analysis	53
2.4	Results and Discussion	58
2.4.1	S1 Simulation results	61
2.4.2	C1 simulation results	65
2.4.3	A recursive mechanism of combinatorial exaptation	66
2.4.4	S2 Simulation results	71
2.4.5	The effect of genotype-phenotype map evolution on the fitness landscape	74
2.4.6	Supplementary results	76
2.5	Conclusions	78
2.5.1	Key Results	79
2.6	Methods	80
Chapter 3:	Homeogenesis	87
3.1	Introduction	87
3.2	Background	87
3.3	An extreme and simple example of homeogenesis	89

3.4 A biological example of homeogenesis	92
3.5 A model of metabolic evolution	94
3.5.1 Aim	94
3.5.2 Methods	95
3.5.3 Results and discussion	99
3.5.4 Wider implications of homeogenesis	106
3.6 Conclusions.....	107
3.6.1 Key Results	109
Chapter 4: Complexity lower bounds.....	111
4.1 Introduction	111
4.2 Structure of this chapter	112
4.3 Generating a hypothesis for a complexity trend generation mechanism.....	113
4.3.1 Exploratory modelling	113
4.3.2 Summary of exploratory modelling.....	125
4.4 A hypothesis for a novel complexity generation mechanism in evolution: Environmental dissociation complexity	126
4.4.1 The Environmental Dissociation Hypothesis.....	130
4.5 Testing the environmental dissociation complexity hypothesis by computational modelling	131
4.5.1 Methods	131
4.5.2 Results	132
4.6 Discussion.....	144
4.6.1 Environmental dissociation and niche construction.....	149
4.6.2 Average path length	150
4.7 Evidence for complexity lower bounds in other models	151
4.7.1 Methods	151

4.7.2 Results	153
4.8 Conclusions	155
4.8.1 Key Results	159
Chapter 5: Summary and Conclusions.....	161
5.1 Research aims	161
5.2 Summary of research findings	162
5.3 Contributions and Implications	165
5.4 Limitations	167
5.5 Further work.....	168
5.6 Concluding remarks	169
List of References	171

List of tables

Table 1. Description of two separate types of exaptation	55
Table 2. Description of selection pressures in the combinatorial exaptation model.....	84

List of figures

Figure 1. A figure illustration of the range of biological combinatorial exaptation mechanisms.....	29
Figure 2. An example of evolving an non-decomposable function (a metabolic pathway) by combinatorial exaptation	57
Figure 3. Genotype-phenotype map for S1 simulation in the combinatorial exaptation model	59
Figure 4. Rules for enzymes interacting in functional interaction space in the combinatorial exaptation model	60
Figure 5. Diagram of the 21 available phenotypic functions in the combinatorial exaptation model	61
Figure 6. Case study results for S1 and S2 simulations of the combinatorial exaptation model	62
Figure 7. Figure illustration of two types of transition (Type-1 and Type-2) possible by combinatorial exaptation	67
Figure 8. Comparison of two approaches to evolving a Type 2 transition	70
Figure 9. Genotype-phenotype map of organisms in the S2 simulation of the combinatorial exaptation model.....	72
Figure 10. Comparison of the distribution of neutral or fitter genotypes in the fitness landscape given different types of genotype-phenotype map in the combinatorial exaptation model.....	75
Figure 11. Plots showing the sensitivity of the S2 simulation of the combinatorial exaptation model to simulation parameters.....	77
Figure 12. Figure illustration of the C ₃ and C ₄ photosynthetic cycles	93
Figure 13. Figure illustration of C ₄ photosynthesis evolution as a process of environmental Internalisation	94
Figure 14. Figure illustration of the chemical reaction network abstraction in the homeogenesis model	95
Figure 15. Description of the genotype-phenotype map in the homeogenesis model.....	97
Figure 16. Figure illustration of chemical environment change in the homeogenesis model.	98
Figure 17. Figure illustration of the sequence of chemical environments used in Simulation 1 of the homeogenesis model	100

Figure 18. Internalised environment frequency results from Simulation 1 of the homoeogenesis model	101
Figure 19. Population size results from Simulation 1 of the homoeogenesis model.....	102
Figure 20. Case study showing internalised environment frequency results from Simulation 1 of the homoeogenesis model.....	103
Figure 21. Simulation 2 results of the homeogenesis model: Comparison of the frequency of internalised sequences of ancestral environments given serial or parallel sequences of environment change.....	105
Figure 22. Figure illustration comparing homeogenesis with other processes of environment maintenance in evolution.....	107
Figure 23. Case study results for Exploratory test 1 of the complexity lower bounds model	116
Figure 24. Schematic illustration of the differences in number of chemical steps required to convert glucose \Rightarrow pyruvate and glucose \Rightarrow glucose 6 phosphate	119
Figure 25. Case study results for Exploratory test 2 of the complexity lower bounds model	121
Figure 26. Case study results for Exploratory test 3 in the complexity lower bounds model	123
Figure 27. Case study results for Exploratory test 4 in the complexity lower bounds model	124
Figure 28. Figure illustration of solving a Rubik's cube.	127
Figure 29. Figure illustration of the state space of a Rubik's cube.	128
Figure 30. Numerical analysis of the state space of a Rubik's cube.....	128
Figure 31. Metabolic complexity case study results of the positive control experiment (<i>without</i> environmental perturbation) of the complexity lower bounds model.....	133
Figure 32. Metabolic complexity case study results of the positive control experiment (<i>with</i> environmental perturbation) of the complexity lower bounds model.....	134
Figure 33. Statistical results, Negative control experiment 1 of the complexity lower bounds model: Comparing the effect of environment change with and without change in environmental dissociation on resulting complexity trends	137

Figure 34. Case study results, Negative control experiment 1 of the complexity lower bounds model: Comparing the effect of environment change with and without change in environmental dissociation on complexity trends.....	138
Figure 35. Statistical results, Negative control experiment 2 of the complexity lower bounds model: Comparing the effects of gradual and punctuated environment change on the resultant complexity trends	140
Figure 36. Case study results, Negative control experiment 2 of the complexity lower bounds model: Comparing the effects of gradual and punctuated environment change on the resultant complexity trends	141
Figure 37. Case study results, Negative control experiment 3 of the complexity lower bounds model: Comparing the effect of the presence and absence of a selection pressure against complexity on the resultant complexity trends	142
Figure 38. Figure illustration of the effect of having an replicator with easy to change functionality on the resultant complexity lower bounds and complexity trends	147
Figure 39. Statistical results of the NAND gate model	154

DECLARATION OF AUTHORSHIP

I,..... [please print name]

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

[title of thesis]

.....
I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission:

Signed:

Date:

Acknowledgements

Writing this thesis has been a fascinating, absorbing and at times difficult process that simply could not have happened without the invaluable input and support of many different people. Firstly, I am deeply indebted to my supervisors; to Richard Watson, for his incredible enthusiasm, support, and sharing of his deep technical knowledge, for the endless hours and lunchtimes he's given me to debate, pull apart and rearrange the intricacies of the ideas in this thesis; and to Alex Kraaijeveld, for his unfaltering eye for detail, sound guidance, technical input and for keeping my theoretical feet on the ground; moreover to both of them for having the willingness, flexibility and belief to let me follow my own path. I'm also deeply grateful to members of the Institute for Complex Systems Simulation – in particular Seth Bullock, Jason Noble and James Dyke for their technical guidance and philosophical discussions, Millie Zedan and Stuart Bartlett for broadening my academic horizons, and Nicki Lewin for making things run as smoothly as they possibly could have. I have also been incredibly fortunate to receive inspirational advice and feedback from a number of external academics, in particular Sean Carroll, Nick Lane, Marc Bedau, Doyne Farmer and Eörs Szathmáry (who suggested the name combinatorial exaptation) – but most notably my Master's Degree supervisor and long-term friend Ros Rickaby, who has been incredibly generous with her time, sharing her always brilliant ideas, many of which have helped shape the research in this thesis.

Finally, I am hugely indebted to my father, Peter Davies, whom through his technical brilliance, deep insight and vast experience has been an irreplaceable scientific sounding board and contributed many of the essential pieces of the scientific puzzle to this work, and my mother, Lindsey Davies, without whose unfaltering support and motivation I simply would not be writing this today. And lastly, for Serena, for giving me the love and support that makes it all worthwhile.

Chapter 1: Introduction and literature review

There is nothing in neo-Darwinism which enables us to predict a long-term increase in complexity.

– J. Maynard Smith, 1969

... macroevolutionary patterns cannot be deduced from microevolutionary principles.

– G. Ledyard Stebbins and Francisco J. Ayala, 1981

1.1 Motivation

Biological evolution exhibits an increasing trend in complexity of the most complex organisms. Even though complexity can be difficult to define, it is hard to deny the earliest prokaryotes are simpler than the single-celled eukaryotes that evolved from them, which are in turn simpler than multicellular organisms that evolved from them, and so on (Bedau 2009; McShea 1991; McShea 1994).

Some researchers argue that the Modern Synthesis (Fisher 1958; Huxley 1942; Wright 1931; Dobzhansky 1970; Haldane 1990), which is the current theory in evolution (Pigliucci 2007), already explains this trend (Bedau 2009). A common argument is that the basic mechanisms of natural selection described in the current theory (e.g. Godfrey-Smith 2007; henceforth *evolution by natural selection*: ENS) are sufficient to provide an infinite space of possibilities to evolution, and therefore, this system will eventually produce a generic trend of progressively more complex organisms (Bedau 2009).

However, most laboratory experiments and computational models that embody those mechanisms have failed to display such long-term, general trends of increasing complexity as observed in nature (Bedau 2009; Lane 2010; Spiegelman et al. 1965; Oehlenschläger and Eigen 1997; Bedau et al. 2000; Watson 2006).

A telling example is provided by the work of Sol Spiegelman in the late 1960s (Spiegelman et al. 1965). Spiegelman took a simple virus of 4500 nucleotide

Chapter 1: Introduction and Literature Review

bases that made up only a handful of genes, many of which produced proteins whose purpose was to subvert the complicated machinery of host cells. He then added the virus to a test tube with a free supply of the RNA replicase enzyme necessary for the virus to reproduce, plus some free nucleotides and salts. Periodically he moved the RNA to a new test tube with fresh solution. The results were dramatic. The virus reproduced steadily, and then gradually started to lose genes – specifically, genes that were necessary to survive in the complicated environment of the host cell, but not necessary in the test tube (for example, genes that subverted the complicated machinery of the host cell). Not only that, but the shorter viruses could reproduce faster, allowing the shorter mutants to prevail. After 74 generations, the original virus with 4,500 nucleotide bases ended up as a dwarf genome with only 218 bases. Over successive generations, he found that successful RNAs become progressively simpler, losing all genes that were unnecessary in the test tube environment. Evolution favoured stripped-down, simple-as-possible organisms because these were the fastest at reproducing. As Nick Lane eloquently summarises (Lane 2002):

‘Evolution selects for beneficial adaptations to a particular environment, and the simplest, fastest or most efficient solution will tend to win out, even if it means excess baggage is jettisoned and organisms become less complicated.’

With the advent of faster computing in the 1980s and 1990s, researchers sought to study evolutionary trends using computational simulations, as part of the field of artificial life.

Many of these simulations actively sought to reproduce what was considered to be ‘open-ended’ evolution observed in the biosphere. In many cases the hope was to provide conditions for evolution that produced progressively more complex and diverse forms over time. Tierra (Ray 1992), Avida (Adami et al., 2000), Polyworld (Yaeger 1994) and Geb (Channon 2001) are examples of these types of models; typically, they have no explicit goal other than survival and reproduction.

Tierra (Ray 1992) is an evolutionary model in which self-replicating digital programs compete for resources (i.e. processor time) on a virtual computer. In

several Tierra simulations, interesting cycling behaviours of parasitism and immunity arose in the system, resulting in a coevolutionary arms race. However, Tierra-like systems all inevitably struggle to continually produce novelty (Channon and Damper 2000). Furthermore, the parasites that evolved did so as a result of a drive towards simplicity: simpler programs could reproduce with less processor time, and so were more efficient. As a result, parasites evolved that did not have their own copying code, but instead hijacked the copying code of other programs.

In Polyworld (Yaeger 1994) and Geb (Channon 2001), digital agents compete for survival in a two-dimensional world. Again, some interesting behaviours result from evolution in this system, such as flocking and foraging – but again evolution in these systems eventually struggles to produce further novelty.

To address this problem of decreasing novelty over time Lehman and Stanley (2011) adopted a different approach, by defining novelty search – a system in which evolution is explicitly rewarded for creating novelty, as opposed to functionality that promotes survival and reproduction alone. They achieved some interesting results, such as showing that novelty search can outperform directed evolution in deceptive problems (i.e. those that typically lead evolution away from the target). However, although novelty search could be useful as an engineering tool, it is not clear how much explicit selection for novelty can tell us about complexity trends in natural evolution, as evolution in the natural world is not known to include such a force.

In a more recent study Auerbach and Bongard (2014) examined how the complexity of the environment can affect evolved complexity. They used a 3d model of organism morphology, similar to Carl Sims' pioneering work on blocky creatures (Sims 1994). They found that when the environment was more complex – in particular, more rugged – organisms selected for a locomotion task generally evolved more complex locomotion mechanisms. This provides evidence that in some cases at least, environments that pose more complex tasks for evolution can consistently result in more complex organisms being evolved.

However, in sum, many artificial life simulations – in particular those without an explicit or directed fitness function – showed the same general behaviour observed by Spiegelman: in many cases organismal complexity generally

Chapter 1: Introduction and Literature Review

decreased over time and settled at some minimum, at which it remained apparently indefinitely (Langton 1984; Ray 1993; Bedau et al. 1997; Sayama 1999; Bedau et al. 2000; McMullin 2000; Suzuki, Ono, and Yuta 2003). Again, this was commonly attributed to simpler organisms requiring fewer resources and being able to reproduce faster.

Another important study on the evolution of complexity is Lenski's *Escherichia Coli* long term evolution experiment. The experiment, which is still on-going, has tracked genetic changes in 12 initially identical populations of *E. Coli* since 1988, making up over 60,000 generations (Lenski 2003). The populations are grown in an incubator in a minimal growth medium, and each day 1% of the population are transferred to a fresh flask of growth medium.

The general results show similarities to artificial life simulations; initially, the populations evolved fairly rapidly to their new environment. All populations produced larger cells in response that were specialised for living on glucose (which was abundant in the medium), resulting in a 70% faster reproduction time. However, after approximately 20,000 generations, the initial rapid changes had dwindled (Lenski 2004). Some novel complex functions were evolved; in particular, one population evolved the capability to metabolise citrate, which was very useful in the highly oxic conditions of the growth medium (Blount et al., 2008). However, despite these changes, the results predominantly show gradually decreasing optimisation to a given niche, and the ability to solve specific problems by evolving new functionality, but not an open ended growth of new forms (Lenski 2004, Blount 2008).

Given these results, we are therefore left with two rather conflicting observations: a general trend of increasing maximal complexity in the biosphere, and a common inherent preference for simplicity observed in artificial evolution experiments. Furthermore, the Modern Synthesis has little to say about what causes trends in complexity in evolution, and the origins of complexity trends remain an open question (e.g. Maynard Smith 1969, McShea 1991; Bedau 2009).

Based on these observations, the key, overarching question that motivates this work is:

How can evolutionary theory be refined to better explain observations that natural evolution exhibits a general trend of increasing maximal complexity, whereas in experiment evolution commonly results in systematic complexity minimisation?

1.2 Theoretical perspective

Before we look into this question further, we must first address another important and subtle issue. In this thesis we define ‘evolution’ as the complete connected set of mechanisms and algorithms that underpins biology, that is only partly understood, and that mapping and understanding this is one of the main goals of research into evolutionary biology. In contrast, we define ‘evolution by natural selection’ (ENS) as the specific algorithm that Darwin (Darwin 1859), and later others (for a thorough review see Godfrey-Smith, 2007) have defined, that is the central component of the Modern Synthesis, and can be generally summarised as variation, heredity, and fitness differences (Godfrey-Smith 2007). Importantly ENS does not explicitly include any specific genotype-phenotype maps, processes of development, niche construction, or other such higher-level processes.

Evolution is a phenomenon that spans multiple levels of organisation, and so may require a different type of theoretical framework than is common in science (Stebbins and Ayala 1981; Watson 2012; Mitchell 2009). For example, a problem with multi-level science is that it is not clear that having a theory that entirely explains phenomena on one level can, even in principle, explain phenomena on levels above (Mitchell 2009; Stebbins and Ayala 1981). Stebbins and Ayala compared different levels in evolution (e.g. genes, phenotypes, ecosystems) to the organisational gap between physics, chemistry and biology. They argued that although the mechanisms of physics and chemistry clearly operate in biological systems, few scientists would argue that complex biological phenomena can be predicted using the laws of physics and chemistry alone. They therefore argue that the same is presumably true in evolution: although the process of ENS clearly *operates* in phenotypes and ecosystems, this does not mean that patterns in those systems can be predicted or explained by ENS alone.

Chapter 1: Introduction and Literature Review

For an analogy, Watson (2012) discusses how several different classes of sorting algorithm all contain some essential algorithmic elements of sorting; but describing these elements alone does not provide a description of how these algorithms function. Sorting algorithms are used to sort lists in to some order (e.g. sorting a list of words alphabetically). The most common class of sorting algorithms is based on repeated compare-and-swap operations, where two records are compared, and swapped if necessary. There are many different compare and swap sorting algorithms (e.g. bubble sort, merge sort, etc.) and they vary greatly in the patterns they produce while sorting, and their efficiency. But they are all simply based on repeated compare and swap operations. What separates them is how those compare and swap operations are organised (e.g. starting at the top of the list and working down, or choosing random positions in the list, etc.). Clearly, if we want to explain why one such sorting algorithm is more efficient than another, or produces different patterns while sorting, we cannot do so with a theory that only describes the compare and swap operation. Such a theory cannot differentiate one compare and swap sorting algorithm from another, because the differences occur at a hierarchical level above the theory itself. Watson argues that evolution is similar: ENS is like the ‘compare and swap’ operation of evolution – it is the bottom level component of the algorithm. As a result, although ENS can explain phenomena on the level of genes, it simply does not contain the information to predict phenomena at higher levels. In other words, ENS is *consistent* with a very large range of possibilities at the levels of phenotypes and ecosystems, but it does not contain the necessary information to differentiate between them. Therefore, ENS alone cannot identify which of these myriad possibilities actually occurs in nature. This is consistent with the view that, as Stebbins and Ayala state, ‘macroevolutionary patterns cannot be deduced from microevolutionary principles [alone]’ (Stebbins and Ayala 1981).

In this dissertation, we build on these ideas. In general, we attempt to search through this space of higher-level algorithms that contain ENS, in an attempt to find algorithms that produce the higher-level patterns that are observed in reality. We do this by using models to reverse-engineer possible solutions. In doing so, we commonly describe phenomena that require algorithms ‘beyond ENS alone’. It is in the particular, multi-level sense described here that we mean ‘beyond ENS alone’. To be clear, by this statement we do not mean algorithms

that are inconsistent with ENS, or in any way disagree with the findings of the bulk of evolutionary theory. This simply implies that some phenomena can only be explained by a more complex algorithm that contains ENS organised in some way at a higher level by some necessary higher-level algorithmic components (e.g. specific types of development processes, etc.) – similar to how some higher-level properties of sorting algorithms cannot be reproduced by compare and swap operations alone, even though they are essential (i.e. no sorting occurs if the compare and swap operations are removed). Furthermore, it is important to note that without considering the multi-level approach we adopt in this dissertation, much of the point of the exercise could be missed. For example, consider the case where we observe that ENS alone fails to produce a given biological phenomenon, but that a higher-level algorithm that contains ENS does produce it. Viewing evolution on a single level, one might consider ENS to be *the* algorithm of evolution; in that case, all that such experiments show is that ENS (plus some proximal details) can cause the biological phenomenon, which just confirms what we already knew, because ENS (i.e. evolution, from this perspective) causes all biological phenomena. Whilst this isn't strictly speaking 'wrong', and moreover ENS is essential to the result, this point of view can miss the bigger picture in the same sense that, for example, saying "adaptation is caused by chemistry" would miss the algorithm of ENS.

1.3 Approach and previous work

Before we discuss the motivating question of this thesis in more detail, we must first discuss complexity itself. In particular, although biological complexity is to some degree intuitive, it has proved to be very difficult to agree on a universally accepted definition for what we mean by complexity (Mitchell 2009). Obviously, this significantly clouds the issue of the evolution of complexity.

Many measures of complexity have been proposed. Some simple proposals suggested complexity could be related to genome size. However, some micro-organisms have genome sizes hundreds of times larger than humans, which seems to disagree with intuitive notions of complexity. Complexity has also been linked to the entropy content of a message (Shannon 1948) such as a genome (Mitchell 2009). However, by this measure, the highest complexity

Chapter 1: Introduction and Literature Review

messages are those that are entirely random – which again seems to disagree with intuitive notions of complexity. Another popular measure of complexity is the algorithmic information content of an object, which is often termed Kolmogorov complexity. Kolmogorov complexity is defined as the length of the shortest computer program that can generate a complete description of an object (Kolmogorov 1965). However, like entropy, Kolmogorov complexity assigns the highest complexity to random objects that those we would intuitively define as complex (Mitchell 2009). A number of complexity measures have been proposed to solve these issues, including effective complexity (Gell Mann and Lloyd 1996), logical depth (Bennett 1995), thermodynamic depth (Lloyd and Pagels 1988) and statistical complexity (Crutchfield and Young 1989); however, although these proposals each have their benefits, none has been universally accepted as being equivalent to what we intuitively mean by biological complexity.

Now let us move on to the motivating question of this work. Two main possibilities are described in the literature to explain why evolutionary experiments commonly fail to generate such long-term trends of increasing maximal complexity as observed in nature. First, there is the possibility that natural evolution contains some factor, missing from the current theory (and hence not included in artificial evolution experiments) that in some cases promotes or necessitates complexity in evolution (complexity drivers; e.g. McShea 1991; McShea 1996; McShea and Brandon 2010). Second, there is the possibility that even in conditions when complexity would be favoured by evolution, there might be some kind of roadblock to complexity that, to bypass, requires evolutionary mechanisms that are not fully described in the current theory of evolution (and hence not included in artificial evolution experiments) but that are present in nature (complexity roadblocks; e.g. Bedau et al. 2000; Watson 2006).

This unpacking gives rise to two more specific research questions that form the central issues addressed in this dissertation; they will take a little more background context to define. The first of these specific questions relates to complexity roadblocks. To define it, we must briefly discuss non-decomposable functions, and mechanisms of their evolution – in particular exaptation and building block processes.

1.3.1 Complexity roadblocks

In the literature, perhaps the most popular candidate for a potential complexity roadblock is non-decomposable functions (also sometimes called emergent, non-additive or irreducibly complex functions; Watson 2006; Schwenk and Wagner 2004; Behe 2009). Non-decomposable functions are a popular candidate for complexity roadblocks for a number of reasons. First, they have a similar structure to biological complexity, in that they contain interactions between their components vital to their functionality; second, there are numerous examples of non-decomposable functions in natural evolution, but they have rarely been evolved in simulations or laboratory experiments; and third, they are difficult to evolve by ENS alone (Lenski et al. 2003; Watson 2006; Bedau et al. 2000; Watson and Pollack 2005). What in particular makes them difficult to evolve is that they cannot be broken down into smaller components without losing their functionality. This makes it difficult to explain how such functions could have been evolved from simpler systems by small, successive changes described in the current theory.

A number of different mechanisms capable of evolving non-decomposable functions have been described in the literature. Exaptation (also termed preadaptation) is one of the most commonly discussed. Exaptation occurs when a trait that has one (or no) function is co-opted for a different purpose (Darwin 1859; Gould and Vrba 1982; Barve and Wagner 2013). This enables non-decomposable function evolution because, for example, a particular trait that was initially evolved for a decomposable function could subsequently be exapted for some other, non-decomposable function, thus explaining how its simpler forms were selected for. In fact, although it is rarely mentioned, logically *all* mechanisms that can evolve non-decomposable functions must involve exaptation, because by definition a non-decomposable function cannot be broken down into simpler components without losing its function (and so therefore evolving a non-decomposable function, however it is done, must at some point involve a change of function). As a semantic note, the term exaptation was initially introduced to refer to the *result* of a co-opted function, and not the *process* of co-option itself (Gould and Vrba 1982). But rather like adaptation, which can refer to both the result of a process and the process itself (Ridley 2009), the term exaptation is increasingly used to also refer to the *process* of co-opting a function to a new use (e.g. Lavialle et al. 2013; Brosius

Chapter 1: Introduction and Literature Review

1999; Bejerano et al. 2006). We use this meaning in this document, and thus refer to functions being ‘exapted’, and ‘undergoing exaptation’.

Exaptation has been widely studied in the fields of paleontology and organismal biology (e.g. True and Carroll 2002; McLennan 2008; Budd 2006). Exaptation has also been the subject of a number of computational models that connect it to a range of related phenomena, including speciation (Graham and Oppacher 2007a), modularity (Mouret and Doncieux 2009), hierarchy (Miglino, Nolfi, and Parisi 1996), the evolution of novelty (Barve and Wagner 2013; Graham and Oppacher 2007b; Lund and Parisi 1995) and others (e.g. Gabora, Scott, and Kauffman 2013; de Oliveira 1994). However, such models tend to be the exception as opposed to the rule; exaptation has received relatively little attention from the computational modelling community compared to other processes of non-decomposable function evolution.

In addition to exaptation, a number of evolution mechanisms for evolving non-decomposable functions have been proposed that create complex functions by combining simpler, existing functions (Watson and Pollack 2005; Goldberg and Holland 1988; Lenski et al. 2003; Jacob 1977; Budd 2006; Gregory 2008; Thornhill and Ussery 2000). These mechanisms have attracted interest in part because they can create evolutionary ‘transition’ like behaviour, in which new evolutionary entities are created from simpler components, which is a common property of biological complexity evolution (Maynard Smith and Szathmary 1997). Furthermore, because they can operate recursively, such mechanisms seem to provide a potential route to ‘open ended’ complexity (Lenski et al. 2003). In this dissertation we build on this approach, and focus in particular on mechanisms of evolution by joining functions. We consider two existing mechanisms in detail: *Building block models* and *tinkering*.

We use the term *building block models* to refer to a well developed collection of computational and mathematical models stemming largely from the evolutionary computation literature that carry out evolution by combining simpler building block functions to make more complex functions (e.g. Goldberg and Holland 1988; Watson 2006; Lenski et al. 2003; Arthur and Polak 2006; Mouret and Doncieux 2009). Building block models have been used to study a range of phenomena including the benefit of sex (Watson 2006), endosymbiosis (Watson 2006), exaptation (Mouret and Doncieux 2009), the

evolution of technology (Arthur and Polak 2006) and the evolution of complexity (Watson 2006; Lenski et al. 2003).

Meanwhile, tinkering refers to François Jacob's conceptual framework for innovation and synthesis in evolution, defined in his now classic 1977 *Science* paper 'Evolution and tinkering', and based on his experience with the actions of regulatory systems and the genetics of development (Jacob 1977). Jacob does not provide a formal definition of tinkering, but describes it conceptually as evolution producing novel functionality by either repurposing or combining existing components – rather like a tinkering engineer who can fashion many different devices from a toolbox of existing components by combining them in different ways. Jacob's view has been strongly supported by subsequent discoveries of evolutionary developmental biology: in particular the strong conservation of developmental pathways between most complex organisms, showing that even very dissimilar species are often the result of simply different combinations of virtually the same underlying 'toolbox' of developmental circuits (Xu et al. 1997; Cohn and Tickle 1999; Abzhanov et al. 2004; Carroll 2005; Müller 2007).

Together, tinkering and the collection of building block models constitute a significant body of work on evolution by combining functions. However, there are a number of remaining issues within these ideas that we will focus on. These are: (a) the lack of an agreed conceptual framework that integrates tinkering, building block evolution and exaptation (a particular problem is the lack of a formal description of tinkering) and (b) recursion (how can building blocks be combined recursively over multiple levels of hierarchy?)

Tinkering is the predominant conceptual framework for evolution by combining functions in organismal and developmental biology, (e.g. Alcock et al. 2010; Fliceck 2013; Laubichler 2007). On the other hand, the ideas contained within building block models provide the most common conceptual framework for combining functions in the field of artificial life (Watson 2006; Watson and Pollack 2005; Forrest and Mitchell 1993; Goldberg 1989). Although tinkering and building block evolution both describe evolution by combining functions, research in these areas remains largely unconnected to each other, and there is no unified theory that integrates them. Furthermore, as we have discussed, because both tinkering and building block evolution are

Chapter 1: Introduction and Literature Review

mechanisms of non-decomposable function evolution, they must also both involve some form of exaptation. However, apart from a few notable exceptions (e.g. Mouret and Doncieux 2009), research into exaptation also remains largely separate from ideas of building block evolution and tinkering. Part of the problem is that tinkering remains a conceptual idea, and lacks a formal analysis (Laubichler 2007; Fliceck 2013; Bock and Goode 2007).

In general, the lack of such a unified theoretical framework for evolution by joining functions makes it difficult to understand the underlying principles of this process, and to apply findings from one area of the field to others.

The second issue that we will address is that there is no consensus on how to enable evolution by joining functions to occur recursively across multiple levels of organisation. In more detail, a particular attraction of evolution by joining functions is that functions formed by combining simpler components could potentially then be used as components for functions on the next level of organisation, and so on (Arthur 1993; Simon 1969). This process would therefore allow recursive scaling up of units of variation within evolution, potentially allowing evolution to transcend multiple levels of organisation and scale to complexity. To explain why this is difficult, we must first understand that virtually all mechanisms that combine functions to evolve non-decomposable functions use some extra ‘evolvability’ machinery (e.g. a specific genotype-phenotype map, or complex genetic change operators) beyond that of ENS alone (e.g. Lenski et al. 2003; Mouret and Doncieux 2009; Watson 2006).

Evolvability is commonly defined as the capability of a system for evolving – that is, not just generating diversity, but for generating adaptive (i.e. fit) diversity (Altenberg 1994; Wagner and Altenberg 1996; Kirschner and Gerhart 1998; Houle 1992; Wagner 2005). Mechanisms that constrain, or direct the effects of variation in useful ways increase evolvability. There are a number of ways in which this can be achieved, such as by having complex genetic operators (e.g. sexual recombination; Watson 2006; de Visser and Elena 2007), or a specific genotype-phenotype map (Wagner and Altenberg 1996; Kirschner and Gerhart 1998; Gerhart and Kirschner 2007; Wagner 2005). Let us consider how genotype-phenotype maps affect evolvability. Through the action of developmental gene regulatory networks, genotype-phenotype maps turn

genetic information into functional phenotypic components (Carroll 2005; Erwin and Davidson 2009; Davidson and Levin 2005). Genotype-phenotype maps therefore also control how genetic variation is turned into phenotypic variation. How a given genotype-phenotype map converts genetic information into a phenotypic component will affect how likely any given genetic change will result in a fit (i.e. adaptive) change to the phenotype. As a result, the makeup of the genotype-phenotype map (such as the particular process of development, structure of gene regulation networks, etc.) will affect the resultant evolvability of its organism (Wagner and Altenberg 1996; Kirschner and Gerhart 1998; Pigliucci 2008). An organism with a genotype-phenotype map that mostly converts genetic variation into useful phenotypic variation that is likely to be adaptive, given the prevailing environmental conditions, will be more evolvable than an organism with a genotype-phenotype map that mostly converts the same genetic variation into non-functional phenotypic components, or phenotypic components that are unlikely to be suitable given the prevailing environmental conditions.

Evolvability has been subject to a great deal of study; there has been particular focus on the evolution of evolvability (i.e. how the capacity for evolvability itself can evolve; Pigliucci 2008; Draghi and Wagner 2008; Pavlicev, Cheverud, and Wagner 2011; Steiner 2012), the relationship between evolvability and modularity (for example, modular gene regulation networks have been shown to be better able to cope than non-modular gene regulation networks when the environment changes in a modular manner; Kashtan and Alon 2005; Kashtan, Noor, and Alon 2007; Parter, Kashtan, and Alon 2007; Kashtan et al. 2009; Mouret and Doncieux 2009; Clune, Mouret, and Lipson 2013; Variano, McCoy, and Lipson 2004), how evolvability promotes functional robustness (Wagner 2005; Aldana et al. 2007; Lenski, Barrick, and Ofria 2006; Palmes and Usui 2005; Whitacre and Bender 2010) and the relationship of evolvability with algorithmic ‘learning’ processes such as the Baldwin effect and Hebbian learning (Watson et al. 2014; Badyaev 2009; Crispo 2007; R. Watson, Buckley, et al. 2010; R. Watson, Mills, et al. 2010).

With respect to evolution by joining functions, the results of a number of related studies imply that some forms of evolvability machinery (in particular, specific types of genotype-phenotype map) are helpful when joining functions because they can provide a mechanism of modular ‘encapsulation’ (Mouret and

Chapter 1: Introduction and Literature Review

Doncieux 2009; Kashtan, Noor, and Alon 2007). Modular encapsulation effectively means having some way of redeploying a complex, multi-component non-decomposable function in the phenotype (such as a complex section of metabolic pathway) as a single, coherent unit. (This is sometimes described as the result of processes of parcellation and integration; Günter P. Wagner, Mezey, and Cakabretta 2001; Mouret and Doncieux 2009). For example, modularly organised gene regulatory networks have been used in this capacity. Given a gene regulatory network that is organised in a modular manner, so that some small changes in the genotype correspond to large, organised changes to whole functional modules in the phenotype (Günter P. Wagner, Pavlicev, and Cheverud 2007), then small genetic mutations to such hierarchical genetic ‘switching’ genes can allow whole phenotypic modules to be redeployed at once. The result is that these complex phenotypic functions are modularly encapsulated. In short, such gene regulatory networks act as hierarchical control mechanisms that allow moving of whole functional modules in the phenotype. This is particularly important for joining functions, because it allows evolution to sample different interacting arrangements of these complex functions (i.e. it allows evolution to join them in different ways) with only small changes in the genotype.

How does this relate to recursive joining of functions? The problem with using such a genotype-phenotype map to enable recursive joining of functions is that the genotype-phenotype map must not only facilitate this type of modular phenotypic change, but the modular level at which the genotype-phenotype map operates at must also change over time. In detail, as phenotypic functions are combined, phenotypic functions are created on a new, higher level of phenotypic hierarchy. Therefore, to continue joining functions recursively on these new levels of hierarchy, the genotype-phenotype map must itself evolve new hierarchical levels of modular control to enable redeployment of these new, higher level functions.

However, evolving the genotype-phenotype map in this manner is particularly difficult because many models show that selection on genotype-phenotype maps is often *second-order*. First order selection occurs on mutations that affect the phenotype, and hence fitness, of the current organism, and hence is generally strong. In contrast, second-order selection occurs on mutations that have no direct effect on the phenotype of the current organism, but instead

increase the ability of that organism to promote beneficial future mutations (i.e. improve its evolvability; Günter P. Wagner, Pavlicev, and Cheverud 2007). Because it has no direct effect on the fitness of the current organism, second order selection is generally significantly weaker than first order selection.

Despite the difficulties with second order selection, recent research has shown that specific regimes of environment change (i.e. that create a strong selection pressure for evolvability), or a cost for links within gene regulatory networks, can generate the kind of aligned, modular gene regulatory networks capable of this type of modular reorganisation (Kashtan and Alon 2005; Kashtan, Noor, and Alon 2007; Clune, Mouret, and Lipson 2013). However, despite showing evolution of modular gene regulation networks, these models have generally not shown the type of multi-level, recursive gene regulatory network evolution that should facilitate the recursive combination of functions over multiple hierarchical levels.

In summary, recursive phenotypic evolution seems to require parallel evolution of the genotype-phenotype map, which has proved to be difficult, and is still unresolved. A further complication to this issue is that the models that demonstrate such multi-level, recursive evolution commonly do not transparently demonstrate the difficulties with this type of evolution very clearly. For example, many models are based on logic circuits (Kashtan and Alon 2005; Arthur and Polak 2006; Kashtan, Noor, and Alon 2007; Parter, Kashtan, and Alon 2008; Mouret and Doncieux 2009; Clune, Mouret, and Lipson 2013), which, although beneficial in many ways, introduce an opaque, hierarchical variational capability to the substrate of the model. Given that recursive evolution is deeply entwined with the ability of evolution to hierarchically redeploy functional components, including such opaque functional hierarchy in the substrate itself obfuscates the issue, making it very difficult to separate the effects of substrate from those of evolution, and in particular to define a control case in which hierarchical variation is not possible.

Given these three key problems with understanding evolution by joining functions, the first, more detailed question that this dissertation addresses is:

What are the evolutionary mechanisms that are necessary and sufficient to enable or facilitate evolution of non-decomposable

Chapter 1: Introduction and Literature Review

functions by combining functions, and what enables natural evolution to do this recursively across multiple scales of organisation?

To address the first component of this question we analyse the existing theories of exaptation, building block models and tinkering, in an attempt to produce a unified conceptual framework for evolution by combining functions. This analysis indicates that tinkering describes two separate processes: one in which an object undergoes a shift in function (functional shift), that is equivalent to exaptation, and one in which multiple objects are combined to make a new function (functional combination), that is equivalent to the core mechanism behind many building block models. This analysis therefore simplifies exaptation, building block models and tinkering to these two core processes (functional shift and functional combination). We then further show that because both of these processes are known to be able to evolve non-decomposable functions, then they must both involve exaptation, as we discussed earlier. To account for this, we show that these two processes in fact constitute two separate *types* of exaptation. Functional shift involves what we term *shift exaptation*, which is exaptation as it is usually defined, in which the new functionality occurs on the same organisational level as the component involved. On the other hand, functional combination involves what we term *combinatorial exaptation*, where the new functionality occurs on a higher hierarchical level than the components involved.

To contrast these mechanisms, consider a collection of enzymes in a chemical system. Each enzyme has a single function, which is to catalyse a specific reaction. *Shift exaptation* would occur when one enzyme was co-opted for use at catalysing some other chemical reaction. In this case, the new function occurs at the level of the individual enzyme. On the other hand, imagine that we randomly rearranged the interactions between enzymes in the system (keeping the function of each *individual* enzyme fixed). Occasionally we might stumble across an *arrangement* of enzyme functions that permits some complex *sequence* of reactions, hence producing a new function of a metabolic pathway at a higher organisational level. In this case, no individual enzyme changes its function, but given the right arrangement of enzymes in interaction with one another, a new functionality springs into existence at a higher organisational level. This springing into existence is combinatorial

exaptation. To concretise this theoretical framework, we provide a computational model that illustrates the key properties of combinatorial exaptation, and how they relate to tinkering, building block models and exaptation. This model, and associated theoretical framework, builds on a range of existing models and conceptual ideas (e.g. Simon 1962; Goldberg 1989; Forrest and Mitchell 1993; Watson 2006; Lenski et al. 2003; Kashtan and Alon 2005; Parter, Kashtan, and Alon 2008; Mouret and Doncieux 2009; Bock and Goode 2007).

To address the second component of the earlier research question – i.e. how evolution by combining functions can occur recursively across multiple levels of hierarchy – we specifically use the model of computational exaptation to identify the key factors in the problem facing recursive mechanisms of combinatorial exaptation. Finally, we use the model to present a solution to this problem, and hence provide a mechanism of evolution by combining functions that can occur recursively and spontaneously over multiple levels of organisation.

In summary, this research provides contributions to the key problems raised earlier in well-defined ways. First, it provides a unified framework for evolution by combining functions that defines how tinkering is related to exaptation and existing building block models, which without such a formal description has not been possible. Second, the model identifies key factors that prevent evolution by joining functions recursively as the presence of severe constraint in the form of ‘burden’ and internal selection (Schwenk and Wagner 2004; Riedl and Jefferies 1978). It shows that these factors occur when carrying out the kind of configurational reorganisation of complex components required by evolution by joining functions given a substrate that does not inherently contain some hierarchical reorganisational capability.

Third, the model shows that these problems limiting recursive evolution can be overcome by a specific type of ontogenetic development process that is present in most complex organisms. The model shows that this process allows evolution to join building blocks recursively over multiple levels of organisation, without any *a priori* information about the way building blocks must be organised at higher levels of organisation.

Chapter 1: Introduction and Literature Review

In summary, in this part of the dissertation (addressing complexity roadblocks) we arrive at the thesis that

Evolving non-decomposable functions by joining functional components can be described as a novel process of exaptation. As joined functions become more complex, increasing ‘burden’ in the form of ‘internal selection’ places limits on evolution by combining functional building blocks, but an ontological development process that occurs by local growth, as present in most complex biological organisms, resolves this problem allowing building blocks to be combined recursively over multiple levels of organisation in a scalable fashion.

We will now move on from complexity roadblocks to discuss the second detailed research question that motivates this thesis. This question relates not to complexity roadblocks, but complexity drivers.

1.3.2 Complexity drivers

In this section, we look at the question of what could cause the general trend of increasing complexity of the most complex organisms in the evolutionary record. Before we begin to discuss this topic, we should clarify that in general, we are interested in understanding what causes complexity to occur in evolution at all – not only the complexity of the most complex organisms. However, biological evolution displays a dizzying array of multifaceted trends of complexity increase, decrease and stasis (Gould and Eldredge 1993; Bird 1995; Uchman 2003; Fedonkin 2003; Newell 1949). Coupled with the difficulty associated with defining complexity (Mitchell 2009; Crutchfield and Young 1989; Kolmogorov 1965; Adami 2002; Edmonds 1995), and hence the interpretation of many complexity trends is highly controversial. Thus to avoid such controversy, we focus on perhaps the most obvious and widely accepted general trend in biological evolution, which is that the complexity of the most complex organisms has generally increased over time.

It used to be a commonly held belief that evolution inherently implies ‘progress’ towards increased complexity (Carroll 2001; Lane 2010). But evidence does not support this; while there is a general trend of increasing

complexity of the most complex organisms, and on the species level some do exhibit apparent trends of increasing complexity, many species have remained in complexity stasis for periods of many millions of years, or have shown long-term trends of decreasing complexity (McShea 1996; Carroll 2001).

Furthermore, as we have discussed, artificial evolution experiments, such as Spiegelman's laboratory experiments and artificial life simulations have generally not shown an inherent preference for increased complexity in evolution, but instead often show a preference for simplicity.

Evidence from the fossil record of stasis and trends of complexity decrease, coupled with these types of artificial evolution experiments, have lead to the idea of evolutionary 'progress' towards complexity falling out of favour (McShea 1996; McShea 1991; Bedau 2009). But without a bias towards complexity in evolution, how do we explain the trend of increasing complexity of the most complex organisms? Clearly, on some occasions, evolution *does* generate trends of increasing complexity. But on what occasions is complexity favoured?

There are a number of proposed theories on what might cause evolution to generate complexity (see McShea 1991; Bedau 2009; McShea 1994 for reviews). They can be separated into three main categories: two categories of driven mechanisms (internalist and externalist) and one of undriven mechanisms.

Internalist theories argue that trends of increasing complexity in evolution are caused by some inherent property of complex systems, or evolution itself (McShea 1991). For example, some argue that as evolution proceeds, the parts of a species that were evolved earlier generally end up having more dependencies placed on them, and hence become harder to change, and therefore harder to remove. The result is that in many cases when simpler solutions are possibly available, the build up of constraints on existing functions prevents them from being evolved. This type of process has been called Generative Entrenchment (Wimsatt 1986; Wimsatt 2001), the Path of Least Resistance (Saunders and Ho 1976; Saunders and Ho 1981) and increasing 'burden' (Riedl and Jefferies 1978); in general such theories can be described as complexity resulting from a build up of evolutionary constraints within an evolving lineage over time (in this dissertation we term these theories

complexity by increasing constraint). But a potential problem with complexity by increasing constraint theories is that they struggle to explain why we observe complexity stasis or trends of complexity decrease in the evolutionary record. If constraints inevitably build up within lineages as evolution proceeds, then we should expect all lineages to steadily increase in complexity over time, which is not generally observed (Carroll 2001; McShea 1996).

Rather than invoke some internal property of evolution, externalist mechanisms look to the external environment as a source of the increased complexity. One example is selection for complexity, where more complex organisms are proposed to be more efficient than simple organisms and hence are more reproductively successful (Rensch 1966). But again, this struggles to explain observed trends of decreasing complexity, and also seems somewhat contradicted by Spiegelman's results. Another externalist theory proposes increases in complexity are produced as a side effect of selection of other features, such as size, for example (Katz 1987; Rensch 1966; McShea 1991). But perhaps the most common mechanism of externalist theories is that increase in organismal complexity occurs due to increased complexity of the environment (e.g. McShea 1991; Knoll and Bambach 2000). As ecosystems become more diverse, new niches become available, prompting new and perhaps more complex organisms – sometimes called an expanding 'ecospace' (Knoll and Bambach 2000). But a problem is that by shifting the cause of complexity to the environment, we are left having to explain what caused the environment to become more complex in the first place, and what specific part of the environmental complexity caused organismal complexity.

The final category of complexity mechanisms is undriven mechanisms. These propose that trends of increasing complexity are not the result of relentless driving forces, but occur passively as evolution progresses. Most undriven theories propose that complexity changes as a random walk (commonly known as 'passive diffusion'): by chance alone, some evolutionary lineages could happen to wander towards higher complexity (Maynard Smith 1972; Fisher 1986; McShea 1991; McShea 1994). Moreover, if there is a complexity floor (some minimum below which no lineage can go) then the mean of all complexity lineages is expected to go up. But again, Spiegelman's results, and the similar subsequent models and experiments that support it, provide at least one example of an apparent bias (for simplicity) inherent within evolution,

because simpler organisms can reproduce more quickly, and require fewer resources to do so. If there were no other force affecting complexity in evolution, as undriven mechanisms propose, then we should perhaps expect that evolution should, just as Spiegelman observed, reduce lineages to their simplest possible incarnations and keep them as such.

Another commonly cited problem with theories for complexity in general is that most remain as verbal arguments, making it hard to verify that an evolutionary system that contained them would actually produce the trends they attempt to explain (Bedau 2009; McShea 1991).

1.3.3 Complexity lower bounds

In this dissertation we argue that there is an important factor that is missing from many of these theories of complexity evolution – the notion of complexity lower bounds. In computer science, there is considerable interest in finding the simplest program, algorithm or circuit that can perform a given function – such as sorting a list, or adding two numbers together (Papadimitriou 2003; Ben-Or 1983; Hastad 1986; Razborov 1990; Smolensky 1987). A known result from this field is that when converting between two states (e.g. converting program input into program output, such as producing a sorted list from an unsorted one), there exist fundamental lower bounds on the complexity of the *possible* functional solutions – and that these lower bounds differ depending on the input / output pair that must be converted between (Papadimitriou 2003).

For example, exhaustive searches have shown that of the many possible logic circuits, the simplest possible circuit of elementary not-and (NAND) gates that can add 2 binary bits contains 5 logic gates – and adding 3 binary bits takes a minimum of 9 NAND gates. Similarly, it has been mathematically proven that sorting a list of length n by successive compare-and-swap operations will always take more than $n \log n$ operations (Papadimitriou 2003).

This might seem rather removed from evolution, but complexity lower bounds also exist in biology. For example, passive diffusion models already include the idea that there is a complexity floor – a complexity lower bound for the simplest organism capable of reproducing, below which no lineage can go (McShea 1991). For another example, consider metabolic networks.

Fundamental chemistry dictates the minimum number of intermediate

Chapter 1: Introduction and Literature Review

chemical reactions necessary to convert from one chemical compound to another. Some compounds can be converted directly, without any intermediate reactions, and so can be mediated by perhaps only a single catalyst. On the other hand, converting between other compounds may require many intermediate steps, and so will necessitate a much more complex metabolic pathway – that is, a chemical algorithm – containing many different catalysts arranged in a specific order. In this way, complexity lower bounds can place fundamental theoretical lower limits on the complexity of possible functional solutions for chemical or biological problems.

It is possible that the set of problems faced by evolving lineages – surviving in any given biological niche – also have *niche* complexity lower bounds that vary depending on the particular problem at hand.

Niche complexity lower bounds are interesting in the context of this thesis because they have the potential to produce trends in biological complexity, when present in an evolutionary system. As new niches were encountered, each niche would necessitate some minimum amount of complexity dictated by the specific complexity lower bound of that niche, which, spread over multiple different niches with different complexity lower bounds, could produce a requisite trend of complexity. But is this not simply saying that complexity increases due to a more complex environment? There are a few reasons, which we will describe in more detail in chapter 4, why citing complexity lower bounds as a cause for complexity is different from most existing theories that simply invoke a more complex environment. One key reason is that the theory behind complexity lower bounds implies that two environments of identical complexity can have very different complexity lower bounds. Complexity lower bounds are not a property of the environment itself; they are a property of the relationship between the organism and the environment – just as the complexity lower bound for a given algorithm depends on the relationship between input and output, and cannot be defined by simply looking at the desired output alone. For example, while initially it might seem that an algorithm whose desired output was the entire works of Shakespeare translated into Mandarin should have to be complex, we should consider that the input might simply be identical to the desired output with, say, a single letter missing from the beginning. In that case, a very simple algorithm would suffice. In this way, a complexity lower bound does not

describe the absolute complexity of the environment, but the complexity of the difference between the environment and some other set of conditions – and hence citing complexity lower bounds as a cause for complexity is fundamentally different from citing an increase in environmental complexity.

To capture the general question of whether complexity lower bounds exist in evolutionary systems, to understand how they might help to explain the apparently conflicting observations of a general trend of increasing maximal complexity in the evolutionary record (McShea 1994; Bedau 2009) and of systematic complexity minimisation in artificial evolution experiments (Bedau et al. 2000; Spiegelman et al. 1965; Lane 2010), and to also understand the relationship between niche complexity lower bounds and environmental complexity, our second more detailed research question is:

What types of environment change require adaptations that are more complex rather than merely different?

If complexity lower bounds are present in biological niches, then when an organism encounters a niche with a higher complexity lower bound than its current niche, this should cause a requisite complexity increase in the organism. We address this problem directly in chapter 4.

1.3.4 The origins of complexity lower bounds in evolution

Complexity lower bounds arise in systems where a set of inputs is required to be converted into a set of outputs. But where does this occur in evolution? One place that this can occur is in the interaction between the chemical needs of an organism's metabolism, and the organism's external environment. First, let us consider this at the level of DNA. Because virtually all life is based on DNA, and DNA replication is necessary for such life (Kornberg and Baker 1992; Ridley 2009) – it follows that producing the *conditions* (chemical compounds, temperature, etc.) that allow DNA replication is also a necessary for life, if those conditions are not already present. In other words, in niches where the conditions for DNA replication are not met, any species capable of survival and reproduction must contain some function that converts the actual, external environmental conditions into conditions that can allow DNA to replicate. In terms of complexity lower bounds, the conditions that allow DNA replication

Chapter 1: Introduction and Literature Review

represent a fixed chemical output state, and the external environment in the given niche represents a chemical input state, and the necessary role of evolution with regards to this is to generate a functional system (e.g. a metabolism) that is capable of converting between the two. In this system, just like in a computational algorithm, there will be complexity lower bounds that dictate what the minimum possible complexity solution is for any given input/output pair. And because the output state is determined by DNA, whose chemical requirements for replication represent a very small range of environments and hence are effectively fixed (e.g. Reaves et al. 2012; Lindahl 1993; Grogan 1998; Marmur and Doty 1962) then all that essentially determines the complexity lower bound in this system is the external environment.

This framework implies that complexity lower bounds are not inherent to a niche, but depend only on how similar or different that niche is to the set of conditions that allow DNA replication; we term this difference *environmental dissociation*. An environment that already has the set of conditions that allow DNA replication (i.e. zero environmental dissociation) will require no metabolism, and hence zero metabolic complexity. On the other hand, an environment that is very chemically different from the set of conditions that allow DNA replication (i.e. large environmental dissociation) would have a large complexity lower bound, because many organised intermediary chemical reactions are likely to be required to convert one state to the other, thus necessitating complex metabolic machinery.

But rather than constantly evolve functional machinery to convert the external environment to the set of conditions that allow DNA replication, why does evolution not simply alter DNA, creating some new replicator that works given the current environmental conditions? Perhaps the most simple explanation is that DNA is very functionally constrained: there are apparently very few other compounds that can be formed by small changes to DNA that work as viable replicators (e.g. Reaves et al. 2012), therefore it is easier to keep DNA the same and change the environmental conditions to fit it. The complexity lower bound model of chapter 4 illustrates that this replicator constraint is a necessary component of the complexity lower bound framework; without it, complexity lower bounds can potentially be circumvented by changing the core replicator.

We also use a separate model system (evolution of logic functions in a system of NAND gates) that is a standard model of evolutionary function in the literature to check for the existence of complexity lower bounds in existing models of functional evolution.

In summary, the complexity lower bound framework, supported by the complexity lower bound model, illustrate how complexity lower bounds occur due to the interaction of a functionally constrained replicator (such as DNA) and a heterogeneous external environment. From these results, we arrive at the general thesis that:

Environmental change motivates evolutionary change, but not necessarily any increase in complexity. However, given

- a. an organism with a replicator that can replicate in some small subset of environmental conditions, and whose replicator cannot feasibly be changed to replicate in conditions outside of this subset;
- b. an environment with heterogeneous environmental dissociation whose conditions change sufficiently gradually;
- c. an inherent selection pressure against complexity such as a cost of resources

then as competition forces evolution to leave the original environment (a), and colonise new environments (b), the magnitude of environmental dissociation of a new environment will dictate the minimum possible complexity of viable organisms in that environment, resulting in a system-wide trend of increasing complexity of the most complex organisms, coupled with local trends of complexity minimisation in individual environments, caused by (c).

The models in chapter 4 show that complexity lower bounds (in conjunction with an inherent selective bias for simplicity in evolution, such as that observed in artificial evolution experiments) produce general, system wide trends of increasing complexity of the most complex organisms coupled with local trends of complexity minimisation within individual niches – therefore helping to reduce tension between complexity trend observations in natural evolution and evolutionary experiments.

1.3.5 Homeogenesis

The model at the heart of the complexity lower bound framework is a particular, boundary case of combinatorial exaptation that we term *homeogenesis*. This boundary version of combinatorial exaptation is interesting in its own right, and is studied separately in chapter 3. Although it is studied in a separate chapter, importantly, homeogenesis also helps to define some of the theoretical framework for complexity lower bounds. As a result, chapter 3 serves as a bridge, connecting the process of combinatorial exaptation to complexity lower bounds.

Before we can define homeogenesis, we must briefly discuss some relevant background context. A widely observed phenomenon in organismal biology is that organisms often appear to contain conditions within their metabolic networks that seem to be similar to conditions in which their ancestors lived (Macallum 1926; Wald 1964; Gross 1998; Mulkidjanian and Galperin 2007; Mulkidjanian et al. 2012). For example, the chemistry of fluids in the cell interior is thought to be comparable to the early oceans, or geothermal vents, in which life began (Wald 1964; Mulkidjanian and Galperin 2007; Mulkidjanian et al. 2012). Another example is that the cytoplasm in the eukaryotic cell is in a highly reduced state (i.e. low oxidation state) even in organisms that live in oxygen rich environments (Mulkidjanian and Galperin 2007; Mulkidjanian et al. 2012). In 1926, Macallum described this phenomenon as what has been termed the ‘chemistry conservation principle’: the chemical traits of organisms are more conservative than the changing environment and hence retain information about ancient environmental conditions (Macallum 1926; Mulkidjanian et al. 2012). So, for example, the early highly reduced biochemical pathways that formed before the atmosphere became oxygenated around 2.2 billion years ago (Hazen et al. 2011) could not be substantially changed after oxygen became common, and so retain information about these low-oxygen ancestral conditions. The idea that ancient environmental conditions somehow become internalised has thus been used to hypothesize about the origins in which life began, (e.g. Mulkidjanian et al. 2012). However, there is no consensus on what processes cause Macallum’s conservation principle, or why it results in events of ‘environmental internalisation’.

In chapter 3 we present a hypothesis that homeogenesis can inherently cause environmental internalisation, and hence could be responsible for some of Macallum's observations.

When environment change occurs, the usually considered mechanism of adaptation by evolution is to make changes to the existing functionality of the species, such as the classic example of the peppered moth (Grant 1999). A second option sometimes considered is that evolution can act to change the environment instead by changing behaviours of species in a process termed 'niche construction' (or an adaptive subset of the 'extended phenotype'; Dawkins 1999; Dawkins 2004) – for example, beavers building dams, or birds building nests (Odling-Smee, Laland, and Feldman 2013; Laland, Odling-Smee, and Feldman 2000). In chapter 3 we illustrate that because of combinatorial exaptation there is a third option for adapting to environment change that is separate from either of these two traditional mechanisms. When the environment changes, instead of changing the existing functionality or the environment itself, evolution can undertake homeogenesis: it can combine the existing functionality with some new, simple 'adapter' function that 'converts' the new environmental conditions back into their immediately previous conditions (that the species' existing functionality 'expects'). This adapter would sit 'in between' the existing functionality and the environment. A simple analogy is adding a voltage adapter to an electrical appliance when taking it abroad: rather than changing the existing functionality of the appliance, or changing the power grid of the foreign country, the easiest solution is to add an adapter that 'converts' the new external environment conditions back into those expected by the existing functionality. We propose the term homeogenesis for this process because, similar to *homeostasis*, it works to preserve the same, or constant (i.e. *homeo*) conditions in the phenotype – but unlike homeostasis, which occurs by changing behaviour, homeogenesis achieves this by creating (i.e. *genesis*) extra phenotypic function.

We use the evolution of C₄ photosynthesis as an illustrative example of homeogenesis. C₃ photosynthesis (an earlier mechanism from which C₄ photosynthesis evolved) becomes less efficient as CO₂ concentrations drop (Edwards and Walker 1983; Ehleringer et al. 1991). In response to millions of years of dropping CO₂ concentrations, rather than change the function of C₃ photosynthesis, evolution simply added an adapter function to C₃

Chapter 1: Introduction and Literature Review

photosynthesis – a new chemical cycle that increases CO_2 concentrations internally within the leaf – thus counteracting the environment change and providing the existing C_3 functionality with the chemical inputs it ‘expects’. The result is a qualitatively new type of photosynthesis (C_4) that consists of a combination of the unchanged existing mechanism *plus* an environmental adapter function that recreates the previous, ancestral high CO_2 environment internally within the metabolism (Ehleringer et al. 1991).

How does homeogenesis relate to Macallum’s observations of internalised ancestral environments? Simply, by creating an adapter function that recreates the previous environmental conditions, evolution effectively creates a version of that ancestral environment internally within the species phenotype, incorporated within its function: the stored environment is the output of the adapter. For example, in the evolution of C_4 photosynthesis, when adding an adapter to recreate the high CO_2 conditions preferred by C_3 photosynthesis, evolution effectively recreated a version of the ancestral, high CO_2 external environment internally within the leaf.

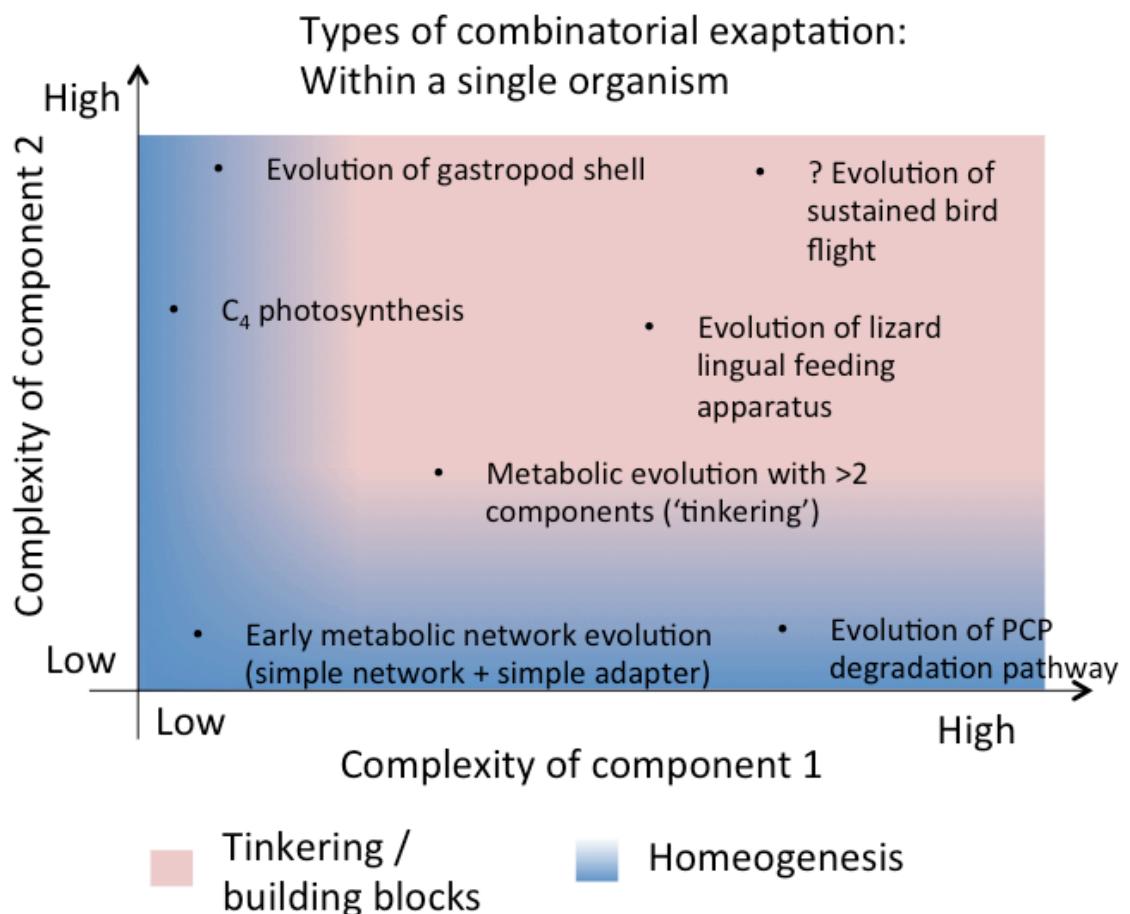


Figure 1. A figure illustration of the range of biological combinatorial exaptation mechanisms that can occur within an organism (as opposed to across multiple organisms, such as endosymbiosis). Combining an existing complex component with a simple component (blue area) will often act to maintain the existing component's functionality by using the new component as an environmental adapter (i.e. homeogenesis). This process internalises the previous external environment conditions, storing a record in the phenotype itself. In contrast, mechanisms of combinatorial exaptation that involve multiple (i.e. >2) components, or similarly complex components less obviously preserve past environments, and are commonly described as 'tinkering', or combining building blocks. In this dissertation we argue that all of these processes are variants of a single process, that we term combinatorial exaptation, in which complex non-decomposable functions are evolved by combining functional components.

Chapter 1: Introduction and Literature Review

How does homeogenesis relate to combinatorial exaptation? Homeogenesis is a type of combinatorial exaptation because it involves the combination of two separate functions (the existing function and the adapter) to create a new function capable of functionality that neither of the components themselves can carry out. But it is a boundary condition of combinatorial exaptation, as it likely involves only adding a very small adapter to a likely large existing function. Thus we can propose a scale on which we can compare combinatorial exaptation and homeogenesis (Figure 1).

In short, homeogenesis is combinatorial exaptation in which a new component is combined with some existing component (i.e. the existing organism) to sit in between the external environment and the existing function. In most examples the adapter function is simple, relative to the existing functionality. One such example is the evolution of C₄ photosynthesis, as previously discussed, where a small metabolic adapter function is added to the existing C₃ photosynthetic pathway (Ehleringer et al. 1991; Edwards and Walker 1983). Another example is the evolution of a novel metabolic pathway for degradation of pentachlorophenol (PCP), a xenobiotic pesticide, in the bacterium *Sphingomonas chlorophenolica*. This novel pathway evolved since the pesticide was introduced in the 1950s; again, this was achieved by the addition of a small metabolic adapter function added to the perimeter of the existing metabolic network (Copley 2000), which effectively recreated the ancestral, PCP free environment within the bacterium's metabolic network. Another potential example of homeogenesis in which the adapter is slightly more complex is the evolution of the hard, body-encasing gastropod shell in response to predation (Haszprunar 1988; Palmer 1979; Vermeij, Schindel, and Zipser 1981), which would have effectively recreated the ancestral predator-free environment within the confines of the shell.

In contrast, with combinatorial exaptation that involves more than two components, or where the components are both of similar complexity, there is often no clear 'existing function' or 'adapter', and is more commonly termed tinkering, or building block evolution. Possible examples include the evolution of the lingual prehension lizard feeding apparatus, in which a number of previously separate musculoskeletal components were brought together to provide a novel function (Günter P. Wagner and Schwenk 2000; Schwenk and Wagner 2001), and the evolution of sustained avian flight due to the

combination of set of complex features including forelimbs, feathers and a novel, highly efficient breathing system (Berner, VandenBrooks, and Ward 2007; Claessens, O'Connor, and Unwin 2009). The diagram above is far from perfect; the relationship is more complex than it is possible to describe in this simple phase space. For example, it does not easily take into account the possibility of combining more than two components. However, its purpose is to illustrate some examples of the relationship between homeogenesis and other types of combinatorial exaptation.

1.4 Summary of major topics of this dissertation

In this dissertation we first describe a theoretical framework that unites the theories of exaptation, building block models and tinkering. This framework describes a new type of exaptation (combinatorial exaptation) that occurs on a higher hierarchical level than the objects themselves. We provide a computational model to concretely illustrate the properties of combinatorial exaptation, and to identify key outstanding problems with evolution by combining functions – in particular with recursively combining functions across multiple levels of organisation. We then present a potential solution to this problem within the model, hence providing a mechanism of combinatorial exaptation that can evolve complex non-decomposable functions recursively across multiple levels of organisation (chapter 2).

We then describe homeogenesis: a boundary case of combinatorial exaptation that enables adaptation without changing the existing function or the external environment by instead adding an environmental ‘adapter’ to the existing functionality. We use our theory of homeogenesis to tackle the common but poorly understood phenomenon that many organisms appear to contain conditions in their phenotypes similar to the ancient environments in which their ancestors lived. We discuss how a well-studied biological example system (the evolution of C₄ photosynthesis, which resulted in the internalisation of an ancestral environment in the phenotype) can be described as an example of homeogenesis. We then use this example, in conjunction with an abstract computational model, to show how homeogenesis internalises environments, and show that it can store whole ordered sequences of ancestral environments (chapter 3).

Finally, using chapter 3 and its model of homeogenesis as a bridge, we describe a further theoretical framework that illustrates how evolution in a system of complexity lower bounds generally results in an overall, system-wide trend of increasing complexity of the most complex organisms over time, coupled with local trends of complexity minimisation within individual niches. We then demonstrate that the complexity lower bound framework robustly produces such trends in an evolutionary context by using a concrete computational model. Finally we show that complexity lower bounds are present in a standard NAND gate model of evolutionary function (chapter 4).

1.5 Implications

The models and framework we provide illustrate the following points.

1. Combining functions is a form of exaptation.

Existing theory views exaptation as a ‘non-adaptive’ process in the sense that it is entirely reliant on chance (hence exaptations are placed in contrast with adaptations). But because exaptation is thought to be a major cause of innovation in evolution, this leaves an uncomfortable amount of explanation of evolutionary complexity down to chance events. By recognising that exaptation can occur not only by shifting the function of an existing trait (i.e. shift exaptation, as the existing theory describes, which *is* entirely reliant on chance), but also by joining the functionality of multiple traits to create a new, higher level function (combinatorial exaptation), we show that there are two distinct types of exaptation – and crucially that the latter is not entirely down to chance: we show that evolvability mechanisms, such as specific genotype-phenotype maps greatly enhance the probability of finding useful novel functionality by combinatorial exaptation. As a result, this resolves an uncomfortable issue in evolutionary theory by providing a mechanism of evolutionary innovation that is directed and systematic, in place of an explanation that relies entirely on pure chance alone.

2. Including a local, ontogenetic development process in evolutionary models can enable them to scale to complexity across multiple levels of organisation.

We have illustrated how constraints that link the logical structure of gene regulatory networks and the physical structure of the phenotype provide a mechanism of combinatorial exaptation that can evolve complex non-decomposable functions recursively across multiple levels of organisation. This observation could help to explain the benefits of development in natural organisms, and help evolutionary algorithms scale to complexity.

3. Organisms can adapt to environment change without altering their existing function or the external environment.

We illustrate that changing an organism's existing functionality (e.g. classically defined adaptation) and altering its external environment (e.g. niche construction) are not the only routes by which organisms can evolve to better fit their environments. Pointing out the logical possibility that organisms can undertake homeogenesis instead provides new possibilities for evolutionary adaptation in situations (e.g. of severe constraint) that might have previously been considered unviable for evolution.

4. Some mechanisms of adaptation systematically store external environments internally within organisms.

By showing that homeogenesis systematically stores previously experienced environments internally within the phenotype, the models in this dissertation provide one of the first mechanistic explanations for the common observation that many biological organisms contain conditions in their phenotypes that appear to represent ancient environmental conditions in which their ancestors lived (e.g. Mulkidjanian et al. 2012; Mulkidjanian and Galperin 2007).

5. Trends of complexity observed in evolution can potentially be explained by complexity lower bounds.

We demonstrate that complexity lower bounds can cause robust trends in complexity in evolutionary systems, including an overall trend of increasing complexity of the most complex organisms, similar to natural evolution. We also provide a framework that shows how complexity

lower bounds can occur in evolutionary systems, and demonstrate their existence in a standard model of evolutionary function.

1.6 Models and approach

To investigate these research questions and support the associated claims we develop conceptual arguments and formal analyses supported by computational illustrations and mathematical proofs.

1.6.1 Modelling approach

Evolution is often impractical to empirically observe, due to its long timescales and significant complexity, among other factors. Modelling is a pragmatic alternative to observing and experimenting with natural biological evolutionary systems.

The approach to modelling followed in this dissertation is to produce transparent computational models that reproduce specific behaviours of biological evolutionary systems.

We choose computational over mathematical models in many cases because they generally require less prior human interpretation about the dynamics of the system, and hence reduce the chance of accidentally oversimplifying or incorrectly modelling its behaviour. We choose transparent models over opaque models because, like for like, transparent models provide more mechanistic understanding than opaque models. A common criticism is that transparent models can seem contrived, and lack the surprising behaviour of opaque models – but that is only because transparent models are simply opaque models with the black box removed. We also choose to keep the models as simple and generic as possible, removing any unnecessary assumptions to make the important dynamics clear. Finally, in some cases we make computational models of verbal arguments. We do this because mechanisms left as verbal models can easily conceal logical contradictions or fallacies that are not possible in a concrete computational model, and so producing a model provides better support for the validity of the argument.

The claims of this dissertation are supported by four models, described below.

1.6.2 Model 1: Combinatorial Exaptation – exaptation by joining functions

1.6.2.1 Aims

The key aims of this model were to:

1. Provide a transparent illustration of the mechanism of combinatorial exaptation;
2. Identify the factor, or set of factors that hinder existing models from recursively combining functions across multiple levels of organisation;
3. To illustrate a potential solution to existing problems of recursively combining functions, thus providing a mechanism of combinatorial exaptation capable of recursive evolution over multiple levels of organisation.

1.6.2.2 Methods

We defined a fitness landscape that contained a set of complex functions that were all non-decomposable. Thus there were no smooth gradients in this landscape, making evolution difficult. Within this landscape, a subset of these non-decomposable functions had structurally related positions. This enabled the in-principle possibility of systematic, guided exaptation within this subset. The specific relationship between non-decomposable functions in this subset was hierarchical: complex non-decomposable functions were composed of specific configurations of more simple non-decomposable functions (after Lenski et al. 2003; Watson 2006). The model also included specific biologically motivated constraints missing from many such systems:

1. Reorganisation of building block non-decomposable functions in the phenotype was not trivial, as all components of a given non-decomposable function had to be reorganised individually, unless some hierarchical control mechanism (such as a development process) had been evolved separately (which was not provided).
2. The most complex non-decomposable functions required combining building blocks across three scales of organisation. At each hierarchical level, the scale of blocks that had to be reorganised was different, thus necessitating the need for a mechanism capable of discovering and

Chapter 1: Introduction and Literature Review

evolving new block sizes and their control mechanisms, as evolution progressed.

We also mapped the entire fitness landscape to understand how these evolvability mechanisms help combinatorial exaptation access potential novelties.

1.6.2.3 Results

The main results of this model are as follows. The model:

1. Provides a novel mechanism of evolution that is capable of evolving novel functionality by combining building blocks recursively over multiple levels of organisation (without being given information about the hierarchical structure of blocks in advance).
2. Identifies that increasing 'burden', in the form of 'internal selection' (Riedl and Jefferies 1978; Schwenk and Wagner 2004) is a key factor responsible for hindering evolution by combining functions occurring recursively across multiple organisational levels, and in turn explains why a mechanism of progressive encapsulation facilitates this form of evolution. (In particular, burden is created by the functional constraints that exist between components of higher-level functions – i.e. internal selection; the number of these constraints increase with the number of components, and the hence complexity of the higher level function, and breaking them causes dramatic loss of functionality).
3. Graphically illustrates how having an aligned, modular gene regulatory network uses existing structure in the search space to enrich the local variational neighbourhood with fit phenotypes (building on existing work in this area; e.g. Parter, Kashtan, and Alon 2008).

1.6.3 Model 2: Metabolic evolution by homeogenesis

1.6.3.1 Aims

The key aims of this model were to:

1. Provide a concrete illustration of homeogenesis;
2. Test the hypothesis that homeogenesis commonly internalises and preserves past environment conditions, and hence can potentially

- explain Macallum's chemistry conservation principle and observations of apparently conserved ancestral environments in biological organisms;
3. Identify how the type of environment change affects the likelihood that environments, or sequences of environments will be internalised by homeogenesis.

1.6.3.2 Methods

We hypothesized that the evolution of C₄ photosynthesis evolved by homeogenesis, and using that example as a starting point, defined an abstract chemical representation of this general class of mechanism.

The model was of metabolic evolution in response to a changing environment. It was based on a simplified chemistry; we defined a chemical network of possible reactions, in which all organism function and environment change in the model takes place. Populations underwent evolution across a gradually spatially heterogeneous environment. Traditional mechanisms of adaptation were not available: the organisms' existing replication functionality could not be changed, and neither could the external environment. But organisms were able to evolve sets of metabolic reactions (i.e. successive 'environment adapters') to convert between the external environment and the fixed chemical needs of their replicator.

Variants of the model explored how different types of environment sequences affected both the structure of the evolved metabolic networks (and hence how previous environments are 'stored'), and the progress of evolution. A variant of this model was also used as the primary tool in chapter 4 to illustrate the capability of complexity lower bounds for generating requisite trends in complexity.

1.6.3.3 Results

The main results of this model are as follows. The model:

1. Illustrates that homeogenesis is a viable mechanism of environment change, and demonstrates a general set of conditions under which it will occur;

2. Shows that homeogenesis causes ancestral environment conditions to become internalised and maintained within the phenotype significantly more often than would be expected by chance;
3. Illustrates how the properties of environment change and the underlying system (e.g. the chemical reaction network that governs environment change and organismal function) affect (a) the likelihood of environmental internalisation occurring, and (b) the nature of environment information that is preserved;
4. Shows that homeogenesis can adapt organisms to environment change without altering either the existing functionality or the external environment.

1.6.4 Model 3: Complexity trends of evolution in a system of complexity lower bounds

1.6.4.1 Aims

The key aims of this model were to:

1. Test the hypothesis that evolution in a system that contains complexity lower bounds in evolutionary niches robustly results in requisite trends in complexity.
2. To identify the characteristic complexity trends produced by evolution in a system of complexity lower bounds, specifically to test if the trends resemble experimental observations of local complexity minimisation, and observations from the biosphere of a system-wide general increase in the most complex organisms.
3. To identify the relative extent to which (a) complexity lower bounds and (b) constraint caused by existing, contingent adaptations affect the resultant complexity of evolved organisms.

1.6.4.2 Methods

The model was based on the model of metabolic evolution described in chapter 2. As with the model in chapter 2 we included a cost of resources that created a continual selection pressure in favour of simplicity throughout the model.

There were two versions of the model used in this chapter. The first version was a simple model that is very similar to the model in chapter 2. It has a simple heterogeneous environment containing a few niches that does not change over time. This model was used to examine the basic properties of evolution in a system of complexity lower bounds. The second version of this model was larger, containing 50 niches, and in some cases included periodic temporal environment perturbations. This model was used to test the environmental dissociation hypothesis (see claim 3 in section 1.7, below).

1.6.4.3 Results

The main results of this model are as follows. The model:

1. Shows that complexity lower bounds cause requisite trends of organismal complexity in evolutionary systems, and provides an evolutionary context that explains how complexity lower bounds apply in evolution
2. Provides a novel mechanism and explanatory framework that enables a principled distinction between environmental change that requires evolution to make organisms *different* from how they used to be and environmental change that requires evolution to make organisms *more complex* than they used be.
3. Shows that evolution in systems of complexity lower bounds commonly produces two characteristic trends simultaneously: (a) a system wide trend of generally increasing complexity of the most complex organisms, and (b) local trends of complexity minimisation within individual niches – and hence can help to explain observations of these types of trends in biological evolution and experiments.

1.6.5 Model 4: Complexity lower bounds in NAND circuit calculations

1.6.5.1 Aims

The key aim of this model was to test whether complexity lower bounds are present in a system commonly used as an analogy for evolutionary function (i.e. a system of NAND logic gates).

1.6.5.2 Methods

The model consisted of a network of NAND logic gates set up to calculate (i.e. logically transform) a set of fixed size Boolean inputs to a given set of Boolean outputs. The space of possible solutions for each input/output pair was sampled by repeatedly evolving populations of solutions with a selection pressure for simplicity, to observe whether different sized lower bounds (i.e. complexity lower bounds) exist for different (but equal size) calculations. Finally, exhaustive searches were performed on some simple calculations, to concretely identify the presence of complexity lower bounds.

1.6.5.3 Results

The main results of this model are as follows. The model:

1. Proves that complexity lower bounds exist in a commonly used existing model of functional evolution (NAND gate logic functions);
2. Shows that transformations between identically complex environment pairs can have very different complexity lower bounds (hence showing that complexity lower bounds are not related to the complexity of the input or output in this system).

1.7 Claims

1. As joined functions become more complex, increasing ‘burden’ in the form of ‘internal selection’ places limits on evolution by combining functional building blocks, but an ontological development process that occurs by local growth, as present in most complex biological organisms, resolves this problem allowing building blocks to be combined recursively over multiple levels of organisation in a scalable fashion.
2. When both the external environment and an organism’s existing functionality are too difficult to change, a third possibility exists for evolution - adapting to environment change by adding an internal environmental ‘adapter’ that converts the new external conditions into those necessitated by the organisms existing functionality – and in

doing so, inherently creates an internal replica of the previous environment within the organism's phenotype.

3. Environmental change motivates evolutionary change, but not necessarily any increase in complexity. However, given
 - a. an organism with a replicator that can replicate in some small subset of environmental conditions, and whose replicator cannot feasibly be changed to replicate in conditions outside of this subset;
 - b. an environment with heterogeneous environmental dissociation whose conditions change sufficiently gradually;
 - c. an inherent selection pressure against complexity such as a cost of resourcesthen as competition forces evolution to leave the original environment (a), and colonise new environments (b), the magnitude of environmental dissociation of a new environment will dictate the minimum possible complexity of viable organisms in that environment, resulting in a system-wide trend of increasing complexity of the most complex organisms, coupled with local trends of complexity minimisation in individual environments, caused by (c).

1.8 Contributions

This dissertation has made the following contributions both for understanding complexity roadblocks and how they can be alleviated, and for understanding complexity drivers:

Understanding complexity roadblocks:

1. Providing a conceptual framework that unites exaptation, tinkering and building block mechanisms of evolution;
2. Identifying 'burden', in the form of internal selection, as a key factor that hinders evolution by combining building blocks occurring recursively across multiple levels of organisation, explaining its context, and showing how this causes mechanisms of functional encapsulation to be beneficial in this mode of evolution;

3. Graphically illustrating how aligned, modular genotype-phenotype maps use existing structure in the fitness landscape to enrich the local variational neighbourhood with fit phenotypes;
4. Providing a mechanism of evolution by joining functions capable of dynamically and recursively rescaling its units of variation, thus allowing evolution to spontaneously cross progressive levels of organisation.

Understanding complexity drivers:

1. Demonstrating the existence of a novel mechanism of adaptation that involves neither changing existing function nor the external environment;
2. Providing a conceptual description and functional model of an evolutionary process of adaptation that can systematically store previously experienced environment conditions in the phenotype;
3. Demonstrating that inherent lower bounds on the complexity of algorithmic problems can cause evolution to generate robust trends of increasing complexity in evolution.
4. Taken together these contributions combine to alleviate the mismatch between the complexity increases observed in natural evolution and the complexity minimisation behaviour observed in evolutionary experiments.

1.9 Scope

The work described here is theoretical, and as such addresses issues of general principle, and not empirical observations about natural biological processes. The work describes the behaviour of abstract representations of biological systems. Conclusions drawn from the behaviours of these systems should be applied to real biological systems with the appropriate amount of consideration and qualification.

However, the results of these models and conceptual frameworks are not ‘arbitrary explorations of possible biologies’ (Watson 2006). None of the models are in any sense ‘unevolutionary’; they only illustrate the capabilities of non-teleological adaptive processes. To the greatest of our ability we include the relevant constraints present in real biological systems. These constraints

limit what mechanisms can possibly produce a given phenomenon – both in real biological systems and models that contain them. Moreover, many of the mechanisms that we illustrate beyond that of ENS alone – such as involving a hierarchical development processes and evolvable gene regulation networks – are biological fact. We simply illustrate their potential capabilities. The other concepts we include all have substantial, if in some cases controversial, history in the biological literature.

In short, the aim of these models is to identify the theoretical capabilities of the evolutionary mechanisms, and clarify the properties of the evolutionary systems that we address – with the hope of guiding the direction of future empirical work to ascertain whether or not these conditions are met in natural systems. Meanwhile the conceptual and theoretic principles behind this work stand independently.

1.10 Structure of this dissertation

Chapter 1 – Introduction and literature review

Describes the central problems, concepts, relevant existing literature and scope of the thesis.

Chapter 2 – Combinatorial exaptation

Describes the conceptual relationship between exaptation, tinkering and combining building blocks, illustrates a transparent working model of this mode of evolution across multiple scales of organisation, and clarifies various intricacies and problems inherent with such mechanisms and how they can be resolved using some known ontogenetic properties of complex organisms.

Chapter 3 – Homeogenesis

Illustrates, by means of a simple computational model and discussing well-studied biological examples, a mechanism of evolution capable of a) adapting to environment change without altering existing functionality or the external environment, and b) systematically storing previously experienced environment conditions in the phenotype.

Chapter 4 – Complexity lower bounds

Chapter 1: Introduction and Literature Review

Describes an evolutionary framework in which fundamental limits on the lower bound of solutions to algorithmic problems can effect evolution, producing robust and multifaceted trends of organismal complexity – in particular a general trend of increasing complexity of the most complex organisms.

Chapter 5 – Summary and conclusions

Provides a summary of arguments, illustrations, experiments and contributions of the dissertation, draws conclusions and discusses possible further work.

Chapter 2: Combinatorial Exaptation

2.1 Introduction

In this chapter, we are interested in understanding how the current theory of evolution can be refined to better explain how evolution is capable of scaling to complexity. A central aspect of this problem is expressed in the state of artificial evolution techniques, such as evolutionary algorithms, and artificial biological evolution: Despite their success as optimization methods, evolution in these formats generally struggles to scale to complexity (Bedau et al. 2000; Mouret and Doncieux 2009; Bedau 2009; Bedau et al. 1997; Spiegelman et al. 1965; Oehlenschläger and Eigen 1997; Lane 2010). In natural evolution, scaling to increasing complexity is often associated with evolutionary transitions, in which new levels of hierarchical organisation are created by joining small, previously separate entities into some new, larger functional entity (Maynard Smith and Szathmary 1997; Watson 2006). This behaviour is commonly missing from artificial evolution (Watson 2006; Goldberg 1989; Bedau et al. 2000). Here we consider the possibility that an evolutionary mechanism capable of crossing such transitional thresholds is an important component to enable the evolution of complexity.

In particular, we consider the possibility that these thresholds are precipitated by the existence of complex non-decomposable functions (i.e. functions that cannot be broken down beyond some threshold without losing their functionality, such as the mouse-trap example we discussed in chapter 1; Watson 2006; Thornhill and Ussery 2000; Behe 2009; Günter P. Wagner and Schwenk 2000). There are a number of proposed mechanisms capable of evolving non-decomposable functions (e.g. Thornhill and Ussery 2000; Watson 2006) – and importantly, some of these mechanisms operate by joining smaller components to make new functions, and hence imply such transitional behaviour (e.g. Jacob 1977; Gregory 2008; Watson 2006; Mouret and Doncieux 2009). We therefore focus on this type of non-decomposable function evolution by joining functions. We will focus on three mechanisms of non-decomposable function evolution in particular: *building block mechanisms* and *tinkering*, both of which describe processes of joining functions, and *exaptation*, that generally

Chapter 2: Combinatorial Exaptation

does not, but is one of the most commonly proposed mechanism of non-decomposable function evolution.

The literature on evolution by joining functions is fairly well developed. However, there are two significant issues in the field that we focus on here. The first issue is that there is no consensus on an underlying theoretical framework that describes how tinkering, building block models and exaptation are linked. Part of the problem is that tinkering lacks a formal theoretical analysis, and remains as a conceptual framework (Alcock et al. 2010). As a whole, the lack of such a framework makes it more difficult to understand the underlying principles of evolution by joining functions, and apply findings from one area of the field to others.

The second issue is that there is no consensus on how biological evolution joins functions recursively over multiple levels of organisation. In more detail, one of the features of evolution by joining functions that makes it an attractive prospect for evolving complexity is that functions formed by combining components could then potentially be used as components at the next level of organisation, and so on. This would therefore provide a recursive, potentially open-ended mechanism of complexity evolution whose units inherently scale as complexity increases. However most models of joining functions fail to achieve this type of open-ended recursive evolution (e.g. Watson and Pollack 2005; Watson 2006; Arthur and Polak 2006). Furthermore, what exactly causes the problem has been hard to identify. A particular issue is that such recursive evolution is thought to be deeply intertwined with the availability of hierarchical variation mechanisms; however, many of the models that achieve such multi-level evolution are systems of logic circuits that due to their intrinsic properties inherently and opaquely introduce hierarchy within the substrate itself (e.g. Lenski et al. 2003; Mouret and Doncieux 2009). This makes it very difficult to separate the effects of the evolutionary processes being studied from the opaque internal properties of the substrate.

The rest of this chapter is structured as follows. We first describe relevant previous work: we give a brief historical background for exaptation, tinkering and building block models, and then link this work to the problem of recursive evolution, which we describe in greater detail. This initial section recaps much of the discussion of the history and problems with evolution by joining

functions from chapter 1, but in more detail. We then analyse the three mechanisms together, and propose a theoretical framework that describes their interrelationships, and characterises what we believe is the central process of evolution by joining functions. Next, we provide a computational model that contains the key principles of joining functions to evolve non-decomposable functions, but that is built on a transparent substrate that avoids many of the problems of logic gate systems. This system enables us to clearly isolate the causal factors that inhibit evolution by natural selection (ENS) from joining functions recursively over multiple organisational levels, and links a number of related but previously unconnected biological phenomena to this process. Using this model system we are also able to map the entire fitness landscape to illustrate the problem graphically. Finally, we then provide a novel solution to the problem of recursive evolution. We show that by using a common and simple type of developmental mechanism that operates by the action of local rules, evolution is capable of evolving new variation operators aligned to new modules as they are evolved, thus enabling spontaneous recursive evolution over multiple organisational levels.

2.2 Previous work

2.2.1 Exaptation

Exaptation, or preadaptation, as it was referred to at the time, arose in response to early criticisms of Darwin (Gould and Vrba 1982; Budd 2006; True and Carroll 2002). Critics argued that some complex traits would be functionally useless if they were broken down beyond some point (such as 2% of a wing, for example), and so could not have evolved gradually for their current purpose as Darwin suggested. Darwin responded by suggesting that traits could change their functions during evolution, and so such a complex trait could have initially been evolved for some other function that was useful even when broken down further (Darwin 1859; Budd 2006; Thornhill and Ussery 2000). Exaptation has subsequently been widely used to explain the origin of complex organismal traits, and is popular in organismal biology (where it originated) and paleontology (e.g. Budd 2006; Gould and Vrba 1982; Gould and Eldredge 1993; True and Carroll 2002). It is commonly defined as the process by which traits are ‘co-opted’ to serve new functions in evolution;

Chapter 2: Combinatorial Exaptation

well-documented examples include heat-shock proteins being co-opted to form part of the eye lens, and lungs of basal fish being co-opted to become the gas bladder (True and Carroll 2002). The ‘complex traits’ that Darwin’s critics described are non-decomposable functions: they are functions that cannot be broken down beyond some point without losing their function. Furthermore, exaptation is usually considered to be a non-adaptive process (i.e. non-adaptive in the sense that it only occurs as a chance by-product of ENS, similar to genetic drift; Barve and Wagner 2013). As such, exaptation requires no extra machinery of evolution beyond ENS.

2.2.2 Tinkering

The term tinkering was coined by François Jacob in his now famous 1977 Science paper ‘Evolution and Tinkering’ (Jacob 1977), although it has roots in earlier theories (Laubichler 2007). Based on his observations with regulatory genes, Jacob developed a conceptual framework describing innovation and synthesis in evolution. He argued that novelty in evolution comes from repurposing or reorganisation of existing parts:

‘Evolution... works on what already exists, either transforming a system to give it new functions or combining several systems to produce a more complex one.’ (Jacob 1977)

Since Jacob’s description, tinkering has been shown to be responsible for evolving numerous existing biological functions (Alcock et al. 2010; Flicek 2013). However, despite these successes, tinkering remains a conceptual heuristic; Jacob did not provide a strict formal or theoretical analysis (Alcock et al. 2010), and as far as we are aware, no such analysis has been subsequently published. Many of the types of evolvability adaptations associated with tinkering (e.g. hierarchical gene regulation networks) have been well researched (e.g. E. H. Davidson 2010; E. H. Davidson and Erwin 2006; Erwin and Davidson 2009; Carroll 2005).

2.2.3 Building Block Mechanisms

Rather than refer to a formal evolutionary mechanism, we use the term building block mechanisms to refer to a collection of computational and

mathematical models, mostly stemming from the fields of evolutionary computation and artificial life, that, for example, attempted to explain the benefit of sex, or the evolution of complexity (e.g. Watson 2006; Mouret and Doncieux 2009; Goldberg and Holland 1988; Arthur and Polak 2006; Simon 1962; Simon 1969; Lenski et al. 2003). As a result, there are numerous computational models of building block mechanisms, and their processes are logically defined within these models. Building block models generally evolve non-decomposable functions by assembling smaller, building block functions into a non-decomposable function, of which they become components.

Building block mechanism research has also significantly contributed to understanding the evolution of complexity by combining functions. Early work was provided by Simon's theory of 'Nearly Decomposable' functions, that included his famous 'watch maker' parable, and later by Goldberg and Holland's genetic algorithm based 'building block hypothesis' (Simon 1969; Goldberg and Holland 1988). However, a significant problem for some of these early models was that they reasoned that to benefit evolution, building blocks within the final hierarchical function being evolved must be effectively separable – i.e. have no significant dependencies on each other (hence Simon's term 'nearly decomposable'). This meant that the hierarchical functions being evolved by these theories were not actually non-decomposable functions (Watson and Pollack 2005), which resulted in two problems. First, it was later shown that ENS alone could evolve such functions equally well, and so building block evolution provided no benefit in this case (Forrest and Mitchell 1993; Watson and Pollack 2005; Watson 2006); and second, without dependencies between components in the complex function being evolved, there was nothing to hold the hierarchical structure together, rendering them indistinguishable from an unordered collection of components, lacking the organised hierarchy observed in natural organisms (Watson and Pollack 2005).

Later models resolved this issue of separability by showing that building block models can evolve genuine non-decomposable functions (e.g. Lenski et al. 2003; Mouret and Doncieux 2009; Watson 2006). In these models, phenotypic building block components need to be found *and* organised to have the right set of interactions between them to evolve the complex functions. (For an analogy, simply evolving the components of a watch is not sufficient to keep time; they must then be organised into the right arrangement). Unlike with

Chapter 2: Combinatorial Exaptation

separable problems, when the problem posed was a genuine non-decomposable function (specifically, one that was modularly hierarchical), combining building blocks was shown to have significant benefits over ENS alone.

In the abstract, models of genuine non-decomposable functions differed from earlier models containing mere collections of components by adding ‘extra combinatorial work’ – that is, some extra step of organisation that must be done to *organise* the components of a higher-level function once those components have been evolved. There were two main approaches to this problem, and understanding them will help to explain the current state of research on recursive evolution. In more abstract models, such as Watson’s Hierarchical If And Only If model (HIFF; Watson 2006), extra combinatorial work was often introduced by having many different low level components that were each individually useful, but only a subset of which could be successfully combined into higher-level functions. In this case, evolution must find the *right* blocks to combine to evolve the higher-level function. (For example, given a selection of watch parts, to build a working watch we must first find a set of watch parts that are theoretically compatible with each other – e.g., are from the same watch). HIFF used an entirely transparent fitness function, and hence it has a great deal of explanatory power.

The second main approach to improving early building block models so that they contained genuine non-decomposable functions was to use real functional entities in the models – in particular, logic circuit systems (e.g. Lenski et al. 2003; Arthur and Polak 2006; Mouret and Doncieux 2009). This had the benefit of forcing the model to contain aspects of reality such as extra combinatorial work because they were built in to the substrate itself. In the logic circuit models, logic gates could be combined into circuits that perform given computational tasks. The fitness functions were built to be hierarchical, where complex tasks could be achieved by combining simpler functional circuits that were also rewarded. However, one downside of such logic gate systems is that they generally have a highly opaque fitness functions, sometimes making it more difficult to understand the results they produce.

2.2.4 Recursive evolution and encapsulation

One aspect common to both HIFF and logic gate models is that to evolve non-decomposable functions by joining functions requires extra evolvability machinery beyond ENS alone, such as complicated genetic operators; e.g. as used by HIFF (Watson 2006), complicated genotype-phenotype maps (Lenski et al. 2003) and community detection algorithms (Mouret and Doncieux 2009).

The reason for this is that to combine already complex components, evolution requires some way of redeploying those components as whole, integrated units. Effectively, it requires some mechanism of modular ‘encapsulation’ (sometimes called parcellation and integration; Günter P. Wagner, Pavlicev, and Cheverud 2007).

As we discussed in chapter 1, one approach to encapsulation is to use a modular genotype-phenotype map (Mouret and Doncieux 2009), in which small modules in the genotype are ‘aligned’ to large modules in the phenotype. This allows small genetic changes (to regulatory ‘switch’ genes) to reorganise whole organised groups of phenotypic traits (Günter P. Wagner, Pavlicev, and Cheverud 2007; Wagner and Altenberg 1996). However, there is no consensus on how such correctly aligned, modular genotype-phenotype maps themselves can be evolved; a particular problem is that selection on genotype-phenotype maps is commonly second order, and hence weak (Günter P. Wagner, Pavlicev, and Cheverud 2007). This problem becomes very important when evolving functions recursively across multiple hierarchical levels. Here, new functions evolved must be used recursively as components at the next level up, and so on – which therefore requires a mechanism of genotype-phenotype map evolution that identifies and encapsulates new phenotypic modules on the fly, as they are evolved in the phenotype.

How this can occur is an open question. Some building block models simply provide the system with a correctly aligned genotype-phenotype map *a priori* and illustrate the capability of evolution from there (e.g. Arthur and Polak 2006). Elsewhere, in the literature on the evolution of modularity, modular genotype-phenotype maps have been evolved under specific environmental pressures for evolvability (Draghi and Wagner 2008; Kashtan and Alon 2005; Kashtan, Noor, and Alon 2007; Parter, Kashtan, and Alon 2008) and cost of connections in the genotype-phenotype map (Clune, Mouret, and Lipson 2013).

Chapter 2: Combinatorial Exaptation

But these methods are not linked to evolution by joining functions, and also do not show modular genotype-phenotype maps evolving open-endedly across multiple hierarchical levels, and so may be limited when enabling recursive evolution.

Solving the problem of how evolution can occur recursively is made more difficult by the details of the models used to illustrate it. In particular, certain aspects of the existing models obfuscate the problem and make it difficult to identify. One particular problem is that because systems of logic gates (e.g. Lenski et al. 2003; Arthur and Polak 2006; Mouret and Doncieux 2009) are directed networks, they are inherently hierarchical, and hence introduce the possibility of small changes to their structure causing large changes to their behaviour. For example, a complex circuit of logic gates could have its output dramatically changed by simply adding or rewiring a single gate. The key point is that this ability to hierarchically change the behaviour of many nodes in the system from only a small structural change is present within the substrate itself. Moreover, this capability of logic gate systems is often highly opaque (e.g. changing a rewiring a single gate in a complex circuit will often result in changes in behaviour that are very difficult to predict without careful analysis). Because the problem of recursive evolution is deeply intertwined with the ability to hierarchically redeploy modules, having a substrate that inherently has this capability can hide (and cause us to underestimate) the problems faced by evolution if such ability is *not* present in the substrate.

For example, the hierarchical nature of logic circuits is used in this way by Kashtan and Alon's model of spontaneous modularity, and some of its derivatives (Kashtan and Alon 2005). Modularity in the model is reliant on the ability of logic circuits to have large changes in function from only few changes to their structure. However, there is no discussion about the likelihood that the substrate will contain this ability; it is simply built into each example system's substrate. What would perhaps benefit this work, and what is particularly difficult using such a system of logic gates, is a control case in which the model system does not intrinsically contain this hierarchical ability. Without such a control case, it is difficult to separate any positive effects of having a hierarchical variation operator within the algorithm of evolution from the effects of hierarchical variation present within the substrate itself. Moreover, it hides that significant problem of how evolution could evolve such a

hierarchical capability (for example, in the form of a hierarchical gene regulation network, or complex genetic operator) if it were not already available within the substrate itself. Here we aim to examine the problems that hinder recursive evolution when such an inherent hierarchical reorganisational ability is lacking. This is similar to the issues addressed by HIFF, although here we focus on using a hierarchical genotype-phenotype map to solve the problem, as opposed to genetic operators (e.g. sexual recombination) as used by HIFF.

In summary, there are two main bodies of literature that describe evolution by combining functions: tinkering and building block models. They remain largely separate; there is no theoretical framework that describes their relationship or relationship with other processes of non-decomposable function evolution such as exaptation. Furthermore, how evolution can recursively join functions across multiple levels of organisation is poorly understood – in particular, how this process relates to encapsulation.

In the next section we develop a framework that incorporates tinkering and building block models with exaptation, and describes a core process of evolution by combining functions. We then address the question of recursive evolution. To do so we define a computational model of evolution by joining functions that requires both finding and organising interactions between modules (similar to logic gate models), but that is transparent, and specifically avoids introducing hierarchical capability within the substrate itself, allowing us to better isolate the problem faced by evolution when recursively joining functions over multiple levels of organisation. Finally, we present a novel solution to the problem of recursive evolution in this system.

2.3 Theoretical analysis

We will start with theoretical analysis to examine the relationship between exaptation, building block models and tinkering. We begin by examining tinkering.

In his definition of tinkering quoted above, it is clear that Jacob actually describes two distinct processes. The first of these – *‘transforming a system to give it new functions’* is logically indistinguishable from the process of exaptation: ENS evolves a function until it takes on a new function. On the

Chapter 2: Combinatorial Exaptation

other hand, the second process that Jacob describes '*combining several systems to produce a more complex one*' shares the same logic as building block models: evolution produces novel, complex functions by combining existing functions.

So on initial inspection it seems that we can condense the three processes of non-decomposable function evolution – exaptation, building block models and tinkering – into two distinct processes:

1. 'Functional shift' (described by exaptation)
2. 'Functional combination' (described by building block models)

This reasoning shows that the second process – 'functional combination', although described in building block models and in part by tinkering, lacks a specific name, making it difficult to discuss. The lack of a name for such a process has recently been highlighted by Gregory (2008). Before suggesting a name for this process, we can make further deductions. Specifically, both functional shift and functional combination have been shown to be capable of evolving non-decomposable functions. And non-decomposable functions by definition cannot be broken down beyond some point without losing their function. Therefore, by definition, any process that is capable of producing non-decomposable functions from components below this threshold must involve a change in function – i.e., exaptation. This implies that both processes in the above list must involve exaptation. As we have described, functional shift obviously involves exaptation, because it is practically the definition of exaptation. But where does exaptation occur in functional combination?

To address this problem, let us consider the evolution of a complex metabolic pathway by the process of functional combination. Initially, many individual enzymes are already present within the organism and are used for their own, separate functions. They do not interact with one another. By mutation, a subset are then brought together and combined into a specific configuration that has a new function as a metabolic pathway: in this new arrangement, the enzymes interact with one another in such a way that produces a fundamentally new function on a higher organisational level. Before this process, the organism did not have the function of the pathway, and afterwards it did. But no individual enzyme changed its function during this

process; only the *arrangement* of the enzymes changed – but that change in arrangement was enough to result in the creation of an entirely new function.

The result is that while functional combination does involve functional change, unlike in ‘traditional’ exaptation (i.e. functional shift), the functional change occurs on a higher hierarchical level. This implies that functional shift and functional combination actually refer to two distinct *types* of exaptation: exaptation as it is traditionally described, which here we term *shift-exaptation*, in which change in function occurs on the same hierarchical level as the physical component, and exaptation that occurs in functional combination, in which change in function occurs on a higher hierarchical level than physical components being combined. We suggest the term ‘combinatorial exaptation’¹ for this process (Table 1).

Type of exaptation	Location of functional shift	Other names
Shift exaptation	Same level as object	Exaptation Preadaptation Tinkering
Combinatorial exaptation	Hierarchical level above objects	Tinkering Building-blocks Collage

Table 1. Description of two separate types of exaptation

Another example of combinatorial exaptation is provided by logic gate systems, which are commonly used as the basis for building block models (Lenski et al. 2003; Arthur and Polak 2006). NAND gates are universal, meaning that every possible logical function can be made out of combinations

¹ This particular name was suggested to us by Eörs Szathmary.

Chapter 2: Combinatorial Exaptation

of NAND gates. Consider, for example, a set of five NAND gates. There are many different ways to combine them, most of which do not represent any particular function. However, two of those possible arrangements produce the XOR function (exclusive-or). If we were to rearrange our NAND gates randomly, eventually we would reach a combination that produces the XOR function. Again, no individual component (i.e. NAND gate) changes its function in this process – but when the components are combined in just the right way, a new function springs into existence at a hierarchical level above the components themselves. This ‘springing into existence’ of a new function at a higher hierarchical level is what we term combinatorial exaptation. An example of combinatorial exaptation occurring recursively across two separate scales of organisation is described in Figure 2.

(As an aside, we should state that the particular ‘high-level’ function that is produced by combinatorial exaptation, such as the XOR function in our NAND gate example, is not necessarily ‘special’, compared to other possible high-level functions of the system. It is only that some such high level functions may satisfy corresponding complex selection pressures in the environment, and in that case are likely to be retained by natural selection. The concept of function itself, and how different rearrangements of components can create different functions is discussed further in chapter 4.)

Below we provide a transparent computational model of combinatorial exaptation. We use the model to clarify the problems associated with recursive combinatorial exaptation – in particular the need for an encapsulation mechanism that can evolve to identify and reorganise new phenotypic modules as they are evolved in the phenotype.

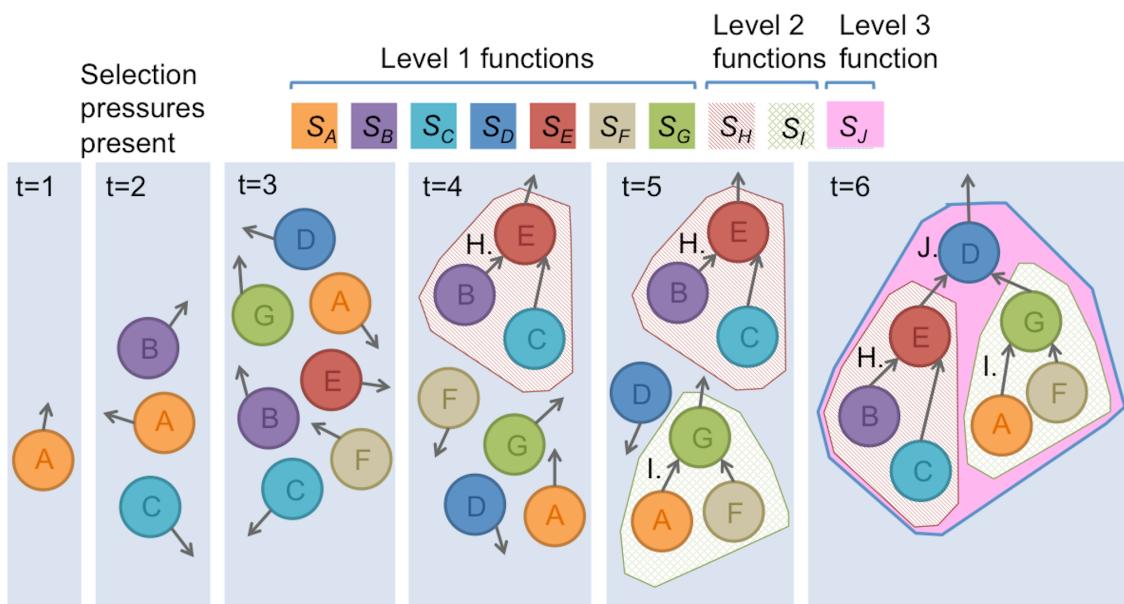


Figure 2. An example of evolving a non-decomposable metabolic pathway by multiple, recursive events of combinatorial exaptation. At t=1, enzyme A has been evolved, satisfying selection pressure S_a (circles represent enzymes, and arrows are their functional outputs). By t=3, more enzymes have been evolved, satisfying corresponding selection pressures S_a-S_g . At t=4 a mutation by chance organises functions B, C and E into an arrangement in which they interact to produce the new metabolic pathway function H, satisfying selection pressure S_h . H is a new function formed on a higher level of organisation than its components. It springs into existence when B, C and E are in the right arrangement. This is combinatorial exaptation. The same occurs to A, F and G at t=5, producing the new metabolic pathways function I, satisfying selection pressure S_i . Finally at t6 D, H and I are reorganised into an arrangement that produces J, which is a further event of combinatorial exaptation. H, I and J are all non-decomposable functions, because they require all of their components to be present and interacting in the right manner to function. They would therefore be difficult to evolve by ENS alone; however, combinatorial exaptation can easily evolve them. In this example, combinatorial exaptation occurs recursively, using functions H and I evolved by combinatorial exaptation as components on the next hierarchical level.

2.4 Results and discussion

The model system was a population of organisms evolving enzymes and interactions between those enzymes to evolve metabolic pathways. There were 16 available enzymes that could be evolved. Those enzymes could be combined into 4 medium complexity metabolic pathways, each containing 4-component enzymes. In turn, those 4-enzyme pathways could be combined into a single 16 enzyme metabolic pathway. In all simulations, this 16-enzyme pathway was the target of evolution. Because this target pathway had 3 levels of internal hierarchy, practically evolving it by joining functions required a mechanism of recursive, multi-level combinatorial exaptation.

We conducted three main simulations: S1, C1, S2. These simulations were identical except for having slightly different genotype-phenotype maps. In all cases, organisms began with a genotype phenotype map (in the form of a gene regulation network) that did not contain any hierarchical structure corresponding to the 4-enzyme or 16-enzyme pathways. The three key simulations in this chapter can be summarised as follows:

1. **S1** was a negative control experiment that sought to test the hypothesis that without such a hierarchical gene regulation network, organisms would be unable to carry out recursive joining of functions across three levels, and hence not be able to evolve the target 16-enzyme metabolic pathway. Therefore in S1, organisms had a gene regulation network without this hierarchical structure, and their gene regulation network could not evolve hierarchical interactions between regulatory genes, thus preventing the evolution of such a structure.
2. **C1** sought to test whether allowing the gene regulation network in S1 to freely evolve hierarchical interactions between regulatory genes would enable the evolution of a hierarchical gene regulation network capable of evolving the target 16-enzyme metabolic pathway. Thus C1 was identical to S1 except that the gene regulation network was allowed to evolve.
3. **S2** sought to test whether having a development process that occurred by local growth, as opposed to in a top-down manner as in S1 and C1, enabled organisms to evolve the target 16-enzyme metabolic pathway. Thus S2 was identical to C1 (i.e. with a gene regulation network that was

allowed to evolve hierarchical interactions between regulatory genes) but in S2 organisms had a development process that occurred by local growth.

To give a more detailed picture of the workings of the model, here we will briefly describe the S1 simulation. More detailed technical description of the model is defined in the methods (section 2.6)

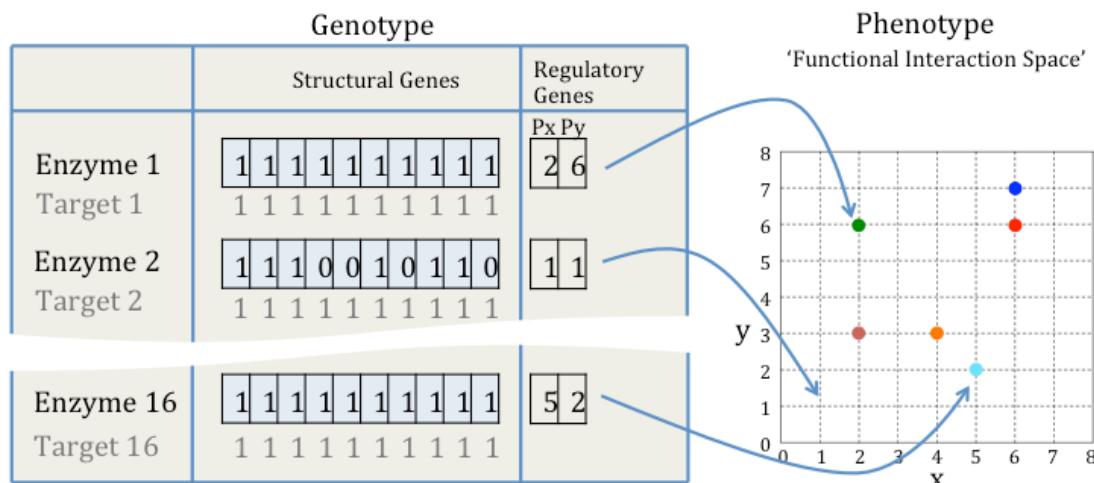


Figure 3. genotype-phenotype map for simulation S1

Enzymes that were neighbouring in functional interaction space were considered to be interacting. In this way, regulatory genes controlled interactions between enzyme functions in a given phenotype.

In S1, each organism had 32 genes; 16 structural genes (each containing 10 binary loci) that encoded 16 possible enzymes (one per structural gene), and 16 regulatory genes (R_p), one corresponding to each structural gene (each R_p gene was defined by two integers $0 < p_x \leq 9$, $0 < p_y \leq 9$). R_p regulatory genes encoded interactions between enzyme functions in the phenotype, permitting the possibility of evolving metabolic pathways. Each organism's phenotype was represented by a 9×9 grid, termed its functional interaction space, which represented the interactions between enzymes within the organism. Once a given enzyme had been evolved (by correctly setting all 10 binary loci in its corresponding structural gene), it was plotted in functional interaction space by a coloured dot, with its position defined by its corresponding R_p regulatory

Chapter 2: Combinatorial Exaptation

gene (p_x, p_y defining the x and y coordinates of the function respectively). The genotype-phenotype map is depicted in Figure 3.

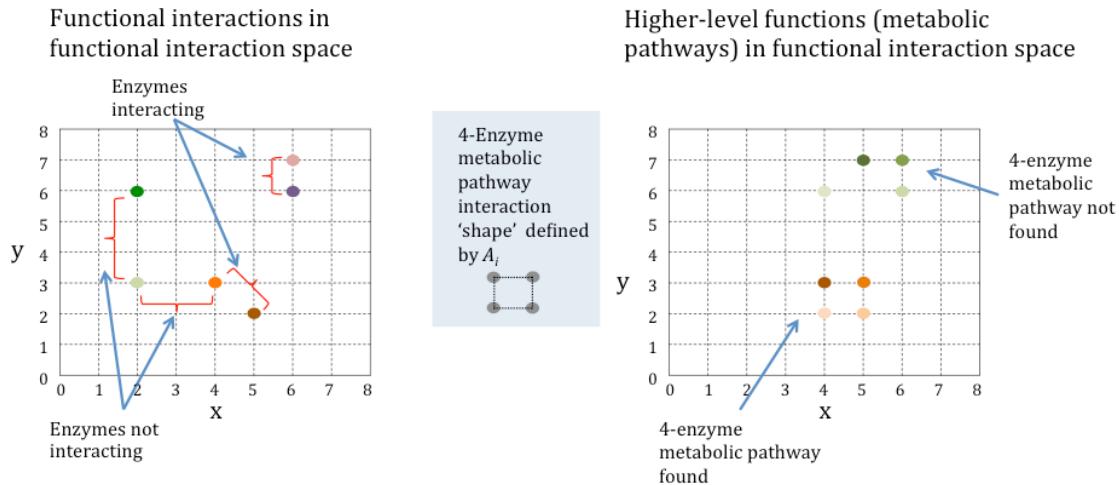


Figure 4. Enzymes interacting in functional interaction space.

The fitness function was hierarchical. Fitness was obtained by evolving any of 21 possible metabolic functions C_a, C_b, \dots, C_u . Evolving any of these functions resulted in the organism being awarded a fixed fitness bonus of +1 for each function evolved. These bonuses represented there being individual selection pressures (S_a, \dots, S_u) for each of the functions. All functions were non-decomposable functions: fitness bonuses were only awarded when functions were found exactly. Some of the functions were more complex than others. The 16 simplest functions were enzymes (C_a, C_b, \dots, C_p). Each enzyme was considered evolved when all loci in its corresponding structural gene matched a predefined fixed 10-bit target. The four next most complex functions were 4-enzyme metabolic pathways (C_q, C_r, C_s, C_t). To be evolved, each of these pathways required four specific enzymes to be organised into a specific arrangement of interactions in functional interaction space. For example, pathway C_q required enzymes C_a, C_b, C_c , and C_d to be arranged in a square formation in functional interaction space (Figure 4, right). The final function, C_u , was the 16-enzyme metabolic pathway that was the target of evolution. It consisted of C_q, C_r, C_s , and C_t organised into a particular arrangement of interactions (a square of neighbouring pathways in functional interaction space; Figure 5).

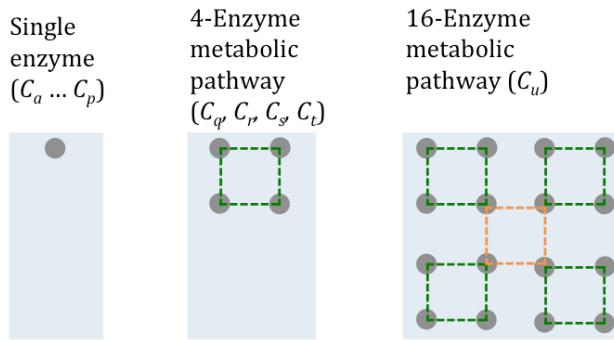


Figure 5. Diagram of the 3 types of available functions in the fitness function.

There were 21 available functions; 16 of the single enzyme type, 4 of the 4 enzyme type, and 1 of the 16 enzyme type. To evolve a function, its respective pattern of interacting enzymes must be formed in functional interaction space. Each 4-enzyme pathway required a specific set of 4 enzymes, but those enzymes could be in any order as long as it satisfied the square arrangement depicted above.

Evolution proceeded as follows: The population size was fixed at $M=50$, and evolution occurred in a generational manner. In each generation, the fittest $L=10$ organisms in the population were allowed to reproduce (i.e. truncation selection); they were copied uniformly at random and placed into a new empty population until it was full (i.e. $M=50$). The population was then mutated. There was a fixed per-locus probability of point mutation (see methods), which caused a bit-flip for binary loci, and selection of a random integer value for integer loci. No crossover or complex mutation operators were allowed. This continued for 10^6 generations or until a maximum fitness phenotype was found.

2.4.1 S1 Simulation results

In simulation S1, organisms had only the simple, direct genotype-phenotype mapping described in Figure 3 (see methods for further detail), and this genotype-phenotype map was not permitted to evolve hierarchical interactions between regulatory genes.

Chapter 2: Combinatorial Exaptation

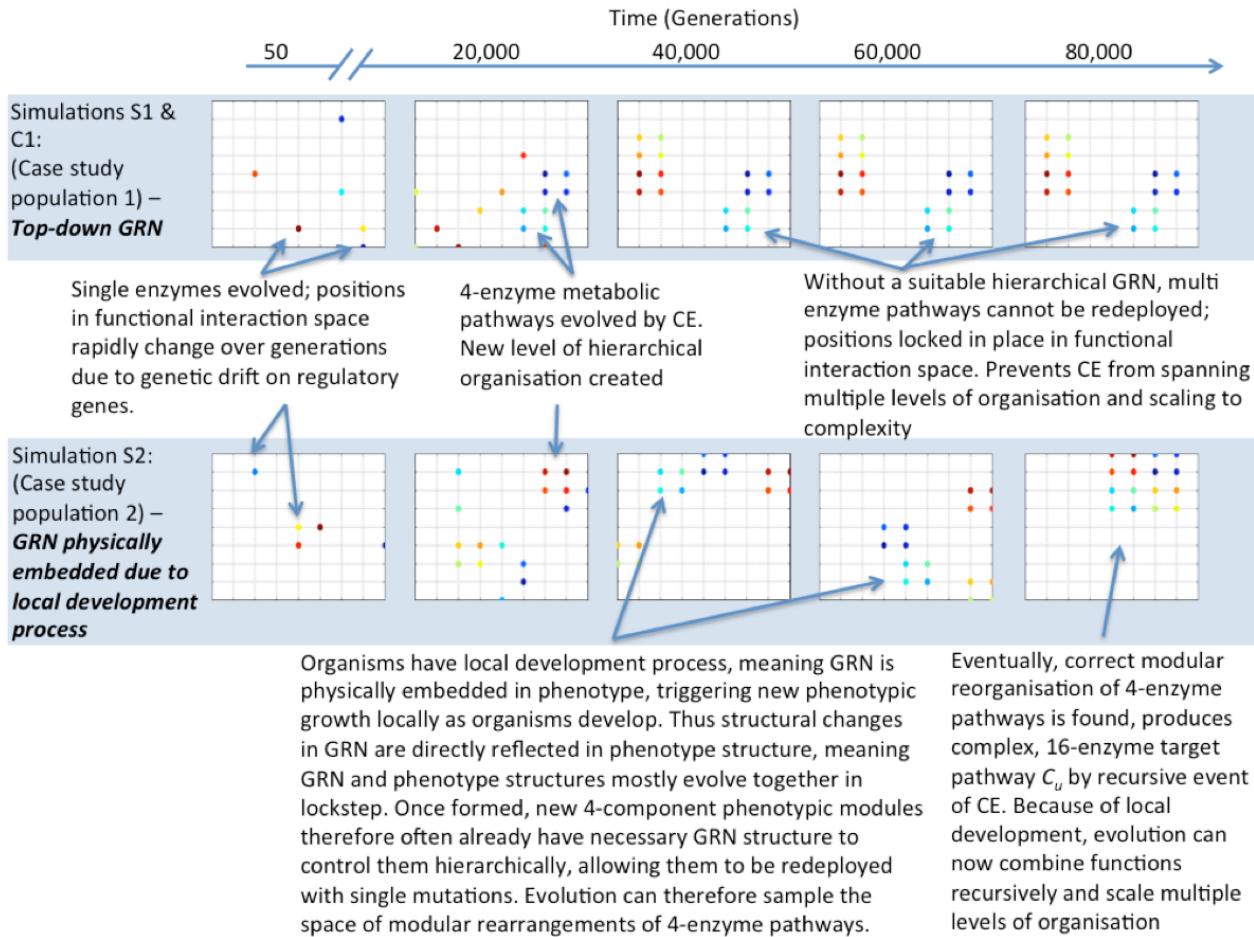


Figure 6. Case study populations show snapshots of the fittest organism in the population at various generations. Top row, (S1 simulation): Top-down development process and the gene regulatory network is not permitted to evolve hierarchical interactions between regulatory genes. Similar results were observed with a top down development process when gene regulatory networks could freely evolve hierarchical interactions between regulatory genes, and hence potentially evolve a hierarchical structure capable of redeploying 4-enzyme pathways (C1 simulation, not shown); in neither S1 nor C1 was the target 16-enzyme metabolic pathway ever evolved. But in S2 (bottom row) organisms had a *local* development process, physically embedding the gene regulation network in the phenotype, and gene regulation networks were allowed to evolve. In this case, the target 16 enzyme metabolic pathway was always evolved (e.g. bottom row, 80,000 generations). CE=Combinatorial exaptation; GP map=Genotype-phenotype map.

We carried out 100 independent repetitions of S1. In all simulations, the fitness of the population increased over time, and all 16 enzymes ($C_a, C_b \dots C_p$) and all four 4-enzyme metabolic pathways (C_q, C_r, C_s, C_t) were evolved within 10^6 generations (mean 57722; standard error 6449). However, the target 16-enzyme metabolic pathway C_u was not evolved in any of the repetitions. The progress of case study population 1 illustrates the typical behaviour of evolution in this system (Figure 6). First, individual enzymes are evolved (e.g. see snapshot at generation 50; here 6 enzymes have been evolved). Once evolved, enzymes remain persistent in the population. However, their positions in functional interaction space change randomly over time, as they are buffeted by genetic drift caused by regulatory mutations: unless they happen to form part of a metabolic pathway, for a given enzyme, any position in functional interaction space is as fit as any other.

By generation 20,000 two simple metabolic pathways (C_q and C_r) have been evolved by combinatorial exaptation. Each event of combinatorial exaptation occurred by random regulatory mutations organising the component enzymes (for example, C_a, C_b, C_c , and C_d for pathway C_q) into a specific square arrangement in functional interaction space. Once its component enzymes were in this correct arrangement, the functionality of pathway C_q sprang into existence, providing a fitness bonus for satisfying selection pressure S_q . In this way, the evolution of pathway C_q is an evolutionary transition: previously separate entities (enzymes) were brought together to form a new, emergent entity (pathway C_q) with a novel function that was not present before the transition.

By generation 40,000, all four of the 4-enzyme metabolic pathways (C_q, C_r, C_s, C_t) have been evolved. There are two particular behaviours of the system to note at this point. First, once found, higher-level functions (i.e. metabolic pathways) are persistent. By generation 40,000, pathways C_q and C_r have remained structurally intact (and hence functioning) for over 20,000 generations. Their particular internal arrangement of interactions (here represented by a square configuration in functional interaction space) is not broken up by regulatory mutations. What keeps these emergent structures together? Simply, because the function of a given pathway is dependent upon the arrangement of interactions between its components, then breaking those interactions (e.g. by regulatory mutation) would result in breaking the function

Chapter 2: Combinatorial Exaptation

of the pathway, and hence a loss in fitness. Therefore, breaking up pathways is strongly selected against, and hence their specific emergent structures are preserved. In short, internal functional dependencies (sometimes termed ‘burden’ (Riedl and Jefferies 1978) or ‘internal selection’ (Schwenk and Wagner 2004)) hold the emergent pathways together. This behaviour is intuitive: in natural organisms, it is also possible that regulatory mutations could reorganise components of important metabolic pathways (e.g. respiration) or physical systems (e.g. the heart) causing them to be broken up, and their functionality lost. In this case, such mutations would also be strongly deleterious, and hence be selected against, thus preserving the structure and function of the multi-component complex trait.

The second behaviour we observe is that once pathways are evolved, they remain stationary in functional interaction space for many thousands of generations: In snapshots at 40000, 60000 and 80000 generations, all four 4-enzyme metabolic pathways C_q , C_r , C_s , C_t have been evolved (each appearing as a square of enzymes), but remain fixed in position in functional interaction space. In contrast, before they are incorporated into metabolic pathways, individual enzymes undergo rapid genetic drift, changing their positions in functional interaction space randomly through time. It is this genetic drift that allows the system to sample many different arrangements of enzyme interactions – and ultimately to find those arrangements that produce fit metabolic pathways C_q , C_r , C_s , C_t and hence evolve them by combinatorial exaptation. Because the system is apparently incapable of performing the type of interaction rearrangement with metabolic pathways as it could with enzymes, this prevents the system from exploring different arrangements of interactions between those pathways, and hence ultimately prevents the system from evolving C_u .

The problem is that the system lacks a mechanism that can ‘encapsulate’ the newly evolved 4-enzyme pathways, and redeploy them as coherent units. Given the structure of the genotype-phenotype map, a single regulatory mutation will only move the position of one component of that pathway (i.e. one enzyme). Because the pathways are non-decomposable functions, doing so will break the required interaction arrangement of the pathway, thus break the functionality of the pathway, and therefore be strongly selected against. Such constraints that work to preserve specific, fit arrangements of functional interactions

within organisms are termed ‘burden’ in the form of ‘internal selection’ (Schwenk and Wagner 2001; Schwenk and Wagner 2004; Günter P. Wagner and Laubichler 2004). The result is that metabolic pathways cannot be moved in functional interaction space one component at a time, but instead require a set of simultaneous regulatory mutations that move each of their component enzymes simultaneously in a consistent direction. Obviously, this will become exponentially less likely as the number of components in a given function increases (and is already extremely unlikely with only 4 components). This behaviour shows why extra evolvability machinery (i.e. a method of encapsulation) is important in enabling recursive combinatorial exaptation across multiple levels. In summary, S1 illustrates that without a method of encapsulation, internal selection causes constraint (i.e. burden) to increase dramatically as the number of components in a non-decomposable function increases, thus preventing joining functions to occur recursively over multiple hierarchical levels.

2.4.2 C1 simulation results

We next sought to test whether allowing the gene regulatory network to evolve its own hierarchical structure could solve the problems posed by internal selection that prevent multi-level recursive evolution.

Simulation C1 was identical to S1, except that the gene regulatory network could evolve hierarchical interactions between regulatory genes (see methods). In short, we found that aligned modular gene regulatory networks sufficient to enable recursive combinatorial exaptation did not evolve. Instead, complicated gene regulatory networks formed rapidly that typically created many interactions between structural genes that could not be combined to form 4 enzyme pathways. As such, these pathways were ‘mis-aligned’ to the modular selection pressures in the environment, hindering further evolution (the difficulty associated with such misaligned genotype-phenotype maps is explored further in section 2.4.5). The results were similar to S1 simulations: in 100 repetitions, all 4-enzyme metabolic pathways were evolved (mean 76463, standard error 6872) but the target 16-enzyme metabolic pathway was never evolved within 10^6 generations. This supports previous similar work outlining the difficulty of gene regulatory network evolution with second order selection (Günter P. Wagner, Pavlicev, and Cheverud 2007).

2.4.3 A recursive mechanism of combinatorial exaptation

Based on the results of S1, we can describe two types of transition by combinatorial exaptation (Figure 7):

- *Type 1*, in which an associated regulatory change *has not* occurred – i.e. where phenotypic components have been combined into a new functional module, but the regulatory circuits that control those components have not been, thus preventing encapsulation and redeployment of the new phenotypic module as a coherent whole. This prevents recursive evolution by combinatorial exaptation, as occurred in the simulation S1 and C1. (Type 1 transitions are also similar to ‘egalitarian transitions’ in social evolution; some new functional symbiosis has been generated, but the system lacks a shared mechanism of genesis).
- *Type 2*, in which an associated regulatory change *has* occurred – i.e. where phenotypic components have been combined into a new functional module, *and* the regulatory circuits that control those components *have also* been combined into a new regulatory module, thus allowing encapsulation and redeployment of the new phenotypic module as a coherent whole. This would theoretically allow multi-level recursive combinatorial exaptation, but we have yet to observe it in simulation here. (Type 2 transitions are also similar to ‘fraternal transitions’ in social evolution, where new functional symbiosis is generated and the system has a shared mechanism of genesis).

So far, we have only observed type 1 transitions, thus prohibiting recursive evolution by combinatorial exaptation. To achieve a type 2 transition (and hence allow recursive combinatorial exaptation), one option is to simply provide the system with a correctly aligned genotype-phenotype map *a priori*, so that the necessary regulatory switches were already in place before new modules are formed, as some earlier models have done (e.g. Arthur and Polak 2006). However, this cannot happen in natural evolution, and so fails to properly explain the phenomenon of recursive evolution.

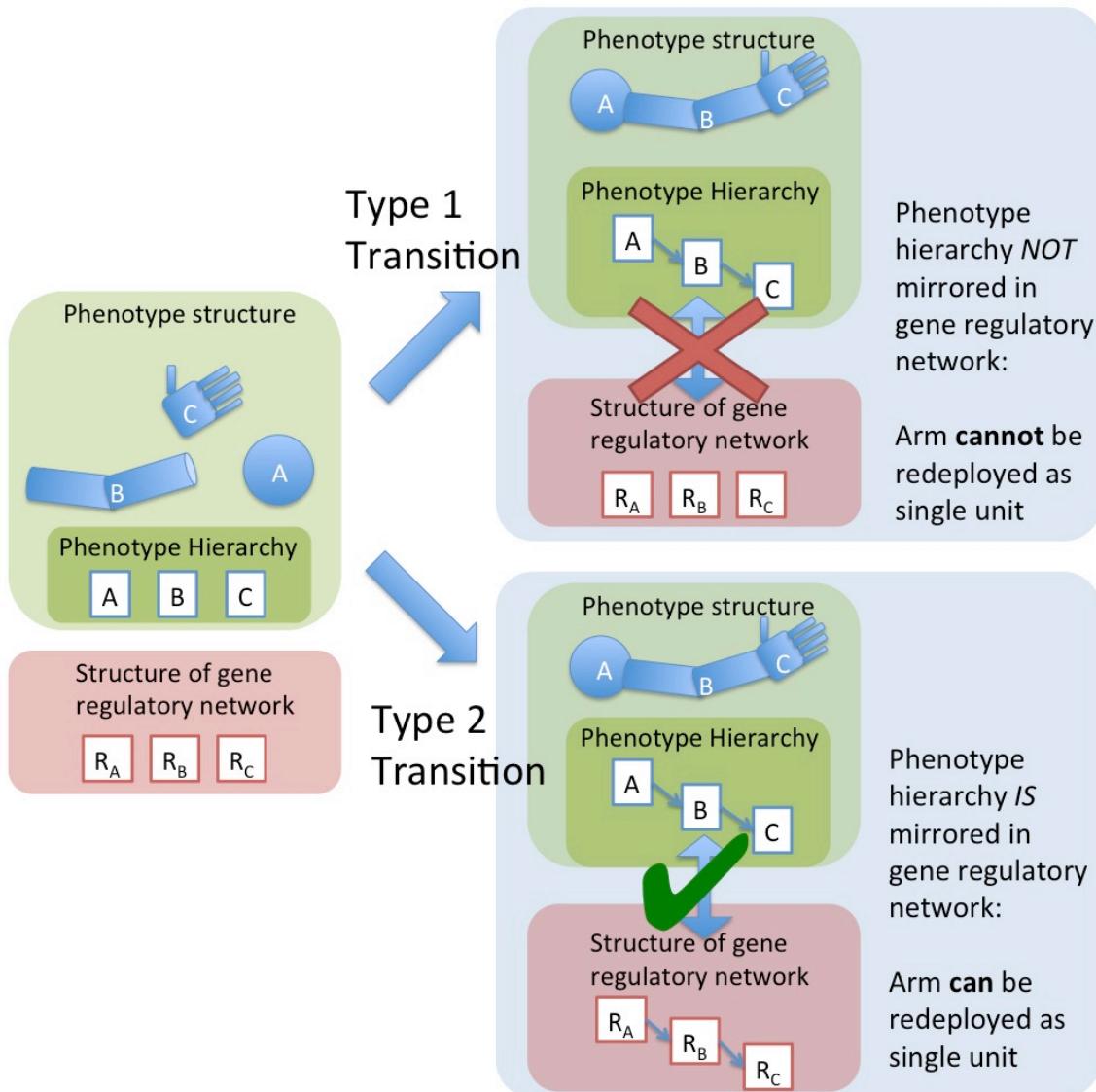


Figure 7. Two types of transition possible by combinatorial exaptation. With Type-2, associated change in the gene regulatory network accompanies the emergence of a new functional unit in the phenotype by combinatorial exaptation, and therefore the new emergent phenotypic function can be encapsulated, allowing combinatorial exaptation to occur recursively and hence scale to complexity.

Rather than further develop the approach used in C1 of using second order selection to attempt to evolve modular, hierarchical gene regulation networks, for example by using strict regimes of environment change, or a cost of connections (Kashtan and Alon 2005; Clune, Mouret, and Lipson 2013; Parter, Kashtan, and Alon 2008), we propose a different method. We suggest that a fundamentally different way to achieve the desired genotype-phenotype map

Chapter 2: Combinatorial Exaptation

structure is if the structural evolution of the genotype-phenotype map and the phenotype were somehow causally tied together, so that structural changes in one were automatically reflected in the other. A similar approach was used by Karl Sims to force virtual creature bodyplans and accompanying neuro-control systems to be aligned in his landmark study of 3-D virtual creature evolution (Sims 1994a; Sims 1994b). As with genetic regulatory systems, it is generally useful for neuro-control architecture to mirror the modular structure of the phenotype. In Sims' system, the phenotypic development occurred via a directed graph. New phenotypic components already had neuro-control systems 'built in', and so blocks of neural control circuitry were replicated along with each instanced part. The result was that the structure of the neuro-control system mirrored the structure of the modular bodyplan as it changed.

We apply this idea to genotype-phenotype map evolution, and in particular the evolution of gene regulation networks that define the genotype-phenotype map. Theoretically, a system with such linkage between phenotype and gene regulatory network structure should enable type 2 transitions and hence recursive combinatorial exaptation, because any new phenotypic module created by combinatorial exaptation would be reflected in the hierarchical modular structure of the gene regulatory network, allowing the new module to be immediately controlled as a single unit (i.e. encapsulated).

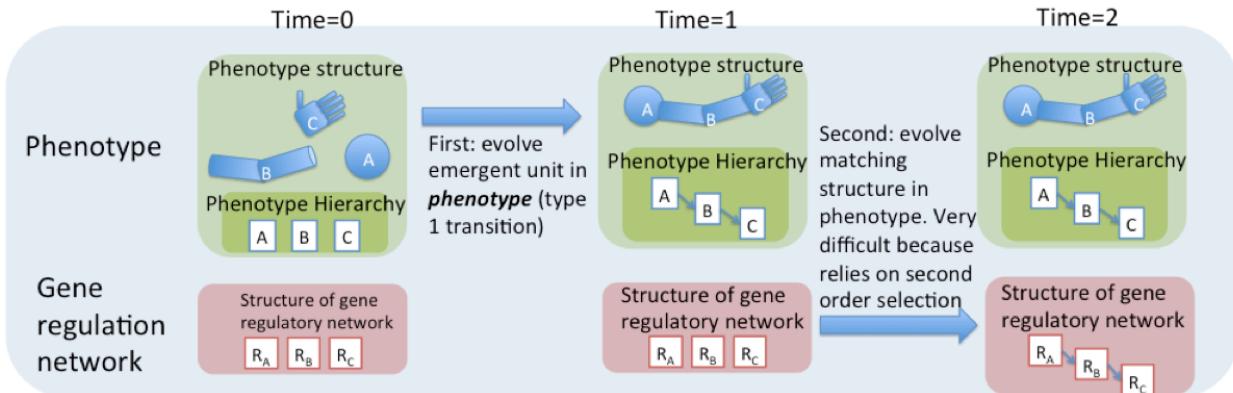
But rather than simply enforce this link, we aim to understand how such a link could exist in biological evolution. One possibility is the role of gene regulatory networks in ontological development. The fact that gene regulatory networks are responsible for encoding a process that builds a physical organism is often ignored in models, because the physical process of development is commonly abstracted away or simplified. But in nature, this necessary requirement of gene regulatory networks places strict constraints on the space of possible genotype-phenotype maps (and associated gene regulatory networks) available to evolution. Moreover, it is possible that the constraints the development place on the space of available genotype-phenotype maps occurs in a way that biases the remaining genotype-phenotype maps to be more likely to have inherent links between the structure of their gene regulatory network and the structure of their phenotype.

For example, if development is allowed to occur in a centralised, top-down manner (e.g. where individual phenotypic components are placed in the phenotype by a series of independent development events, such as in S1), then this potentially permits genotype-phenotype maps where the internal structural arrangement of gene regulatory network circuits (that determine the order in which events occur during development) does not affect the resultant phenotype structure, because phenotype construction is a series of independent events. In contrast, in real organisms, development occurs by a sequence of local interactions, and hence is inherently a process of many contingent steps. As a result, the physical position of components in the phenotype is determined to some extent by the order in which they occur, which is determined by the position of their triggering circuits in the logical structure of gene regulatory networks. The result is that gene regulatory network structure directly affects phenotypic structure in natural organisms (Erwin and Davidson 2009; E. H. Davidson 2010). Genotype-phenotype maps that operate by unrealistic development methods (as in S1) are simply not allowed in natural evolution, thus constraining the search space of gene regulatory networks in potentially useful ways.

To test this hypothesis, we restrict the model so that ontological development has to occur by a decentralised process commonly observed in natural organisms, and then allow the gene regulatory network to evolve hierarchical interactions between regulatory genes within the constraints that this development process implies. The particular development process we adopt is similar to Sims' directed graph mechanism: phenotypic components contain embedded gene regulatory network circuitry that triggers the local growth of further phenotypic components, that also have embedded gene regulatory network circuitry, and hence trigger further local growth, etc. For example, in human limb development, local expression of hox genes at the end of the zeugopod (i.e. the forearm) triggers the development of the autopod (i.e. hand and wrist; Tamura et al. 2008). In this process, every component is built by some other neighbouring phenotypic component (with which they also therefore interact), according to the hierarchical sequence defined in the gene regulatory network.

Evolving a Type 2 transition

- 1. Traditional approach:



- 2. Approach in S2: (Single step)

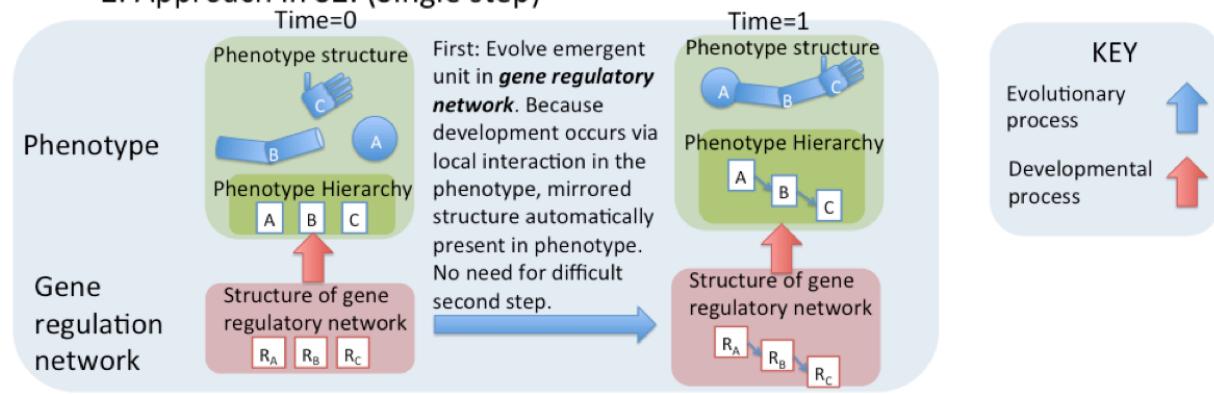


Figure 8. Comparison of two approaches to evolving a Type 2 transition. With the traditional approach, the first step is to evolve an emergent, higher unit in the phenotype by combinatorial exaptation. Here it is hard to redeploy the unit as a single unit, because the gene regulatory network (GRN) does not contain a hierarchical structure mirroring that of the new phenotypic unit. The next step is to evolve the necessary structure in the GRN; however, this is difficult because selection on the gene regulatory network is only second order. In contrast in S2, the new emergent unit is evolved in the GRN first. If development occurs by local interaction of GRN and phenotype, then structure in the GRN will often be automatically mirrored in the phenotype, because regulatory circuits that trigger each other will often produce physically neighbouring (and thus interacting) phenotypic components. The result is that often the corresponding unit will be automatically created in the phenotype due to this development process, thus avoiding the need for a difficult step of second order evolution.

In the example of human limb development, because the gene regulatory network circuitry that triggers the hand to develop exists within the forearm, the hand is formed adjacent to (and hence functionally interacting with) the forearm. In contrast, if the gene regulatory network structure was changed so that the circuit causing hand development was now triggered by the circuit that caused leg growth, then the hand would form adjacent to (and hence functionally interacting with) the leg, instead. In short, in this type of development process, phenotypic components that are hierarchically related in development are commonly hierarchically related in their phenotypic interactions. (However, we must also note that with this type of development complications can occur that mean that gene regulatory network structure will not always be exactly mirrored by phenotype structure. Even in a system where most phenotypic interaction comes about as a result of hierarchical interactions between their respective triggering circuits in the gene regulatory network, there is still the possibility that phenotypic components that were not triggered by directly related gene regulatory network circuits still come to interact.)

2.4.4 S2 Simulation results

To include this type of decentralised development process in the model, we introduce the requirement that every enzyme have some local parent structure in the phenotype that contained the regulatory circuit that triggered its development (a development trigger module). To encode this, we added two extra types of regulatory genes, R_{dtm} and R_r . R_{dtm} (an integer value $0 \leq R_{dtm} < 32$, one corresponding to each structural gene) encoded the developmental trigger module of the given structural gene. Developmental trigger modules could either be one of the other 15 enzymes ($R_{dtm} < 16$; the specific value denoted which enzyme; enzymes could not be their own parent and feedbacks were not allowed), or some other structure in the genotype ($16 \leq R_{dtm} < 32$). In accordance with the development system, enzymes were placed in functional interaction space neighbouring their developmental trigger module (and hence were interacting with it), in a direction defined by the enzyme's R_r gene (see methods). If an enzyme's developmental trigger module was not another enzyme ($16 \leq R_{dtm} < 32$), then it was assumed that the enzyme's development was triggered by some other structure in the phenotype that was effectively

Chapter 2: Combinatorial Exaptation

developmentally unrelated to other enzymes or structures in the model. In this case enzymes were placed in functional interaction space according to their regulatory genes p_x and p_y , as in S1.

Development of an organism therefore occurs by the following process (Figure 9):

- Plot each enzyme that has been correctly evolved and has a developmental trigger module that is not another enzyme (i.e. $16 \leq R_{dtm} < 32$) in functional interaction space according to its corresponding p_x and p_y regulatory genes (i.e. as in S1).
- Plot each correctly evolved enzyme with a developmental trigger module that is another enzyme next to its developmental trigger module enzyme, with a direction defined by its corresponding R_r .

In all simulations, the gene regulatory network began with the same setup as S1 and C1 (i.e. no enzymes having another enzyme as a developmental trigger module).

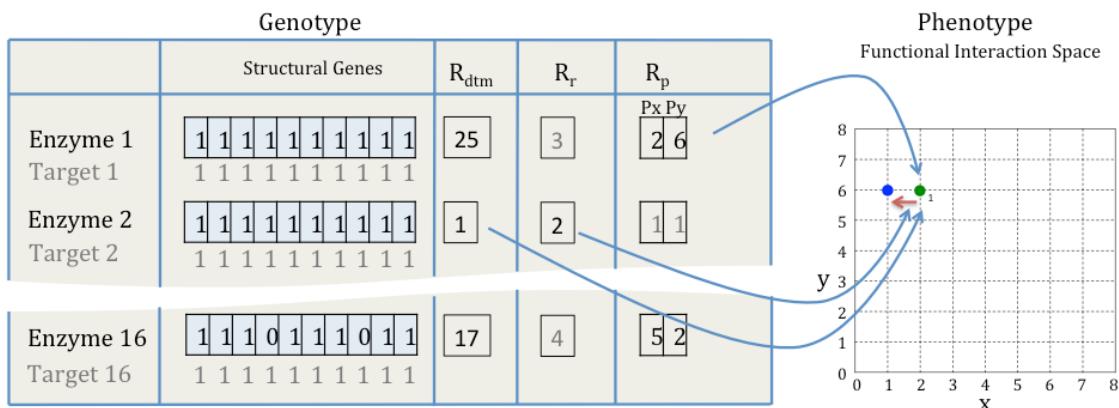


Figure 9. Genotype-phenotype map of organisms in S2. R_{dtm} of enzyme 1 is >15 , hence its developmental trigger module is not another enzyme. Its position in functional interaction space is therefore determined by R_p . In contrast, R_{dtm} of enzyme 2 is 1, indicating that its development was triggered by enzyme 1. It is therefore placed adjacent to enzyme 1 in a direction defined by its R_r (in this case $R_r=2$, signifying a direction of -1,0).

We carried out 100 independent repetitions of the simulation. In all simulations the complex target C_u was evolved within 10^6 generations (mean 118252,

standard error 7215) by a process of multi-level combinatorial exaptation. This is statistically significant when compared to either S1 or C1 simulations in which C_u was never evolved within 10^6 generations ($P < 0.001$, one sample Student's t-test). The progress of case study population 2 illustrates the typical behaviour of this system (Figure 6). In contrast to the previous simulation, in this case when 4-enzyme metabolic pathways are evolved they are not always fixed in place in functional interaction space. This therefore allows evolution to use these metabolic pathways recursively as new units of variation, changing their arrangement of interactions with each other until the complex pathway C_u is evolved. Given the constraint to the genotype-phenotype map that this development system implies, evolution can evolve a gene regulatory network that allows combinatorial exaptation to occur recursively and carry out type 2 transitions, crossing multiple levels of organisation. Let us be clear: By this mechanism, in each simulation the resulting gene regulatory network is hierarchically modular. That is, when the 16 enzyme pathway is formed, the system *already contains* a hierarchical regulatory module that corresponds to the 16-enzyme pathway, allowing the 16 enzyme pathway to be encapsulated immediately upon it being formed. This would thus allow the system to continue to further levels of hierarchy (by combining this 16-enzyme pathway with other pathways) without any increasing impediment. As such, this hierarchically modular regulatory module also contains hierarchical subcomponents that allow it to also encapsulate the smaller, 4 enzyme subcomponent metabolic pathways too, and would continue to do so as the regulatory and structural modules evolved, in concert, over progressively further levels of hierarchy. Again, this is due to the constraints in the system that the gene regulatory network must carry out development, which has the effect of 'locking' the evolution of the regulatory network and the phenotype together. Interestingly, this system generates both type 1 and type 2 transitions. As in S1, phenotypic interaction can be caused without any associated linking of developmental regulatory circuits, resulting in a type 1 transition. In this case the resultant emergent function cannot be encapsulated and used as a search unit at the next hierarchical level. But phenotypic interaction can also be caused by regulatory interactions that then result in phenotypic interactions. It is this possibility that a system of decentralised, local development introduces, and ultimately that allows combinatorial exaptation to scale to complexity in this system.

Chapter 2: Combinatorial Exaptation

In summary, we have suggested that a mechanism that links the logical structure of the gene regulatory network with the physical arrangement of phenotypic structures can provide an alternative solution to producing an aligned, modular gene regulatory network that does not rely on second order selection of the gene regulatory network. We have further shown that this process could enable evolution by joining functions to occur recursively over multiple organisational levels by providing a recursive mechanism of modular encapsulation. Furthermore, we have suggested a possible way in which constraints of natural systems, in the form of an ontogenetic development process that occurs by local growth, can provide such a link between gene regulatory network and phenotype structures.

2.4.5 The effect of genotype-phenotype map evolution on the fitness landscape

The differences between simulations S1, C1 and S2 show how having an aligned, modular genotype-phenotype map can enhance evolvability, and enable combinatorial exaptation by enabling encapsulation of complex phenotypic functions. Previous work shows that such a genotype-phenotype map enriches the local genetic neighbourhood with fit phenotypes that are modular reorganisations (e.g. Parter, Kashtan, and Alon 2008).

To study this further, we use our model to directly observe how different genotype-phenotype maps (modular aligned, non-modular and modular misaligned) reorganise the resulting fitness landscape (simulation GP1). To achieve this, we carried out evolution in the conditions of simulation S1 (i.e. with a non-modular genotype-phenotype map). We allowed evolution to continue until an organism in the population (O_{test}) had evolved all four 4-enzyme metabolic pathways. At this point, as in S1 simulations, the lack of an encapsulation mechanism prevented evolution from redeploying any of those pathways in functional interaction space, thus preventing combinatorial exaptation to continue recursively. To understand what, in terms of the fitness landscape, was preventing O_{test} from continuing combinatorial exaptation recursively, at this point we froze the simulation and then took a snapshot of the distribution of phenotypes in the local genetic neighbourhood (centred on O_{test}) that were equal or greater than the fitness of O_{test} (Figure 10).

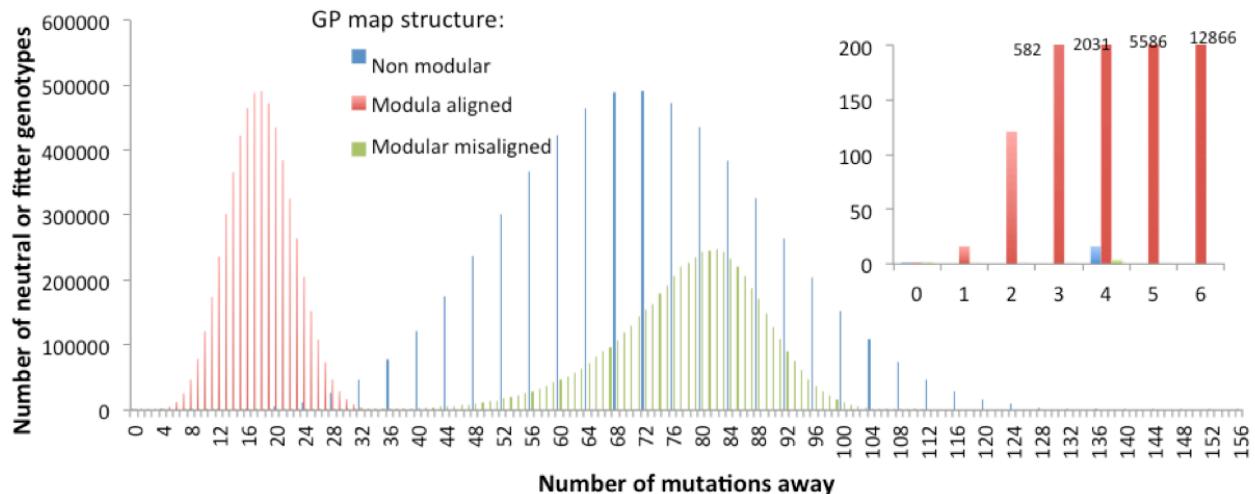


Figure 10. Main chart: Frequency distribution of neutral or fitter genotypes in the fitness landscape from a starting genotype that had evolved all 4 enzyme metabolic pathways (C_q , C_r , C_s , C_t) given three different genotype-phenotype maps. Given a ‘direct mapping’ genotype-phenotype map that does not alter the existing landscape (Non modular, blue columns), the fitness landscape is highly rugged, but with a clear structure: all fit or neutral genotypes are 4 mutations apart, because each 4-enzyme pathway requires 4 simultaneous mutations to redeploy. Given a hierarchically modular, aligned genotype-phenotype map (red columns), single mutations can have hierarchical effects on the phenotype, moving whole integrated 4-enzyme pathways as single units. This removes the need for four simultaneous mutations to redeploy 4-enzyme pathways. As a result, the landscape is transformed, condensing the distributed pattern of fit phenotypes into the local neighbourhood, and removing its ruggedness. This illustrates that a modular aligned genotype-phenotype map uses a heuristic to exploit existing structure in the fitness landscape to remove ruggedness and enrich the local neighbourhood with fit, modular reorganisations. Finally, given a modular genotype-phenotype map in which genotypic changes are not hierarchically aligned to phenotypic modules, the fitness landscape is transformed to be even less hospitable than having no map at all (green columns). Top right: Zoomed in section of the same chart for the local mutational neighbourhood.

Chapter 2: Combinatorial Exaptation

The blue columns in Figure 10 show the distribution of phenotypes that are as fit or fitter than the current phenotype, O_{test} , as we move away from O_{test} in genotype space. It is clear that there is a distinct pattern to this distribution: neutral or fitter phenotypes only occur at multiples of 4 mutations away. This is because at this stage of evolution, the only fitter or neutral forms are those that involve modular redeployment of whole 4-enzyme pathways, which with a non-modular gene regulatory network, requires 4 simultaneous mutations. To understand how a modular, aligned genotype-phenotype map resolves this problem, we then replaced the genotype-phenotype map of O_{test} with a modular, aligned genotype-phenotype map that enabled redeploying 4-enzyme pathways with single mutations (see methods), and then took a new snapshot of the distribution of phenotypes in the genetic neighbourhood (red columns, Figure 10). The results graphically illustrate how a modular, aligned genotype-phenotype map reorganises genotype space: It uses inherent structure in the fitness landscape to systematically remove its ruggedness, enriching the local neighbourhood with fit modular reorganisations. Finally, to observe the effect of genotype-phenotype map alignment on the fitness landscape, we replaced the genotype-phenotype map of O_{test} with a modular genotype-phenotype map, but where the modules were purposefully misaligned with phenotypic modules (green columns, Figure 10). The results show that having a misaligned modular genotype-phenotype map actually decreases evolvability in the fitness landscape, shifting the distribution of fitter phenotypes to a greater distance away.

2.4.6 Supplementary results

Experiments were conducted to test the sensitivity of the results to simulation parameters, keeping all other parameters fixed (Figure 11). Parameter values were set to those described in all other experiments (population size $S=50$, truncation point $L=10$, and mutation rate $Pm=0.006$) unless described otherwise, and with a fitness function, selection pressures and genotype-phenotype map as described in S2 experiments. Each data point represents the mean of 10 simulations, and error bars represent standard deviation. Simulations were also carried out using a hill-climber algorithm rather than a population for all experiments. This had no qualitative effect on the results when compared to the results with a population.

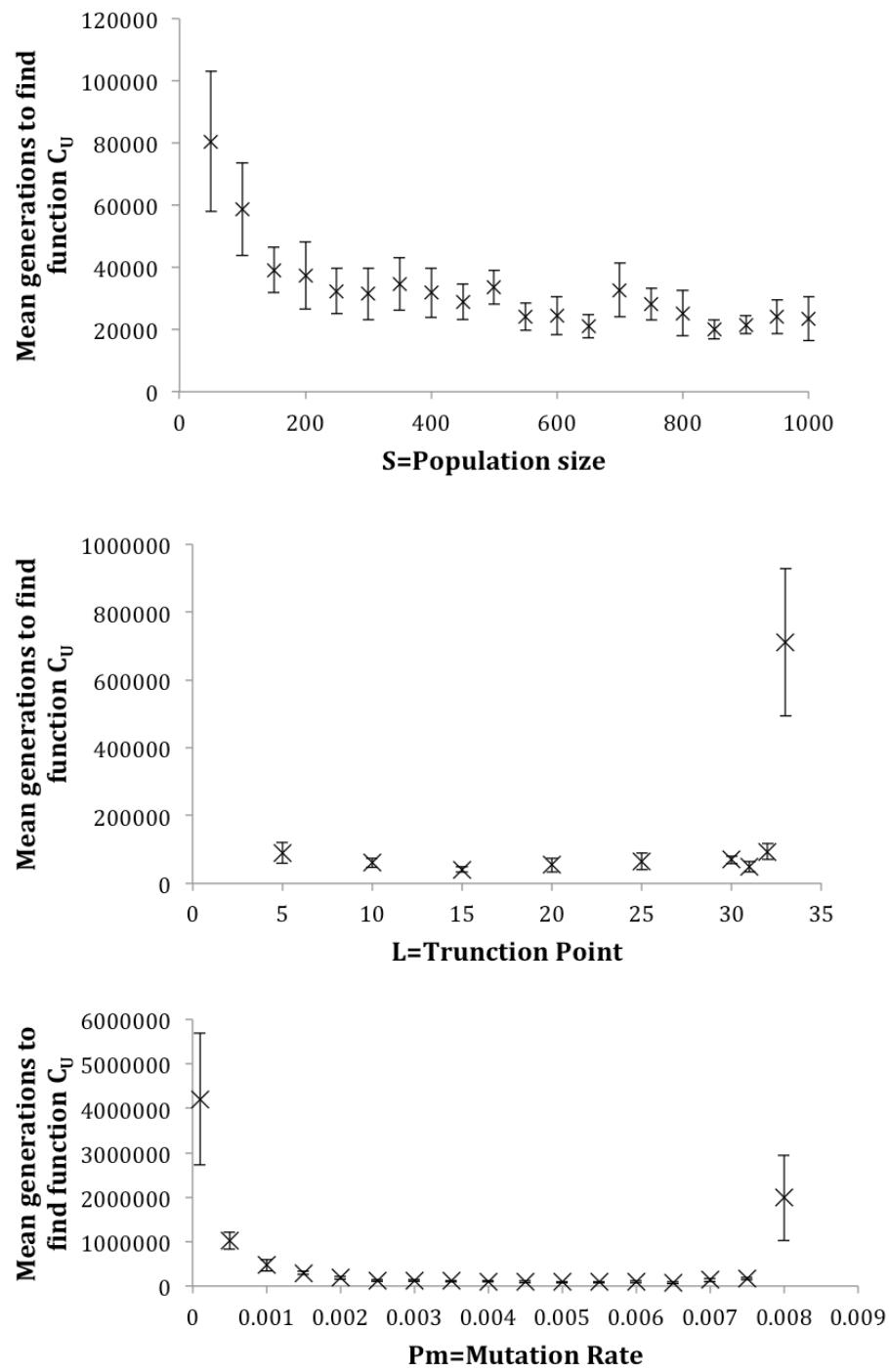


Figure 11. Charts showing the sensitivity of S2 results to simulation parameters.

2.5 Conclusions

In summary, by analysing existing theories of evolution of complex functions, and in particular, evolution by combining functions, we defined a new theoretical framework for evolution by combining functions. This framework builds numerous bridges between the largely separate theories of exaptation, building-block evolutionary algorithm models, and tinkering. It identifies a new type of exaptation (for which we suggest the term ‘combinatorial exaptation’) that we propose is the central mechanism behind processes of evolution by combining functions.

We developed a transparent computational model of combinatorial exaptation to complement our theoretical framework. This model supports previous work showing that joining functions requires some extra evolvability machinery beyond that of ENS alone. Specifically, we showed that evolution requires some mechanism of encapsulation that allows new emergent modules to be reorganised in the phenotype as a single, coherent unit – and that this need is caused by the actions of burden/internal selection. The model shows that an aligned, modular genotype-phenotype map can enable this by removing systematic ruggedness in the fitness landscape, thus creating a local genetic neighbourhood of fit, modular phenotypic reorganisations – importantly, some of which may produce emergent higher level functionality, and hence allow combinatorial exaptation.

The model also shows that combinatorial exaptation can occur recursively, scaling across multiple levels of organisation, if genotype-phenotype map evolves in parallel with combinatorial exaptation in the phenotype, enabling continual, recursive reorganisation of the fitness landscape. We then showed that this problem can be resolved if there is some factor that causes the evolution of gene regulatory network structure and phenotype structure to be (to some extent) causally linked, because in this case new regulatory modules can potentially be formed in concert with new phenotypic modules. We then illustrated that this can occur if development is constrained to occur in a simple, decentralised manner of local growth, as commonly occurs in natural organisms. As an aside, this therefore also provides a new mechanism to explain the modular nature of biological gene regulatory networks that does not rely on a modularly varying environment.

In general, this work implies that a sufficient mechanism of encapsulation could be an important and under-represented component from the current theory of evolution that helps enable the evolution of complexity in nature – and in particular, that a decentralised process of ontological development may be sufficient to provide such a mechanism.

Although we have presented a possible mechanism by which combinatorial exaptation can occur recursively, and given some brief argument as to how this might come about in nature, it would be useful to model this in more detail, where development occurs more explicitly, thus allowing comparison of how different types of development affect the evolvability of the resulting gene regulatory networks. It would also be interesting to explore the effects of combinatorial exaptation when selection pressures are not all present in a single environment, but are spread over a heterogeneous spatial environment.

2.5.1 Key Results

The key claim of this chapter is that

As joined functions become more complex, increasing ‘burden’ in the form of ‘internal selection’ places limits on evolution by combining functional building blocks, but an ontological development process that occurs by local growth, as present in most complex biological organisms, can resolve this problem allowing building blocks to be combined recursively over multiple levels of organisation in a scalable fashion.

This claim is supported by the following results:

- Evidence that ‘burden’ in the form of ‘internal selection’ places limits on evolution by combining building blocks is provided by S1 simulation results, in which the 4-enzyme metabolic pathways were evolved, but were never able to be combined to find the complex target function, and illustrated in Figure 6.
- Evidence that an ontological development process that occurs by local growth can resolve this problem is provided by S2 simulation results, in which evolution with such a development process was

consistently able to evolve the target 16-enzyme metabolic pathway by combining level 2 components, and illustrated in Figure 6.

- Control experiments showing that such an ontological development process is the key factor that enables evolution of the target 16-enzyme metabolic pathway, (and that simply allowing the gene regulation network to evolve its own hierarchical structure given a top-down development process is not sufficient) is provided by simulation C1.

2.6 Methods

Genome structure. (S1, GP1): Organisms had 16 structural genes, each encoding one enzyme, and 16 regulatory genes (R_p) that encoded the interactions between those enzymes. Each structural gene had 10 binary loci; an enzyme was evolved when the 10 loci in its structural gene matched a predefined bit string. Each structural gene had a corresponding R_p regulatory gene that determined the position its respective enzyme in functional interaction space. Each R_p gene consisted of two integers, $0 < p_x, p_y \leq 9$. The values of p_x and p_y corresponded to the x and y position respectively of the function of their structural gene in functional interaction space.

(C1): Genome structure in C1 was identical to that in S1 except that each structural gene had one extra corresponding regulatory gene (R_{dtm}). Hence in C1 each structural gene had two regulatory genes (R_p and R_{dtm}), resulting in a total of 32 regulatory genes per organism. R_{dtm} encoded which structure in the phenotype triggered the local developmental circuitry to build the given enzyme during development. It was represented by an integer $0 \leq R_{dtm} < 32$.

(S2): Genome structure in S2 was identical to that in S1 except that each structural gene had two extra corresponding regulatory genes (R_{dtm} and R_r). Hence in S2 each structural gene had three regulatory genes (R_p , R_{dtm} and R_r), resulting in a total of 48 regulatory genes per organism. R_{dtm} encoded which structure in the phenotype triggered the local developmental circuitry to build the given enzyme during development. It was represented by an integer $0 \leq R_{dtm} < 32$. R_r was an integer ($0 < R_r < 5$) that encoded the direction in functional interaction space in which the given enzyme was placed adjacent to its developmental trigger module (see below).

Phenotype and genotype-phenotype map: (S1) The phenotype of each organism was represented by a 9x9 grid (i.e. functional interaction space) that displayed the interaction of its evolved enzymes. Enzymes plotted as a coloured dot in functional interaction space once they were evolved. Enzymes that had been evolved and were neighbouring in functional interaction space were interacting, and hence could possibly form metabolic pathways (Figure 3). Multiple enzymes could not occupy the same position in functional interaction space (mutations that caused this were disallowed). The genotype-phenotype map for S1 simulations is described in Figure 3.

(S2): Phenotype and genotype-phenotype map structure in simulation S2 was identical to that in S1 except for the action of the extra regulatory genes, R_{dtm} and R_p . These genes were included to account for a development process that occurred by local growth, and hence locally triggered developmental circuits. In S2 every enzyme had a developmental trigger module, defined by R_{dtm} . Enzymes were placed adjacent to (i.e. interacting with) their developmental trigger module in functional interaction space. This represented the notion that because development was triggered and then occurred locally, enzymes would be built next to, and hence interacting with, the phenotypic structure that triggered their development (i.e. their developmental trigger module). The direction in which the enzyme was placed with respect to its developmental trigger module was determined by the enzyme's corresponding R_p regulatory gene. The value of R_p (1, 2, 3, or 4) corresponded to a shift in functional interaction space of (+1,0), (-1,0), (+0,1), (-0,1) respectively with respect to the enzymes developmental trigger module. If an enzyme had a corresponding $R_{dtm} > 15$, the developmental trigger module was not another enzyme, and hence functional interaction space position of the enzyme was determined by R_p in an identical manner to S1.

C1: Phenotype and genotype-phenotype map structure in the simulation C1 was identical to that in S1 except that hierarchical regulatory interactions were allowed to evolve in the gene regulatory network. This was achieved by allowing regulatory genes to evolve regulatory interactions with other regulatory genes, in a similar manner to S2: Each enzyme was allowed a single hierarchical parent enzyme to be linked to in the gene regulatory network, not allowing feedbacks. The enzyme then calculated its position in functional interaction space (using its own P_x and P_y genes relative to its parent's position

Chapter 2: Combinatorial Exaptation

in functional interaction space, instead of the origin. (Mutations that caused enzymes to have positions outside of functional interaction space were not allowed). The result was that hierarchical gene regulatory networks could evolve, in that changing the P_x or P_y regulatory genes of a given enzyme could affect multiple other enzymes in a coordinated manner, similar to S2. However, the key difference between S2 and C1 is that in C1 there was no enforcing of local growth, and so, for example, enzyme A that was linked in the gene regulatory network to enzyme B was not forced to develop in the phenotype adjacent to that enzyme, and hence was not placed adjacent to it in functional interaction space. As a result, C1 lacks the inherent link between gene regulatory network structure and phenotypic interaction present in S2.

GP1: The non-modular genotype-phenotype map used in this simulation was identical to that in S1 simulations. For the aligned modular genotype-phenotype map, we began with the non-modular genotype-phenotype map used in S1 simulations, and then for each group of 4 enzymes in a given 4-enzyme metabolic pathway, we assigned one as a ‘master’ enzyme that the others in the group took their positions in functional interaction space relative to. The result was that regulatory mutation of any of these four master enzymes caused systematic redeployment of the other three enzymes in the respective metabolic pathway of the master enzyme. For the misaligned modular genotype-phenotype map, we began with an aligned modular genotype-phenotype map, but then ensured that each of the three enzymes that took their locations from a given master enzyme were not in the same metabolic pathway as the master enzyme.

Mutation and selection: Mutation and selection occurred in same manner in all simulations. Mutation occurred by point mutation, according to the per locus mutation rate $P_m=0.006$. Mutation occurred caused a bit flip for binary loci, and random reassignment to a new value (within the valid bounds) for integer loci. Selection occurred via truncation selection, the population size was fixed at $M=50$ and evolution occurred in a generation manner. Specifically, for each generation, the fitness of each organism was calculated, and the fittest $L=10$ organisms were selected for reproduction, and the remaining organisms were discarded. (If all organisms in the population had equal fitness then organisms were chosen at random from the population for reproduction.) An empty population was then created. Reproduction occurred by randomly

choosing an individual from the L fit organisms and placing a copy in the new population, after which the new organism underwent mutation. This was repeated until the new population reached size $M=50$. In each simulation, evolution continued for 10^6 generations, or until C_u was evolved.

Fitness function:

In all cases, the total fitness, F , of an individual takes the general form:

$$F = \sum_{i=1}^N f_i \quad \text{Equation 1}$$

$$f_i = \begin{cases} 1, & \text{if } |C_i| = 0 \\ A_i \prod_{j=1}^{|C_i|} (f_{C_{i,j}}), & \text{otherwise} \end{cases} \quad \text{Equation 2}$$

where N is the total number of selection pressures present in the environment, f_i is the fitness contribution for the i th function (i.e. that satisfies the i th selection pressure), C_i is the set of components that make up the i th function, $f_{C_{i,j}}$ is the fitness contribution of the j th component of the i th function, and A_i is a Boolean function that specifies the interaction arrangement that the set of components in C_i must satisfy in functional interaction space to operate correctly; $A_i=1$ if satisfied and $A_i=0$ otherwise. Details of the specific selection pressures are detailed in table 2, below.

Selection pressures	$\{s_a, s_b \dots s_u\}$
16-enzyme metabolic pathway	$C_u = \{C_q, C_r, C_s, C_t\}; A_u = \square$
4-enzyme metabolic pathways	$\begin{cases} C_q = \{C_a, C_b, C_c, C_d\}; A_q = \square \\ C_r = \{C_e, C_f, C_g, C_h\}; A_r = \square \\ C_s = \{C_i, C_j, C_k, C_l\}; A_s = \square \\ C_t = \{C_m, C_n, C_o, C_p\}; A_t = \square \end{cases}$
Enzymes	$\begin{cases} C_a = \{x_{a,1}, x_{a,2} \dots x_{a,10}\}; A_a = 1111111111 \\ C_b = \{x_{b,1}, x_{b,2} \dots x_{b,10}\}; A_b = 1111111111 \\ \dots \\ C_p = \{x_{p,1}, x_{p,2} \dots x_{p,10}\}; A_p = 1111111111 \end{cases}$
Structural gene loci	$C_{x_{\alpha,n}} = \{ \};$

Table 2. Description of selection pressures in all simulations.

The fitness function was the same in all simulations. It contained 21, hierarchically organised selection pressures (S_a - S_u) that corresponded to 21 possible fit functions. For each of these functions that a given organism was able to perform, the organism received a (+1) fitness bonus. All functions were non-decomposable; i.e., fitness bonuses were only awarded if the conditions necessary to carry out the given function were met exactly.

The first 16 of these functions (C_a - C_p) were the simplest. They corresponded to evolving individual enzymes. To receive the fitness bonus for evolving any of these 16 functions, all loci in the structural gene corresponding to that particular enzyme had to exactly match a predefined target 10-bit string.

The four next most complex functions (C_q - C_t) were four-enzyme metabolic pathways. To receive the fitness bonus for any of these given pathways, an organism must have correctly evolved all four of its component enzymes and

have them organised into a particular arrangement of interactions in functional interaction space (specifically, a square formation – Figure 5). The final function, C_u , was the target of evolution. To receive the fitness bonus for C_u , organisms were required to have all 16 enzymes evolved and arranged into a particular arrangement of interactions in functional interaction space (specifically, a hierarchical 16 enzyme square composed of four neighbouring smaller squares of 4 enzymes each – Figure 5). This interaction arrangement represented having each of the 4-enzyme metabolic pathways (C_q - C_t) arranged in a particular neighbouring arrangement.

Chapter 3: Homeogenesis

3.1 Introduction

In this chapter we temporarily change our focus from the general question of how evolution evolves complexity to instead focus on exploring the properties and behaviours of a particular boundary case of combinatorial exaptation.

Although this is a sidestep from the general question of complexity, it addresses some other deeply related open questions, and will help to provide a foundation for further theoretical work on the origins of complexity in chapter 4.

The particular phenomenon that we focus on this chapter is that many biological organisms contain conserved internal conditions within their metabolisms that appear to correspond to ancient environments in which their ancestors lived. For example, the chemistry of the cell interior is thought to be comparable to the early oceans, or geothermal vents, in which life began (Macallum 1926; Mulkidjanian et al. 2012). So far, however, there has been little discussion about how such internalised ancestral environments are incorporated during evolution, or why they are preserved. Here we attempt to address this problem. We present a hypothesis that a form of combinatorial exaptation can cause internalised ancestral environments, and hence could help to explain this phenomenon. We support this hypothesis with biological examples and a computational model of metabolic network evolution.

3.2 Background

In 1926, Archibald Macallum noted that although many organismal fluids, such as blood and lymph, have similarities with seawater, indicating that the first animals emerged in the sea (Mulkidjanian et al. 2012; Macallum 1926), the inorganic composition of the cell cytosol dramatically differs from that of modern sea water. Macallum thus insightfully reasoned that “the cell... has endowments transmitted from a past almost as remote as the origin of life on earth” (Macallum 1926). Macallum’s insight has been summarised as a ‘chemistry conservation principle’: the chemical traits of organisms are more conservative than the changing environment and hence retain information

Chapter 3: Homeogenesis

about ancient environmental conditions (Mulkidjanian et al. 2012; Mulkidjanian and Galperin 2007). For example, the highly reduced state of the cytoplasm, even in organisms that dwell in oxygenated habitats, indicates that the major biochemical pathways were fixed before the atmosphere was oxygenated as a result of the activity of cyanobacteria approximately 2.4 billion years ago, so that substantial modification of these pathways in response to the oxygenation of the atmosphere was impossible (Mulkidjanian et al. 2012). Instead, cellular life forms have evolved numerous energy-requiring membrane transport systems to sustain redox and electrochemical gradients between their interior and the environment (Mulkidjanian et al. 2012). Thus Macallum's work has resulted in a view that a major trend in evolution is the development of increasingly sophisticated mechanisms whereby the internal environment is protected from the external world (Gross 1998). This idea of a conserved internal environment over evolutionary timescales has echoes of homeostasis – indeed, Macallum was strongly influenced by the work of Claude Bernard, who first described the concept (Bernard 1865; Gross 1998). However, unlike homeostasis, which is controlled by behaviours and small functional changes, and occurs in the lifetime of a single organism, Macallum's chemistry conservation principle occurs over multiple generations, and is controlled by largely unknown factors within evolution. What causes Macallum's chemistry conservation principle, or how it results in the incorporation of internalised ancestral environments during evolution is poorly understood.

The main objective of this chapter is to present and investigate a hypothesis that a particular type of functional adaptation based on combinatorial exaptation – that we term *homeogenesis* – could be responsible for internalised ancestral environments in some biological organisms. In short, homeogenesis is similar in concept to homeostasis, in that it is a biological process that maintains a constant environment within the organism – but unlike homeostasis, it occurs over evolutionary time, and operates by evolving extra functionality for the phenotype to maintain its conditions, as opposed to maintaining conditions by functional changes that occur within a single phenotype (e.g. temperature regulation, pH regulation etc.; Cannon 1935; Cannon 1929; Bhagavan 2002).

The remainder of this chapter has been divided into 4 parts. In part 3.3 we give a verbal description of the mechanism of homeogenesis, and describe how it can systematically cause environmental internalisation. In part 3.4 we present the evolution of C₄ photosynthesis as a case study for evidence that homeogenesis has caused internalised ancestral environments in biological evolution. In part 3.5, we concretise the verbal model of homeogenesis by providing a transparent, abstract computational model that illustrates some conditions in which internalised ancestral environments are incorporated in evolution. Finally, in part 3.6, we draw conclusions and discuss possible further work.

3.3 An extreme and simple example of homeogenesis

To address the question of how internalised ancestral environments could be generated in evolution, in this section we will describe the process of adaptation by homeogenesis and examine its capability for generating internalised ancestral environments.

Homeogenesis is best understood by considering an extreme and unrealistic case in which the dynamics are very clear. Imagine a simple organism whose metabolism requires a certain set of chemical inputs. An example could be a bacteria living in a hydrothermal vent. Suppose that its metabolism has so many internal dependencies that it is very difficult to change without breaking it. Now suppose that the makeup of the environment changes slightly, but enough that the organism's metabolism will no longer function given the new inputs. In this worst case, there is no reasonable evolutionary path to change the existing metabolism to make it work with the new inputs, because any small change breaks dependencies causing the metabolism to break. A better option is to leave the existing metabolic functionality alone and instead change the metabolic *inputs* – i.e. the environment – back to a state in which the existing metabolism can use them (e.g. change them back to their previous state).

One way to achieve this is by altering the external environment – that is, by ‘niche construction’ (Odling-Smee, Laland, and Feldman 2013). However, in many cases this will not be possible. For example, given the constant flux of materials in our hydrothermal vent example, any changes to the external

Chapter 3: Homeogenesis

chemical environment caused by the organism would be quickly swept away. However, there is another way to change the external environment. The organism could evolve a new, internal metabolic function that acts as an internal environmental ‘adapter’ between the external environment and the organism’s existing metabolic network (which is left alone). The new adapter function (for example, a new enzyme) ‘converts’ the new, inhospitable environmental conditions back into the old, hospitable conditions within the organism immediately before they are then used as inputs for the organism’s existing metabolic network. After all, the existing metabolic network already works with this old input, and so providing an adapter will ensure its continued functionality. The principle is similar to taking an electrical appliance abroad, where the electricity voltage is different: rather than changing the fundamental internal workings of the appliance, it is much easier to simply add a voltage adapter to the end of the power cable, providing the existing functionality with the input environment that it ‘expects’. In our biological example, the bacterial organism could evolve a simple catalysis step that internally converts the new, offending chemical constituents back to their previous state so that they can be used with the existing metabolic network. By doing this, the negative effect of the environment change has been nullified, but the existing metabolic network left alone, and the external environment has not been changed.

How does homeogenesis imply internalised ancestral environments?

Importantly, for successful adaptation by homeogenesis, the output of the new ‘adapter’ function must match the relevant conditions of the previous external environment, because these are the conditions required by the unchanged, existing metabolic network. The result is that adding an adapter function to ‘undo’ a recent environment change has the effect of making an internal re-creation of the organism’s previous external environment within the organism’s metabolism. In other words, homeogenesis systematically creates internalised ancestral environments.

To flesh out the details of homeogenesis, we can ask a number of further questions. First: How does homeogenesis compare to existing mechanisms of adaptation? Homeogenesis is subtly different form of adaptation than ‘traditional’ adaptation (in which the existing functionality is changed (Ridley 2009; Orr 2005) – e.g. as described in the classic example of evolutionary adaptation, the peppered moth (Grant 1999), because during homeogenesis

the organism's existing functionality remains unchanged. Nor is homeogenesis niche construction, because during homeogenesis the external environment (external to the organism) remains unchanged. Accordingly, we can estimate when homeogenesis is likely to occur: because it offers an alternative route to evolution when neither the environment nor the existing functionality can reasonably be changed, it stands to reason that homeogenesis is more likely to occur in situations of high environmental and functional constraint.

Finally, how does homeogenesis relate to combinatorial exaptation? In our example of homeogenesis, an adapter function is joined to the existing metabolic network. Although the adapter function is small, and the metabolic network large, this still represents a case of combinatorial exaptation: we combine two functional components in a specific manner of interaction, and they produce a new, emergent functional entity capable of functionality that neither of the individual components were capable of.

In summary, homeogenesis is a mechanism of adaptation that occurs by adding internal adapter functions at the interface between the existing metabolism and external environment. In doing so, it generates internal recreations of environmental conditions within the metabolism. These internalised ancestral environments are then preserved within the organism because they perform necessary metabolic functions, given the new external environment.

The verbal description of homeogenesis provides a conceptual framework that describes how such a mechanism could theoretically occur in biological evolution. To strengthen this case, in section 3.5 we formalise this verbal argument into a computational model that we use to test the hypothesis that evolution under certain types of constraint will result in the incorporation of internalised ancestral environments. A more substantial case for homeogenesis would require examples of the process occurring in biological evolution. In the next section, we will describe evidence that the evolution of C₄ photosynthesis occurred by a process of homeogenesis².

² (We thank Ros Rickaby for the suggestion that C₄ photosynthesis evolution could be an example of homeogenesis).

3.4 A biological example of homeogenesis

Plants assimilate carbon by one of three photosynthetic pathways, commonly called the C₃, C₄ and CAM pathways (Edwards and Walker 1983). The C₄ photosynthetic pathway represents a modification of C₃ metabolism that is most effective at low concentrations of CO₂. C₄ plants are thought to have evolved in response to a reduction in atmospheric CO₂ that began during the Cretaceous (~100 million years ago) and continued until the Miocene (~20 million years ago; Ehleringer et al. 1991). Carbon fixation in C₃ plants occurs via a cycle of chemical reactions called the Calvin cycle. At low CO₂ concentrations the Calvin cycle becomes increasingly inefficient due to increased photorespiration – an unwanted alternative reaction pathway that apparently has no useful function (Edwards and Walker 1983). Thus the long-term reduction in atmospheric CO₂ represented a serious problem for C₃ plants (Ehleringer et al. 1991).

Here niche construction is not possible; a single plant cannot hope to change the global CO₂ concentration. Another potential solution to this problem is traditional adaptation – i.e. changing the existing photosynthesis metabolism. If it were possible to fundamentally alter the Calvin cycle to somehow be more efficient at low CO₂ concentrations, this would solve the problem of photosynthesis with decreasing CO₂. But the Calvin cycle is a complex cyclical chain of contingent chemical reactions; it is deeply constrained by what alternative reactions are available by mutation, or even possible according to chemistry. Furthermore, it is a process of carbon fixation, and CO₂ is the source of this carbon. It is in this sense highly dependent on the concentration of CO₂ (Ehleringer et al. 1991; Edwards and Walker 1983).

Irrespective of whether it is possible to change the Calvin cycle in this way, this was not the solution adopted by evolution. Rather than alter the Calvin cycle, the C₄ pathway instead added a new cycle of reactions that sit ‘in between’ the external low CO₂ environment and the Calvin cycle. These new reactions have the effect of dramatically increasing the CO₂ concentration internally within the leaf, thus providing a new, high CO₂ internal ‘input’ environment for the normal C₃ photosynthetic cycle (Edwards and Walker 1983; Ehleringer et al. 1991; Figure 12).

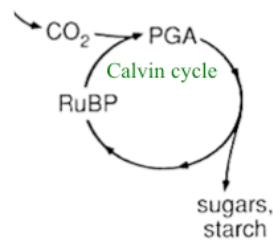
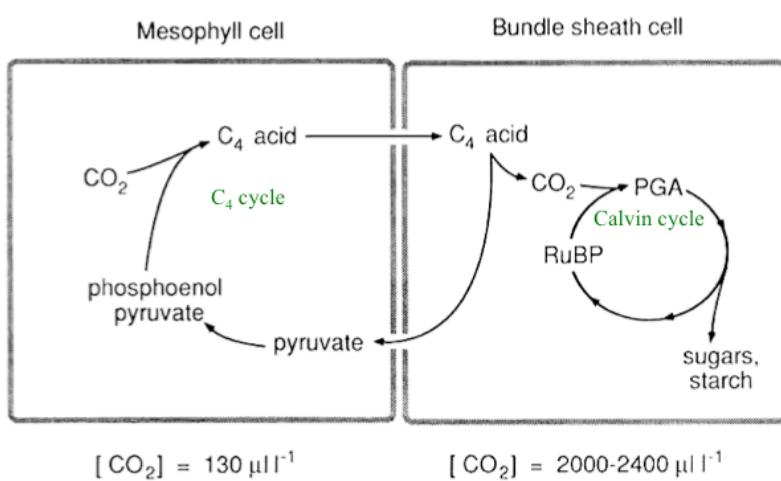
C₃ Plants**C₄ Plants**

Figure 12. The C₃ and C₄ photosynthetic cycles. The C₃ cycle evolved first. It is a simpler process, but is less efficient at times of low atmospheric CO₂ concentrations. The C₄ cycle evolved in response to a long-term reduction in atmospheric CO₂ concentrations. It introduces a new cycle of chemical reactions that change the input to the existing C₃ cycle, while the C₃ cycle itself is left unchanged. The new cycle of reactions have the effect of dramatically increasing the input CO₂ concentration to the C₃ cycle, making C₄ plants more efficient at low atmospheric CO₂ concentrations.

Effectively, C₄ plants responded to environment change by evolving a new environmental ‘adapter’ that recreated their previous environment internally - and then used this internalised environment as an input to their existing functionality. C₄ plants undertook homeogenesis. As a result, they contain an internal record of a previously inhabited environment, stored within their metabolism (Figure 13).

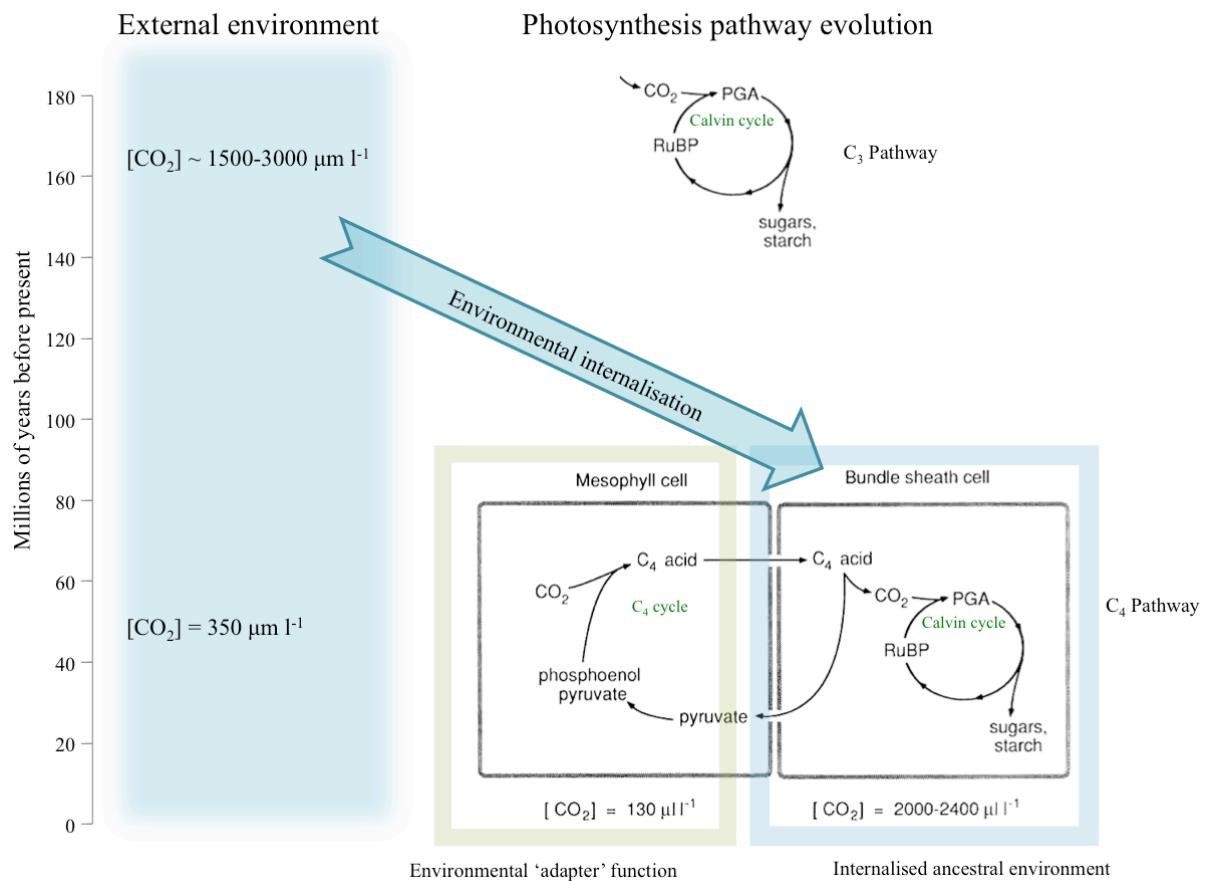


Figure 13. C_4 photosynthesis evolution as a process of homeogenesis causing environmental internalisation. The new chemical cycle evolved in C_4 photosynthesis (green box) increases the CO_2 concentration internally within the leaf, providing new, high CO_2 input environment (blue box) for the existing C_3 cycle. This occurred in response to decreasing atmospheric CO_2 concentrations. C_4 plants evolved machinery to recreate a past fit external environment (i.e. with a higher concentration of CO_2) internally within the leaf. This is therefore an example of environmental internalisation.

3.5 A model of metabolic evolution

3.5.1 Aim

We have simulated a simple and extreme example of this kind of interaction between environment change and evolution. The aim of this model is to provide a concrete illustration of homeogenesis, and to explore its properties.

3.5.2 Methods

The model is of simple organisms (e.g. bacteria) evolving their metabolic networks to cope with environment change in a spatially heterogeneous environment (e.g. a hydrothermal vent). Chemistry in the system is based on a simplified chemical reaction network of possible reactions that contains 36 chemical compounds (S_1 – S_{36}) and 120 (one-way) reactions that convert between those compounds. The network is laid out in a grid formation. (Figure 14).

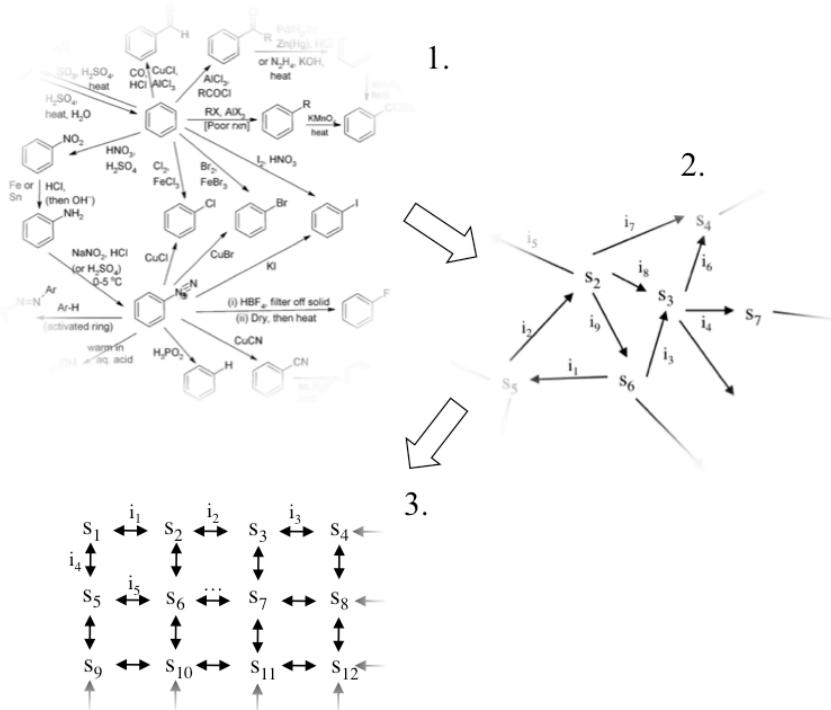


Figure 14. Figure illustration of the chemical reaction network abstraction in the model. Chemical reaction networks define which chemical reactions are possible according to reaction chemistry. (1) An example section of an organic chemistry reaction network. (2) Such networks are commonly abstracted to a network of states (e.g. compounds, $s_1, s_2 \dots s_n$) and reactions ($i_1, i_2 \dots i_n$) that determine transformations between states. (3) In the model we use a simplified chemical reaction network that can be described as a matrix of compounds and reactions.

All organisms were based on the same, fixed replicator (analogous to how all known life has DNA as a central replicator) that requires a specific set of 18

Chapter 3: Homeogenesis

chemical compounds to reproduce. This represents a fixed output requirement for all organisms' metabolisms that remains the same regardless of the chemical makeup of the organisms' external environment. Organisms evolved across a heterogeneous spatial environment, consisting of 7 neighbouring niches (N_1-N_7) arranged in a line. Each niche contained a different set of 18 chemical compounds from the 36 available. Most niches (all but the initial niche, N_1) did not contain the specific 18 compounds required by the organisms' replicator. As a result, to survive in each niche, organisms had to evolve a suitable metabolic network that could produce the chemical requirements of its replicator from the compounds in the external environment.

Each organism had a linear genotype containing a variable number of genes. Those genes specify a metabolic network in the following manner. Each gene codes for one of 120 possible catalysts (one per possible chemical reaction). The genome is transcribed in order, one gene at a time, proceeding along its length. The set of chemical compounds in the organism's niche is used as the input to its metabolic network. As each catalyst is transcribed, the specific chemical reaction that it enables is carried out on this set of compounds, if the input molecule is present. Thus the organism's metabolic input is therefore changed, sequentially, by the sequence of catalysts that the organism's genotype produces. The result is a metabolic network of sequential chemical reactions.

If the output of the organism's metabolic network was the specific fixed target set of 18 compounds required by the organisms' replicator, then the organism could survive and reproduce in that niche. Fitness was calculated according to:

$$F = \max (0, F_s - F_n)$$

Where F is the fitness of the organism, $F_s=40$ if the organism's metabolic network produces the target set of compounds and 0 otherwise, and F_n is the number of genes expressed in the phenotype, representing the energetic cost of producing each catalyst. Each gene also contained a binary switch that determined whether or not it was transcribed. Genes that were not transcribed did not incur this energetic cost.

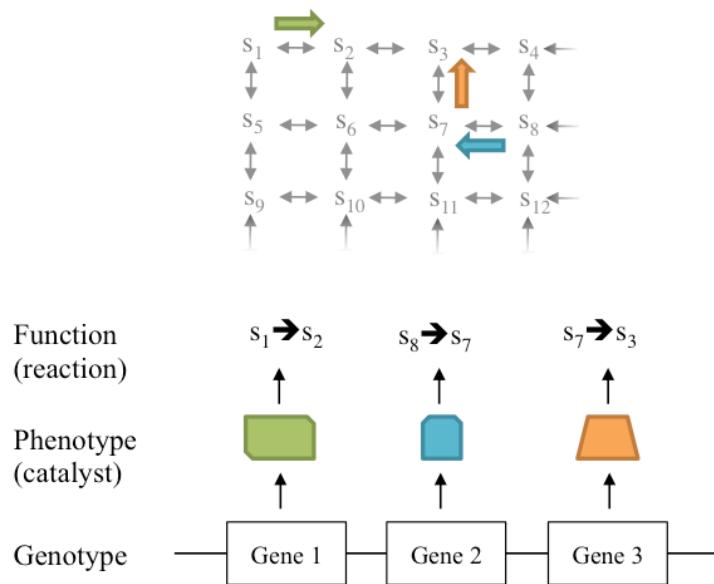


Figure 15. Mapping of genotype to phenotype in the model. Each gene encodes a different catalyst, and each catalyst catalyses a single (unidirectional) chemical reaction from the chemical reaction network.

Each niche had a maximum carrying capacity of 200 organisms. Each simulation began with a single organism in niche N_1 , situated at the end of the chain of 7 niches. N_1 always contained the 18 chemical compounds required by the replication machinery. Accordingly, no metabolic network was required to survive and reproduce in N_1 . The initial organism began with an empty genome. New niches were occupied by there being a fixed probability of 0.2 that any given organism's offspring would be placed in a randomly chosen (with uniform probability) neighbouring environment.

Each gene i was represented by an integer ($0 < X_i < 120$) that corresponded to a specific chemical reaction in the network of possible reactions (defining which reaction the gene catalyses) and a binary genetic switch B_i , whose value determined whether or not the gene was transcribed (0=not transcribed, 1=transcribed).

Chapter 3: Homeogenesis

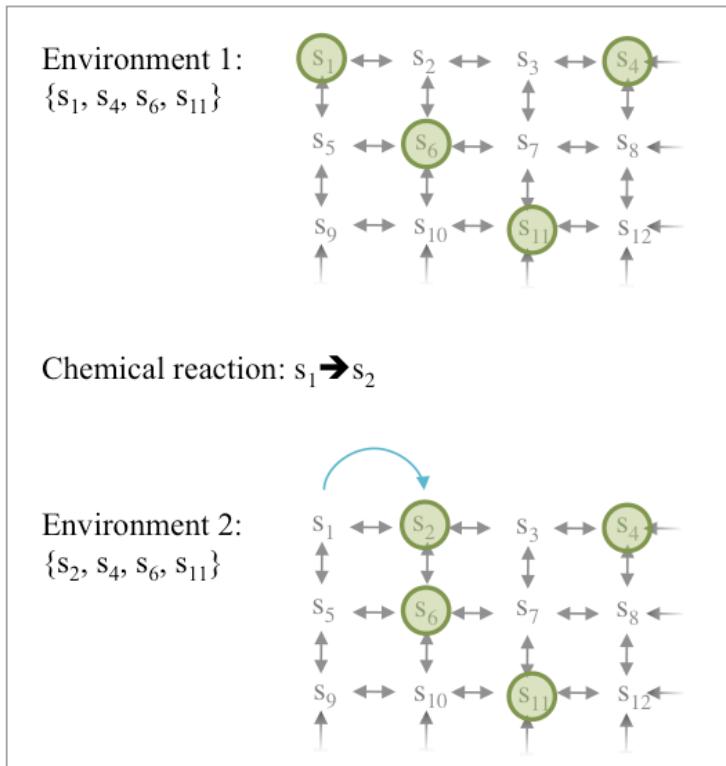


Figure 16. In the model an environment was defined by a particular set of chemical compounds (green circles are present compounds; in the environment 1, compounds s_1, s_4, s_6 and s_{11} are present). During development, genes are read in sequence from the genome. If the chemical reaction that the catalyst made by the current gene is possible, given the current environment (i.e. if the relevant compound is present), then the reaction will occur and environment change will have occurred. For example, given environment 1 in the presence of a catalyst that catalyses the reaction $s_1 \Rightarrow s_2$, compound s_1 is present in the environment 1, and so this reaction occurs, resulting in environment change, the result of which will be environment 2.

There was a fixed per-genome mutation rate of $R_m = 0.01$. With equal probability, mutation either (a) added a new gene in a random position in the genome, (b) removed a random existing gene, or (c) randomly altered an existing gene. Gene alteration involved, with equal probability, either randomly selecting a new value for X_i with uniform probability, or performing a bit flip on B_i . During each generation, for each niche, if there were any organisms in that niche with non-zero fitness, then those organisms were selected for reproduction according to (linear) fitness proportional selection, until $n=200$

offspring had been produced. All non-offspring organisms were then discarded.

Simulations were conducted using two extreme types of environment change: Serial, where sequential environment changes are a series of interdependent events, such as a chain of reactions (e.g. $A \Rightarrow B$, $B \Rightarrow C$, etc.), thus making the difference between adjacent niches interdependent, and parallel, where sequential environment changes are not dependent on each other (e.g. $A \Rightarrow B$, $X \Rightarrow Y$), making the difference between adjacent niches independent of each other.

3.5.3 Results and discussion

3.5.3.1 **Simulation 1: Metabolic pathways evolved via a sequence of environments contain more ancestral environments than random pathways**

We used the simulation to explore homeogenesis in two main ways. First, we looked to test the extent to which homeogenesis caused and preserved internal ancestral environments. To achieve this, we defined a sequence of gradual, serial environment change between niches 1-7 (Figure 17). The shortest possible metabolic pathway that could convert the set of chemical compounds in niche 7 into those in niche 1 (i.e. the target set of compounds) and hence allow survival and reproduction had 6 reaction steps. There were 20 possible 6-step pathways. We began the simulation with a single organism in niche 1. After all niches were populated, we then sought to measure the mean number of internalised ancestral environments – that is, steps in those organisms' metabolic pathways that corresponded to the precise chemical makeup of niches they had previously visited (i.e. niches 2-6). We did not include niche 1 or 7 as internalised ancestral environments because all viable organisms would by definition contain the chemical makeup of niche 1 in their metabolisms (because it was the target) and niche 7 was the external environment. As a control, we analytically calculated the expected number of internalised ancestral environments contained by a randomly selected metabolic pathway from the 20 possible pathways. To obtain a value for evolved pathways, we carried out evolution from a single organism in niche 1; in all runs, we waited until 1000 generations after niche 7 had been populated,

Chapter 3: Homeogenesis

and then measured the mean number of internalised ancestral environments of all organisms with 6-step metabolic pathways in that niche. We carried out 100 repetitions of this simulation and averaged the result across these repetitions. Results are displayed in Figure 18. We found that evolved organisms generally had a much greater number of internalised ancestral environments than found in random viable pathways of the same length (evolved pathways = 4.47, standard error 0.064; analytical expectation from random pathways = 1.4; $P < 0.001$ one sample Student's t-test).

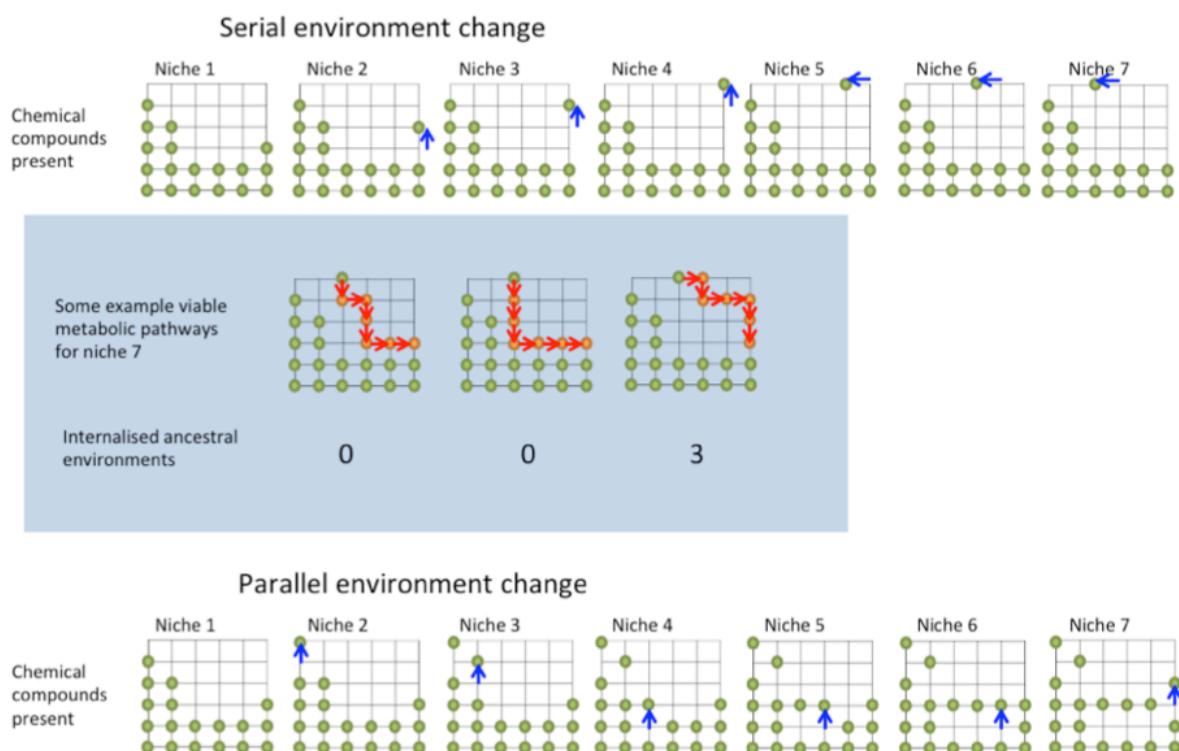


Figure 17. Diagrams showing the two spatial environment sequences used in the model. Each grid represents the chemical reaction network of available reactions, and the green circles are chemical compounds that are present. As such, each grid represents a specific set of chemicals present in a given environment. With serial environment change, only one chemical compound is changed as the environment is traversed, but that compound becomes progressively further away from its original location on the chemical reaction network. The blue box illustrates 3 of the 20 possible 6 step pathways that can convert the environment of niche 7 back into niche 1 (and hence allow survival and reproduction). In the far right example, three of the intermediary steps correspond to the chemical

makeup of environmental niches previously visited (niches 2, 3 and 6), which are therefore counted as internalised ancestral environments. The effects of parallel environment change was compared to the effects of serial change in simulation 2.

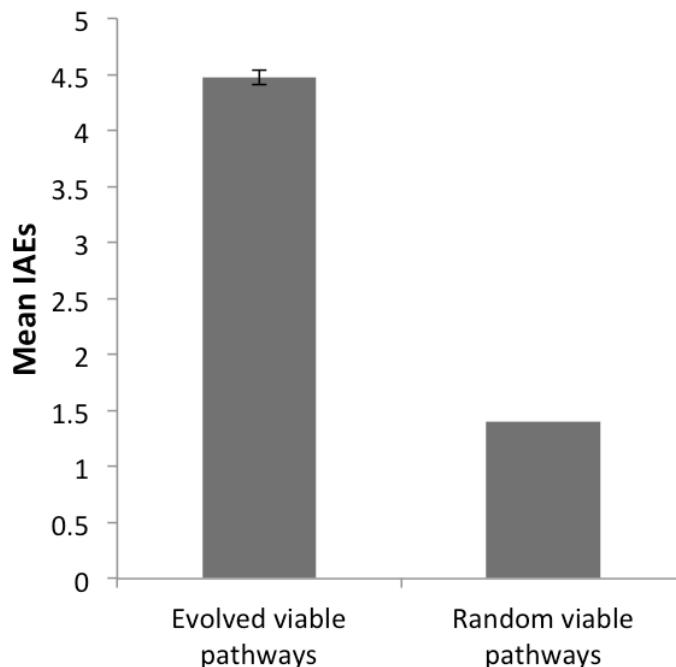


Figure 18. The mean number of internalised ancestral environments in evolved viable pathways is much greater than the analytically calculated expected value of a randomly selected viable pathway of the same length ($P<0.001$, one sample Student's t-test). This chart compares 6-step pathways viable in niche 7; the value for evolved pathways is the mean in the population 1000 generations after niche 7 was populated, averaged over 100 simulations. The random pathway value is the expectation (i.e. mean) internalised ancestral environments of all 20 viable 6 step pathways in that niche. Error bars show standard error. There are no error bars for the expectation from random variable pathways because this is an analytical result.

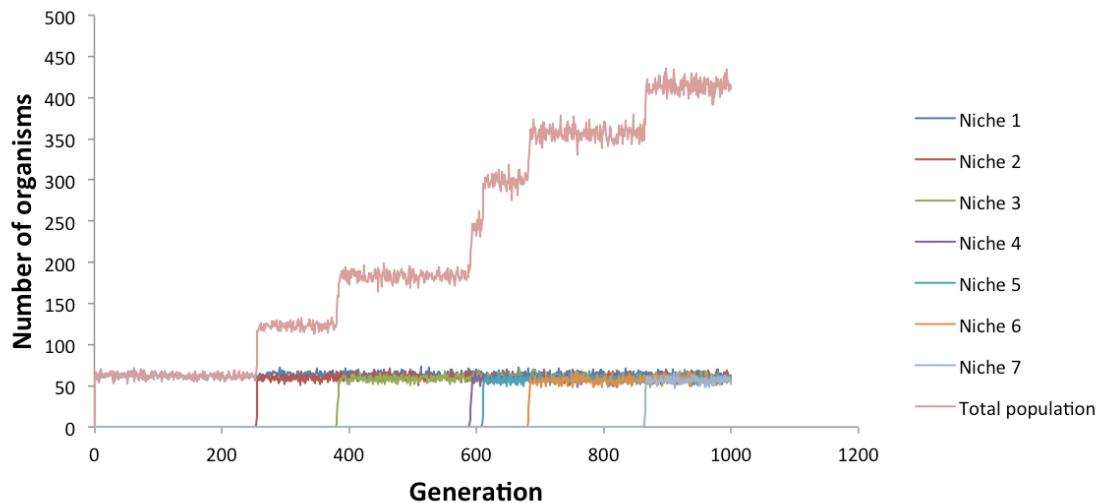


Figure 19. Changes in population size over time in the case study population.

Population size (of organisms with >0 fitness) increases in a stepwise manner as successive niches are colonised in events of adaptive radiation.

This simple illustration shows that non-decomposable functions that are evolved by homeogenesis are likely to preserve the environmental conditions over which they evolved, thus potentially forming a record of past environmental conditions.

A case study population of this simulation is described in Figure 19 and Figure 20. In the case study, organisms with an empty genome filled the initial environment (niche 1) to its carrying capacity (200) within one generation. Of this total population in each niche, only a subsection (60-70 organisms, on average) had >0 fitness, and hence were capable of reproduction. At all times, competition makes empty niches potentially attractive. Initially niche 2 is empty, and organisms that are placed into that niche by migration are very unlikely to be able to replicate given the different chemical environment of this niche, and so it remains uncolonised. Eventually (at generation 442) a new mutant is produced that can survive in niche 2, and there is a rapid radiation as the offspring of this new mutant colonise niche 2. The successful mutant organism achieves this by evolving a catalysis step that converts the single chemical compound in niche 2, not present in niche 1, back into the compound that was present in niche 1, but not present in niche 2. In other words, the mutant evolves a mediated chemical reaction that ‘undoes’ the environment change that has occurred between niche 1 and niche 2. This new catalysis step

in the organism's metabolism thus acts as an 'adapter', converting the external environment into one in which the organism's replicator can reproduce. By doing so, the organism therefore recreates its previous environment (niche 1) internally within its metabolism. A similar process occurs sequentially across all environments, with successive events of adaptive radiation and evolution of new genes. This builds up a metabolic pathway of previously internalised environments. This pathway sits 'in front' of the replication machinery.

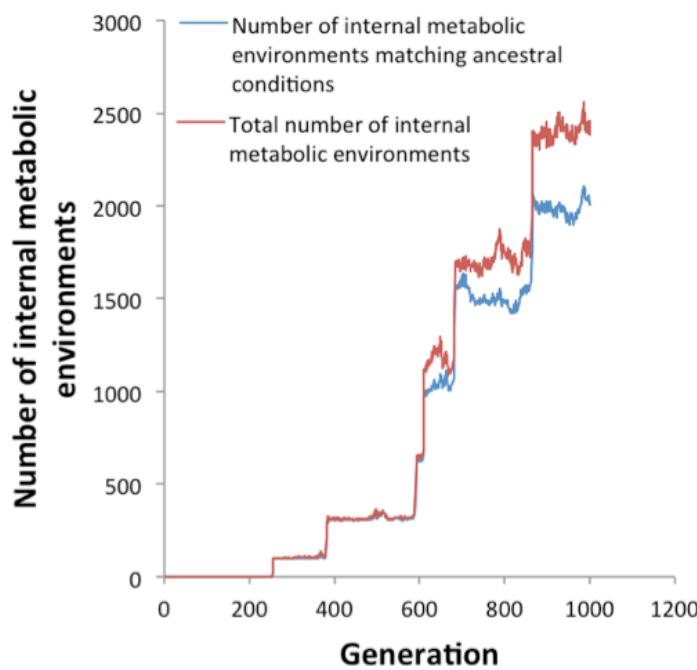


Figure 20. Total number of intermediary steps in metabolic pathways in all organisms in the case study population (red line), and the number of these steps that correspond to previously visited environments (i.e. other niches; blue line). As new niches are colonised, organisms require metabolic pathways with more steps. This increases the number of viable pathways available, hence increasing the likelihood that pathways different from that defined by the previously visited sequence of niches will be evolved, and hence decreasing the proportion of internalised ancestral environments.

At each event of homeogenesis, the new adapter is effectively functionally combined with the existing functionality (including any existing adapters) by combinatorial exaptation, resulting in a single, novel composite function. As the environment becomes increasingly different from the initial environment, this forces evolution to add progressively more steps to its adapter pathway to

cope with the larger transformation required, thus becoming progressively more complex.

3.5.3.2 Simulation 2: Serial environment changes are much more likely to be preserved in sequence than parallel environment changes

We next looked at whether homeogenesis can store *sequences* of ordered environment change within phenotypes, and under what conditions this is likely to happen. Specifically, We aimed to test the hypothesis that serial sequences of environment change, because they are a sequence of dependent events, would more likely be conserved as ordered sequences of internalised ancestral environments in the metabolic network than parallel sequences of environment change.

To achieve this, we conducted simulations using two different sequences of environment change: one serial, and one parallel (the environments used are displayed in Figure 17; similar results were found for many different arbitrary sequences of serial and parallel change). In both cases, we measured the frequency in the population of organisms that contained the precise sequence of internalised chemical environments corresponding to the sequence of environments visited encoded into their phenotypes (i.e. a sequence of ordered catalytic steps in the phenotype that correspond to niche7⇒niche6, niche6⇒niche5, niche5⇒niche4, niche4⇒niche3, niche3⇒niche2, niche2⇒niche1). We carried out 100 simulations each for serial and parallel environment change, and measured the frequency with which such a phenotype occurred in the population. Mean results are displayed in Figure 21.

We found that the specific sequence of environment change is much more likely to be preserved when environment change is serial as opposed to parallel (comparing frequency of preserved sequence in populations at 2500 generations, $P<0.001$, Student's t-test). The simple explanation is that serial chemical changes are contingent on previous reactions, and hence are dependent. Therefore, reordering of sequences in the phenotype by genetic drift is strongly selected against, because reordering of a contingent chemical reaction chain will very likely cause it to cease functioning. In contrast, parallel sequences of environment change are not sensitive to the order in which they occur, and hence are readily reorganised in the phenotype by genetic drift with no effect on fitness.

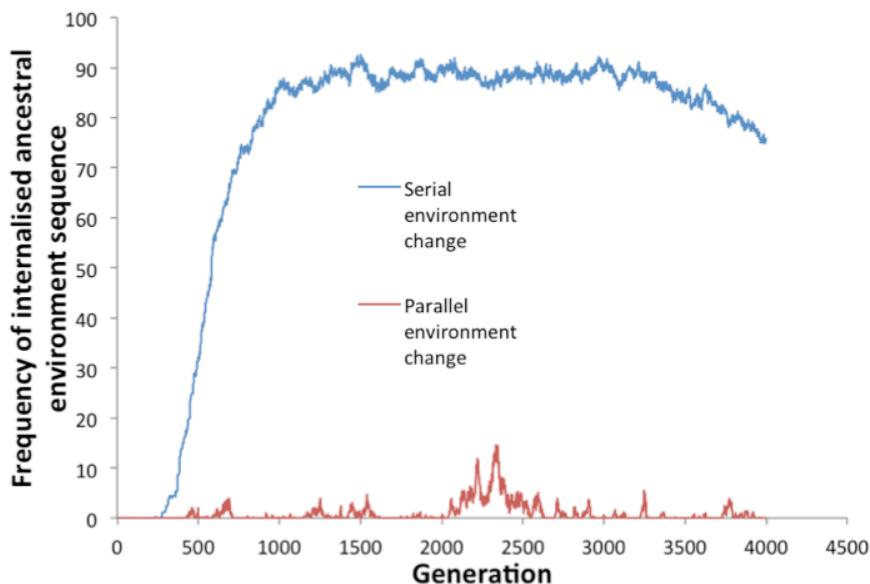


Figure 21. This chart shows the frequency of metabolic networks that contain the specific sequence of internalised environments that represent the sequence of previous environments visited (i.e. $7 \Rightarrow 6 \Rightarrow 5 \Rightarrow 4 \Rightarrow 3 \Rightarrow 2 \Rightarrow 1$) for sequences of serial environment change (blue line) and parallel change (red line) in the whole population. Results are the mean frequency in the population averaged over 100 simulation runs for each type of environment change. The chart shows that the *sequence* of environment change is much more likely to be preserved within phenotypes with serial environment change than with parallel change (comparison at 2500 generations, $P < 0.001$, Student's t-test).

For example, if environment change happens in the sequence $(A \Rightarrow B, X \Rightarrow Y)$ then because neither of these reactions are dependent upon each other, then a metabolic pathway that 'undoes' this change such as $(B \Rightarrow A, Y \Rightarrow X)$ can be reorganised in the phenotype to occur in a different order (i.e. $Y \Rightarrow X, B \Rightarrow A$) without affecting the overall chemical transformation. In contrast, an environment sequence that occurs as $A \Rightarrow B \Rightarrow C$ (i.e. serial change) would require a metabolic pathway of $C \Rightarrow B, B \Rightarrow A$ to 'undo' the change. This pathway cannot easily be reorganised in the phenotype: If the reactions are reordered to $B \Rightarrow A, C \Rightarrow B$, then this reaction chain will not work, as $B \Rightarrow A$ cannot occur until $C \Rightarrow B$ has happened, because $C \Rightarrow B$ provides the input material for $B \Rightarrow A$. As a result, the sequence information of serial environment change is much more likely to be preserved in the phenotype than the sequence information of parallel change.

Chapter 3: Homeogenesis

The point of this particular simulation is to (a) illustrate that homeogenesis can, in some cases, generate metabolic networks that represent whole sequences of past environmental conditions, and to illustrate the conditions that affect the likelihood of this occurring.

3.5.4 Wider implications of homeogenesis

The generic nature of models in this chapter implies that homeogenesis could occur in other areas of evolution beyond that of metabolism and chemical reaction networks. For example, one phenotypic character that could be interpreted as homeogenesis is the evolution of a hard, body-encasing shell in response to the introduction of predators (e.g. in shelled gastropods and chelonians). This would represent a physical example of environmental internalisation (as opposed to chemical) in which the previous, predator-free environment has been recreated internally within the confines of the shell. Another example that could be interpreted as homeogenesis is the circulatory system of a large multicellular organism, which internally recreates the oxygen-rich environment of smaller organisms with diffusion based respiration. Accordingly, the mechanism of environmental internalisation could potentially be applied to explaining the more general organisation of adaptations in the phenotype.

Furthermore, many of the features of homeogenesis are similar to the concept of 'counteractive' niche construction, in which organisms change their external environments to recreate previous environments in which they were fit (Odling-Smee, Laland, and Feldman 2013), but homeogenesis occurs within the organism itself, and counteractive niche construction occurs in the external environment. For example, earthworms go to great lengths to recreate an aquatic environment, to which their phenotype is suited, in the soil in which they live (Laland, Odling-Smee, and Gilbert 2008). Some research on counteractive niche construction does suggest that a similar process could potentially occur within organisms, in a manner similar to homeogenesis (Laland, Odling-Smee, and Gilbert 2008). On a wider level, we can speculate that a similar process might be at work in the evolution of technology: clothes, central heating and farming could all be seen as examples of technologies that attempt to recreate ancestral environments (warm, food-rich) to which the human phenotype has spent a long time adapting to, and hence expects. Such

behaviour could again be seen as long timescale counteractive niche construction. Taken together, we can speculate on a range of similar processes that all act to conserve environmental conditions, but that occur on different timescales and in different situations:

1. *Homeostasis* occurs on short timescales, acts within the organism, and conserves internal conditions by organism behaviours, or short-timescale functional changes;
2. *Counteractive niche construction* occurs on potentially longer timescales (potentially multiple generations if the environment modifications are preserved), acts on the external environment, and conserves external conditions by organism behaviours;
3. *Homeogenesis* occurs on long timescales (i.e. evolutionary time), acts within the organism, and conserves internal conditions by evolutionary adaptations.

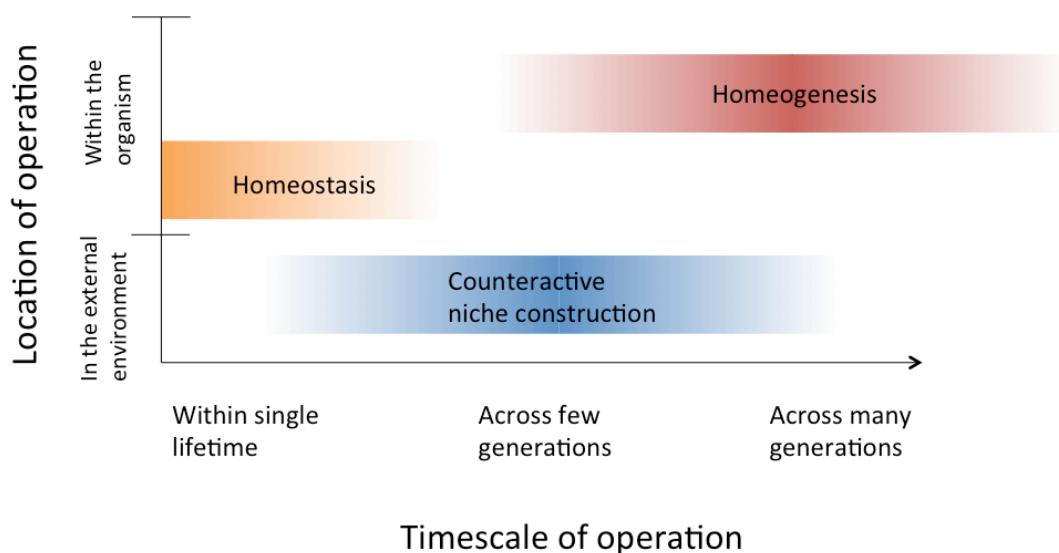


Figure 22. Chart showing the relationships between homeogenesis and other known biological processes in which organism's seek to maintain the environmental status quo.

3.6 Conclusions

In this chapter we have put forward a hypothesis to explain the observation that many organisms appear to contain internal conditions that are similar to the environments in which their ancestors lived. This observation has been summarised as Macallum's chemistry conservation principle – that organism's

Chapter 3: Homeogenesis

internal chemistry is more constrained than the external environment – but few mechanistic explanations have been put forward (Macallum 1926; Mulkidjanian et al. 2012).

We have shown how a boundary case of combinatorial exaptation, that we term homeogenesis, inherently internalises and preserves ancestral environments within the phenotype, and therefore could explain this phenomenon. In brief, homeogenesis occurs when organisms have highly constrained existing functions, and then undergo environment change that is very difficult to counteract by niche construction. Instead, organisms can add some simple, internal ‘adapter’ function to their existing functionality that ‘undoes’ the environment change, thus providing the existing function with the input it expects. In doing so, evolution creates a preserved, internal version of the previous environment. We have supported this hypothesis with a well-studied biological example (the evolution of C₄ photosynthesis, which we show is an example of homeogenesis), and a transparent computational model.

The model uses the domain of metabolic network evolution as an example system. In addition to illustrating the viability of homeogenesis as possible evolutionary mechanism, it provides two main results:

1. First, it shows that metabolic networks that evolve by repeated events of environmental internalisation are much more likely to contain internalised ancestral environments than random viable networks, illustrating the capability of environmental internalisation for internalising and preserving previously experienced environments.
2. Second, the model shows that environmental internalisation can not only preserve individual environments, but can also preserve whole ordered sequences of past environments within metabolic networks, and that this is much more likely to occur if the environment change in question was a sequence of inter-dependent events (i.e. serial change).

This work therefore has significant potential for impact in understanding metabolic evolution. First, it can potentially explain the common but poorly understood phenomenon of organisms apparently containing internalised versions of ancestral environments caused by Maccallum’s chemistry conservation principle. Furthermore, by providing mechanistic support for the

idea that these observed internalised conditions stored within natural organisms really do represent ancient ancestral environments, the results in this chapter also provide much needed support for research that uses these internalised conditions to infer ancient environment conditions (e.g. Mulkidjanian et al. 2012).

Environmental internalisation also has wider implications for evolutionary theory. Although we chose to introduce the model in terms of metabolic network evolution, there is nothing specific to metabolic networks in the model itself. Taken generally, environmental internalisation and the results in this chapter simply describe a mechanism by which complex non-decomposable functions can be evolved by breaking them down into sub components, where (unlike in the previous chapter and model of combinatorial exaptation) the selection pressures necessary to evolve each subcomponent are distributed across a sequence of neighbouring physical environments, as opposed to being in a single environment.

In future work it would be interesting to map out how different levels and types of evolutionary constraints determine under what environmental change conditions evolution is more likely to opt for environmental internalisation over functional change or niche construction. Finally, it would be interesting to explore the implications of environmental internalisation, and in particular, the metabolic model in this chapter, for creating a complexity driver that could perhaps explain some complexity trends. This possibility will be explored in detail in chapter 4.

3.6.1 Key Results

The key claim of this chapter is that

When both the external environment and an organism's existing functionality are too difficult to change, a third possibility exists for evolution: adapting to environment change by adding an internal environmental 'adapter' that converts the new external conditions into those necessitated by the organisms existing functionality – and in doing so, inherently creates an internal replica of the previous environment within the organism's phenotype.

Chapter 3: Homeogenesis

This claim is supported by the following results:

- Evidence that a third possibility for adaptation exists for evolution when both the external environment and the organism's existing functionality cannot be changed is provided by Simulation 1 results (section 3.5.3.1) that show organisms that have a replicator that cannot be changed, in an external environment that cannot be changed, are capable of adapting to novel environments by adding functional adapters to their existing functionality. This is further supported by existing evidence from C₄ photosynthesis (section 3.4).
- Evidence that this process of adaptation (i.e. homeogenesis) creates an internal replica of the previous environment within the organism's phenotype is provided by Figure 18, which shows that organisms evolved via homeogenesis contain significantly more internalised ancestral environments than would be expected by an unbiased process of adaptation.

Chapter 4: Complexity lower bounds

4.1 Introduction

As discussed in chapter 1, it is widely accepted that the biosphere contains a general, long-term trend of increasing complexity of the most complex organisms (Bedau et al. 2000; Bedau 2009; McShea and Brandon 2010). However, most evolutionary experiments have generally failed to reproduce this trend, and instead commonly show robust trends of complexity minimisation (Bedau et al. 1997; Spiegelman et al. 1965; Bedau et al. 2000). In this chapter, we return to address the problem of these conflicting observations. Specifically, the key question we ask is how evolutionary theory can be refined to better explain these conflicting trends. In chapter 2 we addressed this issue by discussing the notion of complexity roadblocks that prohibit access to complexity without some particular mechanism that can work around them. However, here we focus instead on complexity drivers – understanding what, in the absence of any roadblocks, causes trends of increasing complexity in the first place.

There are two components to this question that we address here. The first component, which has received considerable attention in the literature (e.g. McShea 1991; McShea and Brandon 2010), is what mechanism or process has causes the biosphere’s general trend of increasing complexity of the most complex organisms. The second component, that has received less attention, is how such a mechanism can also explain (or how it can be compatible with) common observations of complexity minimisation in evolutionary experiments. In this chapter we attempt to deal with both components of the question: We seek a mechanism capable of generating robust, general trends of increasing complexity of the most complex organisms that is also compatible with the observations of consistent trends of complexity minimisation in evolutionary experiments.

Many potential mechanisms have been proposed to explain the biosphere’s general trend of increasing complexity of the most complex organisms (e.g. McShea 1991a; Carroll 2001; Bedau 2009). These include (a) driven mechanisms, such as the notion that constraints within organisms inherently

Chapter 4: Complexity Lower Bounds

increase over time, thus causing build up of historic information and increased complexity – i.e. complexity by increasing constraint theories – (Wimsatt 1986; Wimsatt 2001; Saunders and Ho 1976; Saunders and Ho 1981); (b) increased environmental complexity necessitating increased organismal complexity (George Ledyard Stebbins 1969; Adami, Ofria, and Collier 2000); and (c) undriven mechanisms, such as passive diffusion, where complexity of lineages changes as a random walk (McShea 1996). However, there is no consensus on whether these mechanisms are sufficient to explain this trend (McShea 1991; Bedau 2009). Furthermore, an outstanding problem is that most of the mechanisms proposed to explain the biosphere’s increase in complexity of the most complex organisms lack strict formal analyses, and remain as verbal arguments (McShea 1991; Bedau 2009). This makes it difficult to verify their proposed behaviours, and to understand whether any of those mechanisms are compatible with experimental observations of complexity minimisation.

In this chapter we describe a novel mechanism that we propose could be responsible for the combined observations of a general trend of increasing complexity of the most complex organisms in the biosphere and apparent complexity minimisation in evolutionary experiments. Specifically, we consider the implications of the model system in chapter 3 (i.e. metabolic evolution by homeogenesis) for complexity generation. Here we consider how over a series of environment changes, repeated events of homoeogenesis could result in the necessary addition of multiple environment adapters to the organism’s phenotype, resulting in a build up of functional complexity, and hence cause a trend of increasing organismal complexity as new environments are colonised.

4.2 Structure of this chapter

The hypothesis that we propose in this chapter requires a reasonable amount of logical unpacking before it can be described in detail. In particular, the mechanism we propose is a number of logical steps removed from homeogenesis. Therefore, to provide the necessary context, in the first section of this chapter (section 4.3) we describe the origin of this hypothesis; specifically, we provide a summary of results from exploratory testing using the model in chapter 3 and relevant associated theory that informed the generation of the hypothesis. Next, in section 4.4 we describe the proposed hypothesis. Following this, section 4.5 contains the key results of this chapter:

we carry out simulations to test if, and under what conditions, the hypothesised mechanism generates both types of trends (i.e. a general trend of increasing complexity of the most complex organisms coupled with robust trends of complexity minimisation). Then in section 4.6 we discuss the model results, limitations and caveats in relation to natural evolution. To assess the scope of the proposed mechanism, in section 4.7 we then carry out further experiments to test if the key factors responsible for generating complexity in our proposed hypothesis are present and affect evolution in a system of NAND gate circuit evolution, which is a common model of functional evolution in the literature. Finally in section 4.8 we summarise the work in this chapter and present our conclusions.

4.3 Creating a hypothesis for a complexity trend generation mechanism

The results from chapter 3 imply that in a system of homeogenesis, as environment change occurs, in some cases functional adapters must be added to the metabolism to enable survival in that given niche. It therefore follows that given a sequence of environment change such a process could result in a trend of increasing metabolic complexity. However, from the results in chapter 3 alone it is not clear what type or sequence of environment change is sufficient to generate such a trend. Specifically, before we can define a hypothesis for a complexity generation mechanism, we must first isolate which factors in the system are responsible for controlling the build-up and removal of metabolic complexity. To achieve this we have carried out some exploratory testing in which we observed the behaviour of the model of metabolic complexity from chapter 3 given a number of different types and sequences of environment change. We describe the key results from this exploratory work below.

4.3.1 Exploratory modelling

4.3.1.1 Methods

The primary tool that we will use to address the capability of homeogenesis for generating complexity trends is the metabolic evolution model described in chapter 3 (see section 3.5.2 for a detailed description of the model).

Chapter 4: Complexity Lower Bounds

Parameters used here are the same as those described in chapter 3 unless described otherwise. Briefly, the model describes the evolution of a population of organisms as they spread across a number of different, neighbouring chemical environments. All organisms have the same core replicator that requires a fixed set of chemical compounds to function. We assume that the replicator is too constrained to change. Therefore, as organisms encounter new environments, to enable survival they must evolve new metabolic machinery capable of converting the external chemical environment in to that required by their existing replication functionality. Given that here we are interested in the complexity behaviours of the model, it is also worth reiterating that the model contains an inherent pressure *against* complexity. Based on the observation from natural systems, the model includes a cost of resources – a fitness penalty proportional to the number of expressed genes (see section 3.5.2 for details). This is expressed as a constant pressure against complexity, similar to that observed in Spiegelman’s experiments and others (Spiegelman et al. 1965; Oehlenschläger and Eigen 1997; Bedau et al. 1997). Furthermore, as introduced in chapter 3, here we again distinguish between serial and parallel types of environment change (see section 3.5.2).

A key further aspect of the model that we must also define is how we measure complexity. A common problem with models of complexity evolution is knowing what to measure, because there is no agreed definition or measure of complexity (Mitchell 2009). However, qualitative examples are common. For example, C_4 plants are regarded to have a more complex mechanism of photosynthesis than C_3 plants (Ehleringer et al. 1991), because the C_4 mechanism contains all of the C_3 pathway *plus* some extra functional sophistication. But this increase in functional sophistication is not expressed as extra physical components; C_4 photosynthesis came about by a reorganisation of existing components already present in C_3 plants (Ehleringer et al. 1991). This highlights some of the problems of measuring complexity – in particular the difference between structural and functional complexity. Following this example from C_4 photosynthesis, in which complexity is defined by the size of the metabolic network, in the below models we will use metabolic pathway size (i.e. number of reactions in the network) as a proxy for functional complexity.

Results in the following section describe the behaviour of case study populations that represent the typical behaviour of evolving populations in their respective conditions.

4.3.1.2 Results and discussion

4.3.1.2.1 Exploratory test 1: Increase in complexity from parallel environment change

This system simulated the evolution of a species whose existing replication functionality is too difficult to change via evolution and that is then subject to a sequence of environments in which environment change occurs by an increasing number of parallel changes.

The environment had six niches laid out in a line. Each niche had an $n=200$ carrying capacity. The initial niche had exactly the right set of chemical compounds to allow the existing replication machinery to carry out reproduction. Each subsequent niche along the line varied by one more type of chemical compound. So for example, niche 2 differed from niche 1 by a single compound; niche 3 differed from niche 1 by two compounds, and so on. The result was a sequence of environments that required an increasing number of parallel functions.

Organisms evolved by a series of adaptive radiations, sequentially filling subsequent niches 1-6. The simplest viable organisms in each successive niche had progressively more complex metabolic networks. The evolved metabolic networks contained internalised environments from each of the previously inhabited niches (Figure 23). These results are straightforward. Competition in the form of density dependent selection makes neighbouring empty niches potentially attractive for successful mutants. This provides a pressure for evolution of new genes allowing expansion into empty niches that necessitate increased complexity, despite there being a constant selection for simplicity within any given niche.

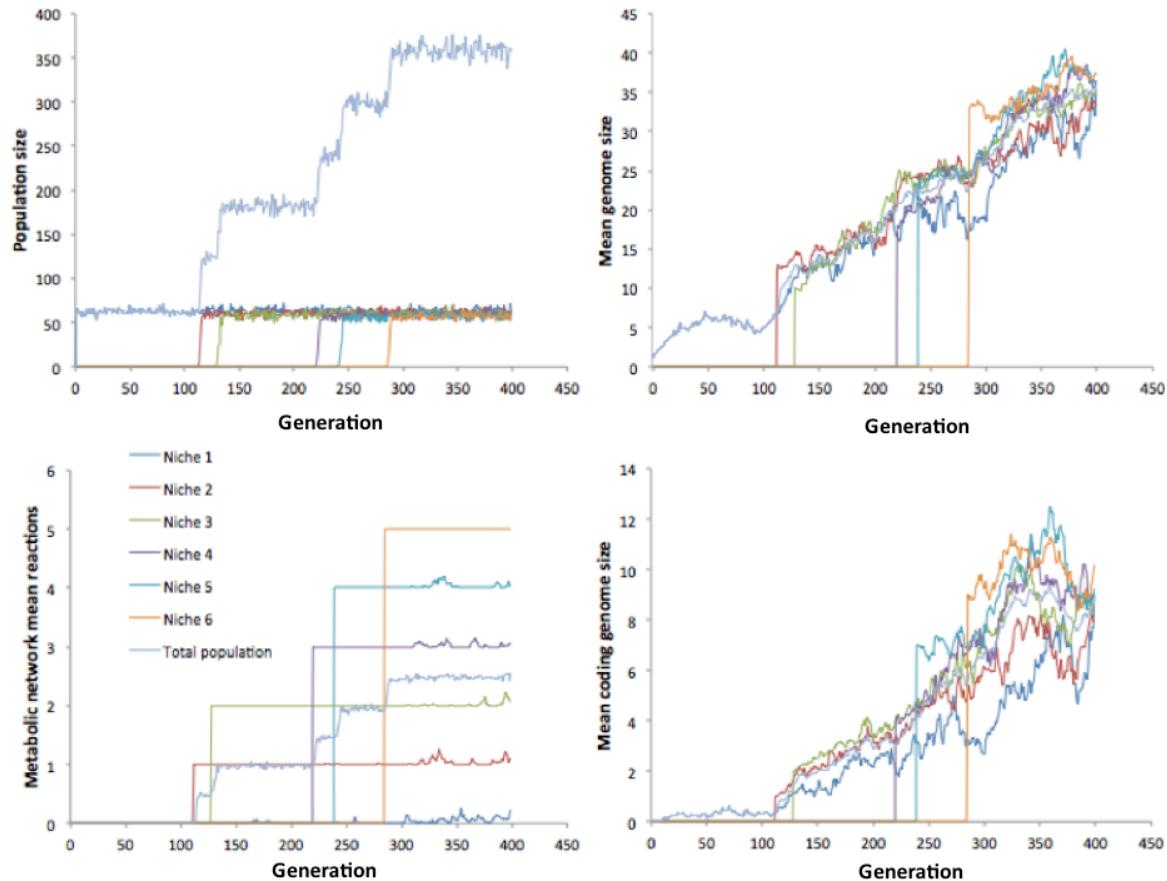


Figure 23. Exploratory test 1. All plots show results of organisms with non zero fitness. *Top left*: Total population size increases in a stepwise manner as each new niche is filled by an adaptive radiation. *Bottom left*: Mean metabolic network size increases with each progressively different niche; organisms in niche 6 have a more complex metabolic network, with more reaction steps, than those in niche 1. This therefore illustrates a simple environmentally mediated increase in metabolic complexity. *Top right, bottom right*: Mean genome size and number of genes expressed increases over time. This increase tails off after ~400 generations, limited by the fitness cost against expressed genes. Although the complexity of the chemical transformation necessary in each niche is different, this is difficult to discern from genome size or number of expressed genes, because many genes are non-coding or have no function in the phenotype.

This illustration demonstrates how certain sequences of environments can cause trends of functional complexity increase in the resulting organisms. In this case, it is environments that differ by progressively more distinct chemical compounds from the set of compounds required by the organism's replicator. In other words, a trend of increasing complexity is caused by an increasing number of parallel environment changes compared to the conditions required by the organism's replicator.

The system illustrates the intuitive notion that 'doing more things' requires more functional complexity than doing fewer things. As Heylingen describes:

'All other things being equal, a system that can survive situations A, B and C, is absolutely fitter than a system that can only survive A and B. Such an increase in absolute fitness is necessarily accompanied by an increase in functional complexity.' (Heylighen 1999)

Given this system, the increase in complexity is not surprising. As more chemical components of the environment become different from the conditions required by the organism's replicator, the number of chemical reactions required to convert one to the other increases, thus so does the minimum number of reactions in any possible viable organism's metabolic network. Selection for simplicity tends to keep the metabolic networks as simple as possible, but this cannot drive the metabolic network to be any smaller than the simplest possible network capable of transforming the environment back into the conditions required by the organism's replicator. The result is an environmentally mediated trend of increasing complexity. Importantly, these simple results contain both a local trend of complexity minimisation (that occurs within any given single niche) and simultaneously, on a system level (i.e. across multiple niches) a trend of increasing maximal complexity, as new niches are occupied.

4.3.1.2.2 Exploratory test 2: Increase in complexity from serial environment change

Some models for the evolution of complexity argue that organisms become more complex not only to cope with parallel environment change, but also to 'break through functional boundaries' (Arthur 1993) – i.e., to achieve more complicated individual tasks. Chemical reaction pathways provide many

Chapter 4: Complexity Lower Bounds

concrete examples of this principle: Some chemical transformations require more steps and intermediaries than others, and in that sense are more complex. For example, consider two different chemical reactions in the EMP glycolysis metabolic pathway. Glycolysis occurs, with some variations, in nearly all organisms, as a key part of cellular respiration (Horton et al. 1996). The overall reaction pathway converts glucose ($C_6H_{12}O_6$) into pyruvate ($CH_3COCOO^- + H^+$) by a 10-step process (i.e. with 10 intermediary molecules). In contrast, glucose ($C_6H_{12}O_6$) can be converted into glucose 6 phosphate ($C_6H_{13}O_9P$) in a single step (which is the first step in the glycolysis pathway; Figure 24).

Let us consider the 10-step glucose \Rightarrow pyruvate reaction pathway and the one step glucose \Rightarrow glucose-6-phosphate reaction as separate chemical functions. Why does changing glucose to pyruvate take 10 steps, whereas changing glucose to glucose-6-phosphate only take a single step? One possibility is that one task is inherently more complex than the other: Converting glucose into pyruvate is a more complicated task than converting glucose into glucose-6-phosphate, and therefore requires a more complex functional solution.

The network of possible chemical reactions that dictates how these conversions *can* possibly occur is determined by physics, which apparently does not allow a direct, single step change from glucose to pyruvate. This agrees with intuition; pyruvate is much *more different* from glucose than is glucose 6 phosphate, and so intuitively requires more changes to transform between one and the other. This implies that glucose \Rightarrow pyruvate is a fundamentally more complex task than converting glucose \Rightarrow glucose-6-phosphate. According to known chemical reaction networks, converting glucose to pyruvate necessitates more serial functional steps, and hence more functional machinery (to both carry out and organise these steps) than the task of converting glucose \Rightarrow glucose-6-phosphate.

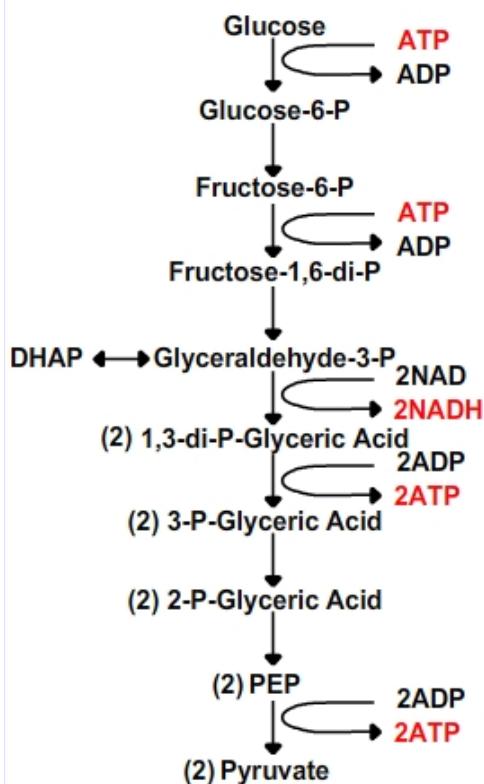


Figure 24. Schematic illustration of the differences in number of chemical steps required to convert glucose \Rightarrow pyruvate and glucose \Rightarrow glucose 6 phosphate. Both are components of the EMP glycolysis pathway. Presumably, because pyruvate is more organisationally different from glucose than is glucose 6 phosphate, the shortest path available in the underlying chemical reaction network to convert between glucose and pyruvate has many more steps than the path converting between glucose and glucose 6 phosphate. The result is that the minimum functional complexity of a metabolism capable of converting glucose to pyruvate must be larger than that converting glucose to glucose 6 phosphate.

In this system we simulate the evolution of a species whose existing replication functionality is under high constraint and that is then subject to environment change that (given this constraint) forces organisms to break functional barriers in this manner, evolving novel function to cope with progressively more different environments.

The environment had six niches laid out in a line. Each niche had an $n=200$ carrying capacity. The initial niche had the right set of chemical compounds so that the existing replication machinery could carry out reproduction. In terms

Chapter 4: Complexity Lower Bounds

of the *number* of compounds different from the first niche, each of the other five niches were the same, having only one chemical compound different.

However, in each subsequent niche along the line, the compound that was different was progressively *more* different (in terms of the chemical reaction network) than the first niche. In each subsequent niche it was one chemical reaction step *further away* (on the chemical reaction network) from its respective compound in the first niche.

Organisms evolved by a series of adaptive radiations, sequentially filling subsequent niches 1-6. The simplest viable organisms in each successive niche had progressively more complex metabolic networks. The evolved metabolic networks contained internalised environments from each of the previously inhabited environments (Figure 25).

The simplest viable organism in niche 6 evolved a 5-step metabolic reaction pathway that converted the different compounds in niche 6 back into their state in niche 1. This pathway sometimes contained a sequence of reaction steps that recreated the sequence of previous environments inhabited (Figure 25). But because the reaction network dictates that there are many possible viable reaction pathways to complete such a large transformation, occasionally, different pathways were evolved that did not contain all of the previous environments.

Again, these results show how certain sequences of environments cause trends of functional complexity increase in the resulting organisms. In this case, the environments were progressively more different from the conditions required by the organism's replicator by a single component (a chemical compound) becoming *more* different. Again, given the system, the general result of increased complexity is not surprising. As before, chemistry dictates that as an environment becomes more chemically different from the conditions required by the organism's replicator (this time in terms of a single component), then the minimum number of chemical reactions in a pathway capable of converting between the two increases, thereby increasing the minimum size of a viable metabolic network in that niche.

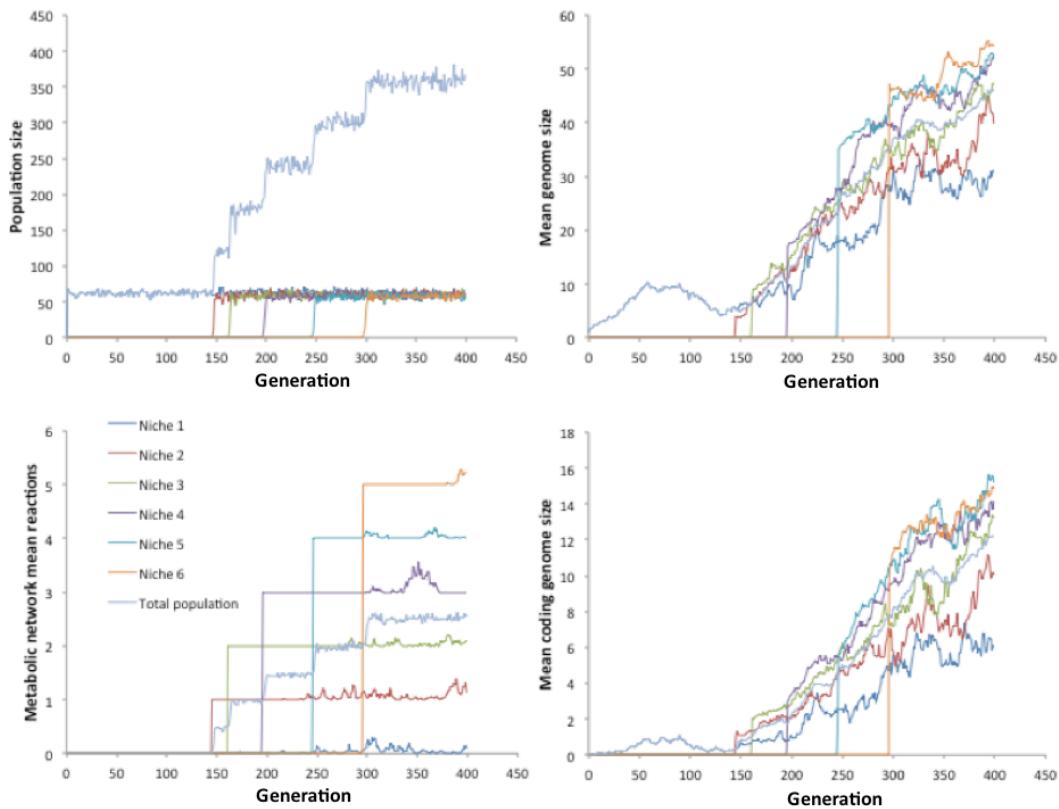


Figure 25. Exploratory test 2. All plots show results of organisms with non zero fitness. *Top left*: Total population size increases in a stepwise manner as each new niche is filled by an adaptive radiation, similar to system 1. *Bottom left*: Mean metabolic network size increases with each progressively different niche. This demonstrates an environmentally mediated increase in metabolic complexity, similar to system 1. *Top right, bottom right*: Mean genome size and number of genes expressed increases over time, eventually limited by the fitness cost against expressed genes. As with system 1, it is difficult to discern which organisms are more complex by genome size or number of expressed genes, as many genes are non-coding or have no function in the phenotype.

4.3.1.2.3 Exploratory test 3: Decrease in complexity by parallel environment change

Given the right sequence of environments, the model system described can also result in decreases in complexity. Again, this is straightforward. We have provided transparent illustrations to concretise the verbal argument, which can

Chapter 4: Complexity Lower Bounds

be summarised as follows: Some sequences of environments provide organisms with chemistry that is closer to the requirements of their replicator than their current environment. This provides the possibility of a shorter metabolic pathway capable of converting these new environments to the conditions required by the organism's replicator, potentially making some of the existing functional steps, and their associated internalised environments, redundant. If this shorter pathway can be reasonably evolved, in some cases it likely will be, especially in the presence of selection for simplicity. The result will therefore be a decrease in complexity.

For simplicity, and to enable separation of the two ways in which this system can bring about loss of complexity, both models are extreme and unrealistic examples that extend the previous two simulations, as before. Both simulations have an extra niche added (niche 7) that neighbours niche 6.

The first simulation extended exploratory test 1 (parallel environment change). Here niche 7 contained an environment that differed from niche 1 by two parallel chemical reactions. As a result, the chemistry of niche 7 was much closer to the set of compounds required by the organisms' replicator than its neighbouring niche (niche 6, which was different by five reactions).

Organisms evolved by a series of adaptive radiations, sequentially filling subsequent niches 1-7. The simplest viable organisms in each successive niche 1-6 had progressively more complex metabolic networks (in terms of number of reaction steps) formed by homeogenesis. The simplest organisms in niche 7 had two reaction steps in its metabolic network, less than half the number of reaction steps than those in niche 6 (which had five), representing a decrease in complexity (Figure 26).

In an environment where a simpler set of metabolic reactions is potentially capable of reproducing the conditions required by the organism's replicator than the metabolic network currently evolved, then because of selection for simplicity this simpler metabolic network will, if possible, be preferentially evolved. Vestigial genes that represent now redundant reactions will likely deteriorate by genetic drift. The result is a potential decrease in complexity.

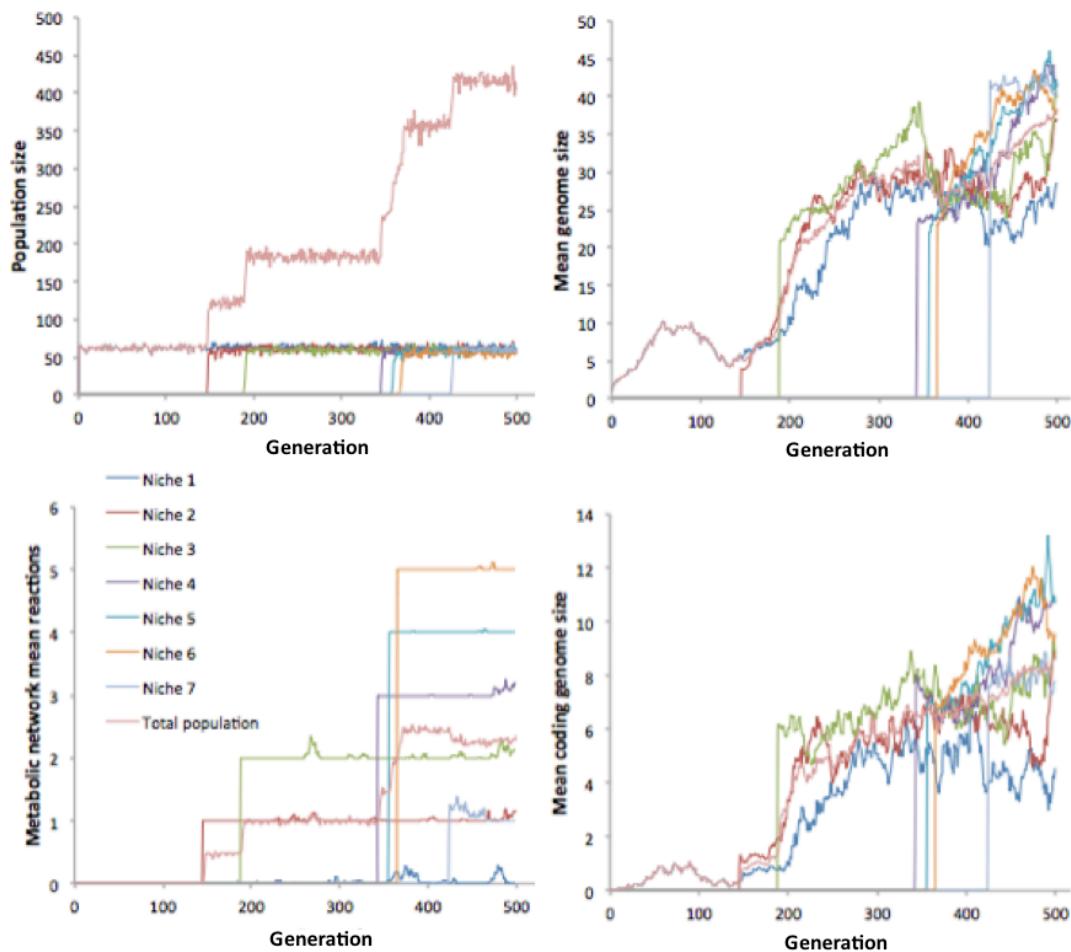


Figure 26. Exploratory test 3. All plots show results of organisms with non zero fitness. *Top left:* Total population size increases in a stepwise manner as each new niche is filled by an adaptive radiation, similar to systems 1 and 2. *Bottom left:* Mean metabolic network size increases with each progressively different niche, but dramatically decreases when niche 7 is colonised. This demonstrates an environmentally mediated increase in metabolic complexity, followed by an environmentally mediated *decrease* in metabolic complexity. This shows one way that environment change can bring about both increases and decreases in phenotypic complexity, given sufficient functional constraint. *Top right, bottom right:* Mean genome size and number of genes expressed follows a similar pattern to previous systems; phenotypic complexity trends are hard to observe in this data.

4.3.1.2.4 Exploratory test 4: Decrease in complexity by serial environment change

This system extended exploratory test 2 (serial environment change). Here niche 7 contained a new molecule type, not present in niches 1-6. This molecule was a neighbouring molecule of the altered molecule in niche 3. It differed from niche 1 by a minimum sequence of 2 chemical reactions.

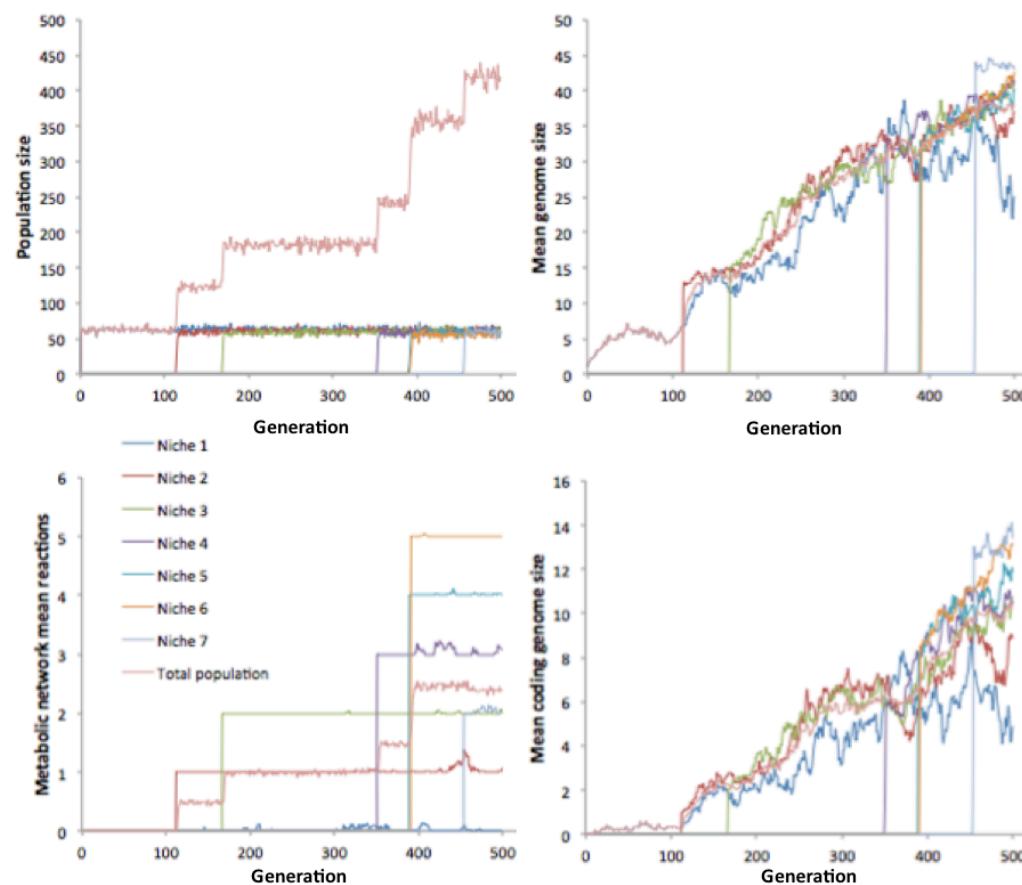


Figure 27. Exploratory test 4. All plots show results of organisms with non zero fitness. *Top left*: Total population size increases in a stepwise manner as each new niche is filled by an adaptive radiation, similar to systems 1, 2 and 3. *Bottom left*: Mean metabolic network size increases with each progressively different niche, but dramatically decreases when niche 7 is colonised, as with system 3. This again demonstrates an environmentally mediated increase in metabolic complexity, followed by an environmentally mediated decrease in metabolic complexity. *Top right, bottom right*: Trends in mean genome size and number of genes.

The results were qualitatively similar to system 3. The simplest viable organisms in niche 7 had metabolic networks with fewer reaction steps (i.e. 2) than those in niche 6 (i.e. 5), thus demonstrating an environmentally mediated trend in decreasing complexity.

These two systems illustrate the two separate ways in which complexity can be lost in this system, and that this can be predominantly controlled by the type and sequence of environments inhabited. They are simply the reverse of the ways in which complexity can be added: by doing fewer things (due to a reduction in parallel environment change), or by doing a less complex thing (being able to substitute a shorter/fewer-step function for a longer/more-step function due to a reduction in serial environment change).

Taken as a whole, the four systems described provide a picture of the two extreme types of environment change (serial and parallel), showing in each case how they can cause increases and decreases in complexity by homeogenesis, given a replicator that is too difficult to change. They demonstrate how sequences of environments can be internalised to cope with environment change, and how this adds functional complexity in terms of parallel change, resulting in parallel functions, or by serial change, resulting in serial functions – even in the face of a constant pressure for simplicity. They also show how this function can be lost, should the environment change to a state where the added function becomes redundant.

4.3.2 Summary of exploratory modelling

The results of exploratory testing have documented how homeogenesis can cause complexity trends in evolution. In this section we will attempt to isolate what key factors cause build-up of complexity by homeogenesis.

First, it is clear that homeogenesis generates robust, repeatable trends in complexity; it is not the result of an undriven random walk: This mechanism is not passive diffusion. Second, the above simulations demonstrate that homeogenesis can cause such trends of increasing complexity without the need for a corresponding increase in environmental complexity: In simulations showing trends of increasing complexity (i.e. exploratory tests 1 and 2), each environment contained the same number of chemical compounds, but evolution across these environments still resulted in systematic increases in

organismal complexity. This stands in contrast to many proposed environmental mechanisms of complexity increase in evolution. Third, the simulations also show that homeogenesis produces robust trends of increasing complexity without the need for a corresponding increase in constraint within the organism. Although some fixed amount of constraint was present (specifically that organisms have a replicator that is assumed to be too constrained to change), no further constraint was artificially introduced from one niche to the next, and yet evolution across these niches nonetheless resulted in robust trends of increasing complexity.

If complexity does not result from passive diffusion, increasing environmental complexity or increasing constraint, what causes it to increase in these systems? In short, the above results imply that the build up of metabolic complexity in this system is controlled not by absolute environment change, but by the number of chemical changes (i.e. shortest path of reaction steps in the chemical reaction network) that separate the given external niche from the environment required by the organism's replicator. As we will explain in the next section, this observation helps us connect the complexity generating behaviour of this system to 'complexity lower bounds', which are a well-known property of algorithm problems in computer science – thus better enabling us to describe a mechanism of complexity generation in this system.

4.4 A hypothesis for a novel complexity generation mechanism in evolution: Environmental dissociation complexity

As discussed in chapter 1, In computer science, it is an established result that for any algorithm that converts a set of inputs to a set of outputs, a specific lower bound will exist on the complexity of any possible solution, and the magnitude of this complexity lower bound will depend on which inputs and outputs are being converted between (Papadimitriou 2003). For example, exhaustive searches have shown that of the many possible logic circuits, the simplest possible NAND logic circuit that can add 2 binary bits contains 5 logic gates – and adding 3 binary bits takes a minimum of 9 NAND gates. Similarly, it has been proven mathematically that sorting a list of length n by successive compare-and-swap operations will require an algorithm that defines a

minimum of $n \log n$ operations (Papadimitriou 2003). (However, no general proof exists for an arbitrary problem, and for many problems / transformations – e.g. sorting, the travelling salesman problem, matrix multiplication, etc. – the simplest known solution is an empirical estimate, based on the best solution available at the time; Papadimitriou 2003).

Similar concepts have also been discussed in biology. For example, recent research has suggested that there is a minimum gene set capable of sustaining cellular life of around 250-300 genes (Koonin 2011); elsewhere, passive diffusion models often include a ‘left wall’, which describes some minimum complexity below which evolution cannot go (McShea 1996). However, despite these examples, complexity lower bounds are not widely discussed in the evolutionary literature.

To explain how this concept relates to the complexity generation behaviour of our model system, it is helpful to consider a simple analogous system that also contains complexity lower bounds. A Rubik’s cube is popular 3-D combination puzzle that was invented in 1974 by Erno Rubik. The puzzle consists of a cube, each side having 9 coloured faces. In its solved state, the cube has all 9 faces on a given side of the cube the same colour. The cube allows users to rotate its sections, allowing each of the faces to be moved individually. Doing this, the cube is reorganised into a random state; the aim of the puzzle is to return the cube back to its original, ‘solved’ state (Figure 28).



Figure 28. Solving a Rubik’s cube. The cube begins in a random, disorganised state (left). The puzzle can rotate in various ways, allowing the faces to be rearranged. The aim of the puzzle is to return the cube back to its solved configuration (right), in which all of the faces on each side of the cube are the same colour.

Chapter 4: Complexity Lower Bounds

In mathematical terms, the goal of the Rubik's cube puzzle is to find a sequence of moves that will transform the cube from its current state into its solved state. Obviously from any given state, there are many possible sequences of moves that can reach the solved state. However, finding these sequences is difficult: despite its simplicity, the Rubik's cube has over 4.3×10^{19} possible configuration states. A key area of mathematical research on the Rubik's cube has been focused on attempting to find the shortest possible sequence of moves (i.e. lower bound) needed to solve the cube from a given state. To address this problem, we can imagine the set of all possible moves from any given state as an expanding network, where the nodes are configuration states of the cube, and the links are individual moves (Figure 29).

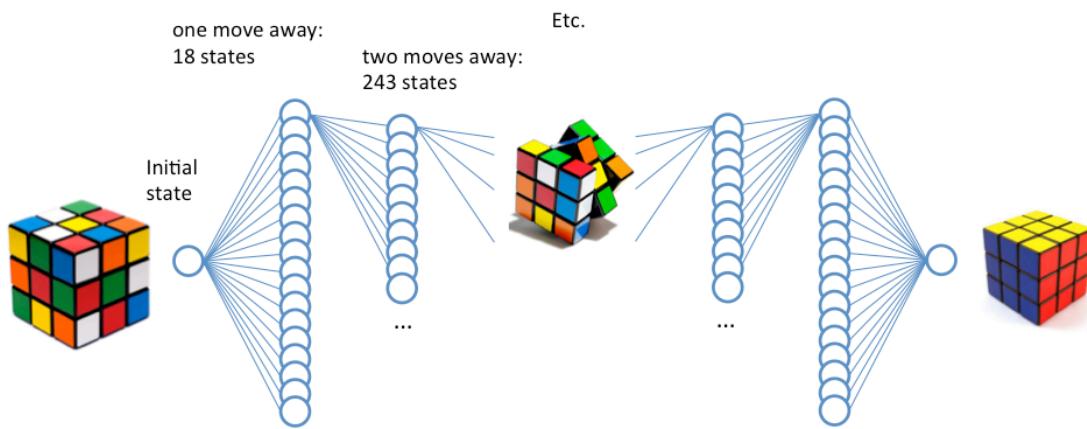


Figure 29. The state space of a Rubik's cube can be described as a network of cube states (nodes) and moves (links).

By mapping out this network, we can know exactly, from any position, the minimum number of moves we are away from the solved state. This network is termed the 'state space' or 'state space search tree' of a system (Russell et al. 1995; Lilius 1998; McMillan and Probst 1995): a network that contains the complete set of possible configurations of the system connected by links that represent available change operators that connect those states. For a Rubik's cube, the nodes of the state space are configurations of the cube, and each link is an individual rotation of one face. Only in the last few years has the state space for a Rubik's cube been fully mapped (Figure 30).

Distance	Count of Positions
0	1
1	18
2	243
3	3,240
4	43,239
5	574,908
6	7,618,438
7	100,803,036
8	1,332,343,288
9	17,596,479,795
10	232,248,063,316
11	3,063,288,809,012
12	40,374,425,656,248
13	531,653,418,284,628
14	6,989,320,578,825,350
15	91,365,146,187,124,300
16	~1,100,000,000,000,000,000
17	~12,000,000,000,000,000,000
18	~29,000,000,000,000,000,000
19	~1,500,000,000,000,000,000
20	~300,000,000
21	None

Figure 30. The number of positions that are a given number of moves from the solved state of a Rubik's cube. For example, there are 18 states that are 1 move away from the solved state, 243 states 2 moves away, and so on. This imparts fundamental lower bounds on the complexity (i.e. number of moves) of possible solving algorithms from a given state.

In state space, a sequence of moves that can solve the cube from a given position corresponds to a path through state space that connects the given current configuration to the solved configuration. The important points to note from Figure 30 are that every possible state has a fixed lower bound on the number of moves in which the solved state can be reached (i.e. shortest path to the solved state), and this is different for different states depending on where they are in the network. In other words, there are inherent lower bounds on the complexity of possible solving sequences from any given state, and those lower bounds vary depending on the state we start from.

In this chapter, we are arguing that complexity in evolution (and specifically, our model system of metabolic evolution) is controlled in a similar manner. One can imagine the set of available chemical environments as being like an

Chapter 4: Complexity Lower Bounds

extremely complicated Rubik's cube. The variation of the environment in time and space represents different states of the Rubik's cube. The solved state of the cube represents the very special environment in which DNA, the base replicator of life, can replicate, which is necessary for life and evolution to continue. In this system, the task of evolution is the same as the task of the puzzle solver with a Rubik's cube: given any environment, it must create a sequence of moves that can transform the current state of the system into the solved state, and to do so can only use the available transformation operators defined over the state space (but instead of sequences of moves, we call these sequences metabolic networks or organisms).

Most importantly, this system provides us with a new perspective on what factors control complexity in evolution. The key point is that in such a system, just as in a Rubik's cube, there will be inherent lower bounds on the complexity of possible solving algorithms (organisms) from any particular state (niche), and these are defined by the shortest path in state space from the current state (current external niche) back to the solved state (the environment required by the organism's replicator). Moreover, different states (niches) will have different lower bounds.

Importantly, this is different from saying that complexity comes about in evolution because one niche is fundamentally more complex than another. In the Rubik's cube, and in our chemical model, all states of the system are identically complex. Rather than absolute complexity, the difference in complexity lower bounds in different niches comes from their different distances in the state space network from the 'solved' state, not from any inherent properties of the niches themselves.

4.4.1 The Environmental Dissociation Hypothesis

We can now describe the main hypothesis of this chapter. Let us term the distance in state space of the current environment from the environment required by the organism's replicator the amount of 'environmental dissociation'. Because this factor effectively controls complexity in the system, we term the mechanism of complexity generation by this process *environmental dissociation complexity*. This mechanism is defined as follows:

Given:

- a. an organism with a replicator that can replicate in some small subset of environmental conditions, and whose replicator cannot feasibly be changed to replicate in conditions outside of this subset;
- b. an environment with heterogeneous environmental dissociation whose conditions change sufficiently gradually;
- c. an inherent selection pressure against complexity such as a cost of resources

then as competition forces evolution to leave the original environment (a), and colonise new environments (b), the magnitude of environmental dissociation of a new environment will dictate the minimum possible complexity of viable organisms in that environment, resulting in a system-wide trend of increasing complexity of the most complex organisms, coupled with local trends of complexity minimisation in individual environments, caused by (c).

4.5 Testing the environmental dissociation complexity hypothesis by computational modelling

4.5.1 Methods

To test this hypothesis, we carried out simulations in an expanded version of the metabolic model described earlier in this chapter and in chapter 3. The model differed from that described in chapter 3 in the following manner. Instead of 7 niches, there were 50 niches, again laid out in a line. Furthermore, rather than having contrived environment configurations in each niche, unless described otherwise we began with a chemical environment generated in the following manner (as before, containing 18 of the set of 36 possible chemical compounds). Each successive neighbouring niche was created by performing one, random chemical reaction from its neighbour (each reaction could be serial or parallel change, there were no restrictions). The result was that the simulation contained 50 neighbouring niches distributed in a line, each varying by one chemical reaction from the previous niche. The chemical makeup of niche 1 was then taken to be the fixed target input required by the initial organism's replication machinery, hence allowing the initial organism to reproduce in niche 1 without any associated metabolism.

Chapter 4: Complexity Lower Bounds

In all cases, unless stated otherwise, the system was set up with all of the conditions required by the hypothesized mechanism, i.e.:

1. Organisms with a replicator assumed to be too difficult to change by natural selection;
2. A heterogeneous environment with gradually varying environmental dissociation;
3. A pressure against complexity in the form of a cost of resources;

We carried out 4 control experiments to test the hypothesis.

1. Positive control;
2. Negative control 1: No change in environmental dissociation;
3. Negative control 2: Punctuated environment change;
4. Negative control 3: No pressure against complexity.

4.5.2 Results

4.5.2.1 Positive control experiment

First we carried out a positive control experiment in which we tested whether the mechanism robustly created the complexity trends claimed (i.e. a system-wide general trend of increasing complexity of the most complex organisms, coupled with local trends of complexity minimisation within individual niches) given all of the conditions stated in the mechanism. Case study results that show typical behaviour of the system are illustrated in Figure 31. In all positive control experiments, the system generated a system-wide trend of increasing complexity of the most complex organisms that corresponded to the colonisation of new niches, coupled with local trends of complexity minimisation in individual niches (e.g. Figure 31), supporting the hypothesis. We carried out 100 repetitions of the positive control, each running for 5000 generations. Results showed a marked increase in mean complexity of all organisms in the population by generation 5000 (mean=10.94, standard error=0.21), and a marked increase in complexity of organisms in the most complex occupied niche (mean=19.89, standard error=0.39); these results are plotted in comparison to a negative control in Figure 35 (left columns).

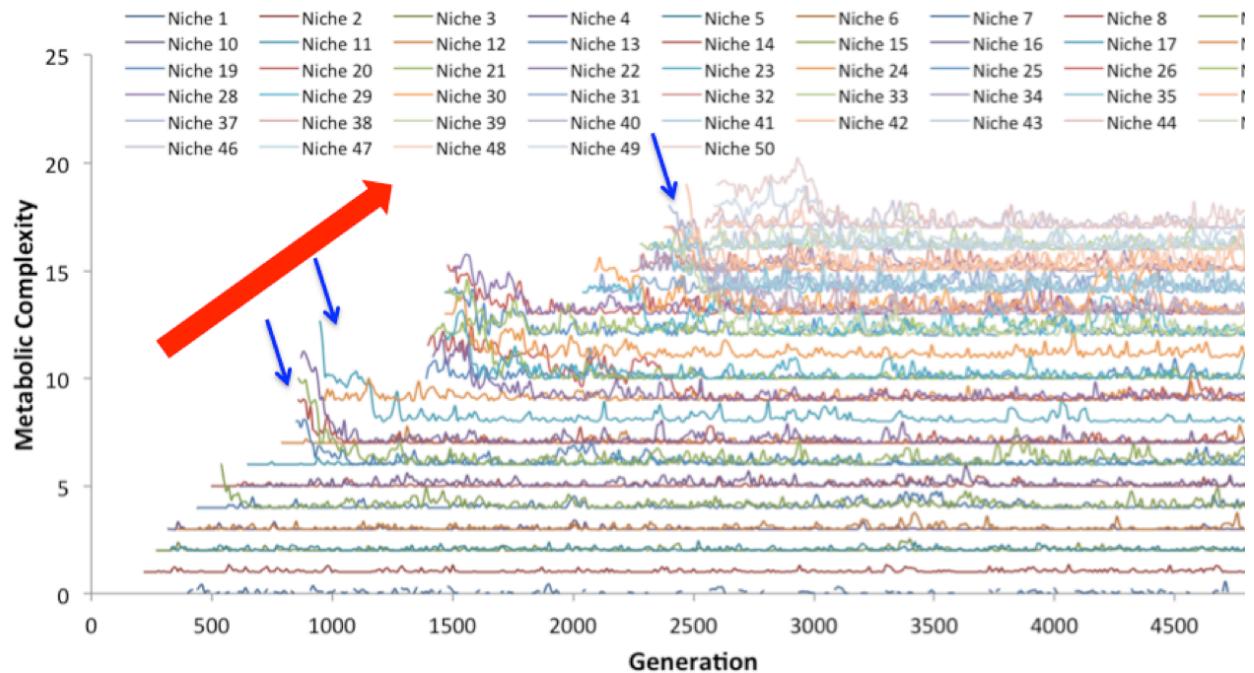


Figure 31. Case study example of the positive control experiment illustrating the typical behaviour of evolution in this system. Each data line shows the mean metabolic complexity of viable organisms in a given niche. Niches are only plotted once they are colonised. At generation 0, a single organism begins in niche 1 with no metabolism. As neighbouring niches are colonised, the mean complexity of organisms in that niche is plotted on the chart. (E.g. at generation 223 niche 2 is populated: red line) New niches very often have greater environmental dissociation, and hence require a more complex metabolism. The result is that as new niches are colonised, the system generates a trend of increasing complexity of the most complex organisms in the system (red arrow). However, within any individual niche, complexity displays a trend of minimisation to the complexity lower bound in that niche. Blue arrows show example characteristic events of local complexity minimisation in individual niches. This result helps to explain conflicting observations of increase of maximal organismal complexity in the biosphere and trends of complexity minimisation in experiments.

Chapter 4: Complexity Lower Bounds

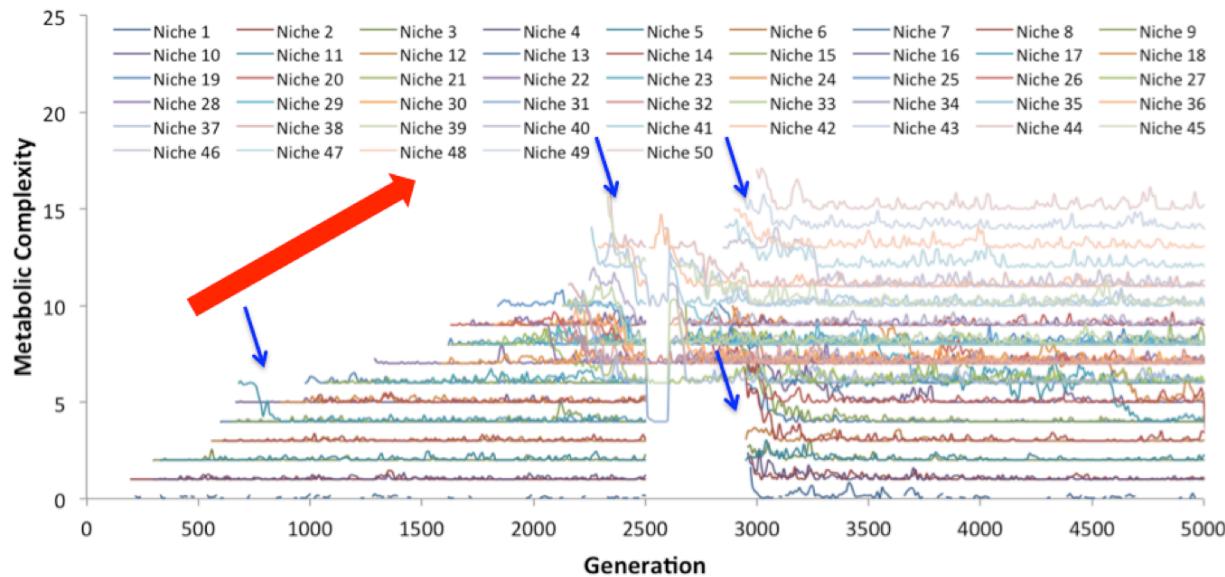


Figure 32. Case study example of the positive control experiment (with perturbation) illustrating the typical behaviour of evolution in this system. This system is identical to the positive control experiment but also underwent environmental perturbation at generation 2500 that lasted for 100 generations. In this instance, only a few species survived the extinction, the rest going extinct. The species that survived were moderately complex, leaving niches 1-6 empty. These niches were subsequently recolonised during an adaptive radiation around generation 2950 by these more complex organisms. The recolonising organisms quickly lost much of their complexity after colonising these niches (~generation 2950-3400), as they were pushed by the inherent pressure against complexity in the system towards the complexity lower bound in their given niche. The results again show a general, system wide trend of increasing complexity of the most complex organisms corresponding to the colonisation of new niches with higher environmental dissociation (red arrow), coupled with local trends of complexity minimisation in individual niches. Blue arrows show characteristic events of local complexity minimisation in individual niches. These are particularly common after the perturbation, when more complex organisms come to recolonise niches with smaller environmental dissociation, thus allowing genes redundant in that niche to be jettisoned and complexity decreased.

4.5.2.2 Negative control experiment 1: No change in environmental dissociation.

We next asked whether increasing environmental dissociation is necessary to produce a general trend of increasing complexity, or if environment change alone is sufficient. To test this we set up two separate simulation cases: a positive control experiment, in which environmental dissociation varied across niches, and negative control experiment that was identical except that environmental dissociation did not vary across niches. In both cases, niches were generated in order 1 to 50 by carrying out two random chemical reactions different from the previous neighbouring niche, but in the negative control experiment environmental dissociation was kept constant. We then measured the resultant increase in metabolic complexity after 5000 generations (Figure 33).

To allow environment dissociation to be kept constant across different niches, each niche was separated by two reactions as opposed to one; a single reaction will always increase or decrease the environmental dissociation in this system, because it will always either step nearer to or further away from the niche required by the replicator. Hence enforcing there to be two reactions between niches allowed the possibility of zero net gain of environmental dissociation.

To ensure that environment change occurred even when change in environmental dissociation did not, in both experiments no two niches were permitted to have the same set of compounds as any other niche in that experiment. An additional necessity to allow environmental dissociation to be fixed across different niches was that unlike in other experiments, here niche 1 was not the same as the environment required by the organisms' replicator. Instead, it was 10 reactions dissociated from this environment: The result was that to survive in niche 1, organisms required a minimum 10-step metabolic pathway that converted the initial niche into the niche required by the organisms' replicator. This requirement was necessary because it is not possible to change the environment from that required by the organism's replicator to any other environment without increasing the environmental dissociation. (For example, it is possible to change a Rubik's cube that is 10 steps away from its solved state and leave it in a new configuration that is still 10 steps from its solved state, thus keeping its environmental dissociation

Chapter 4: Complexity Lower Bounds

fixed; however, it is not possible to do this if the Rubik's cube begins in its solved state, as any change to a new configuration will increase the distance from this state.) In each simulation, the initial organism was provided with such a 10-step pathway, enabling it to survive and reproduce in niche 1. Accordingly, the results measured the change in complexity evolved from this starting point of 10.

We carried out 100 repetitions of each case, each time using different randomly generated sequences of niches according to the above method. Results are shown in Figure 33. In the positive control experiment, the mean increase in complexity across all organisms in the system was very significantly higher than observed in the negative control experiment (positive control: mean=7.38, standard error=0.40; negative control: mean=0.22, standard error=0.02; $P<0.001$ Student's t-test). The mean increase in complexity of organisms in the most complex occupied niche was also very significantly higher in the positive control experiment than in the negative control experiment (positive control: mean=12.49, standard error=0.46; negative control mean=0.86, standard error=0.08; $P<0.001$ Student's t-test). The results support the hypothesis that the increase in complexity observed in the positive control experiment was caused by the increase in environmental dissociation across niches (positive control experiment: mean change in environmental dissociation from niche 1 to niche 50= +11.88, standard error 0.28; negative control experiment: change in environmental dissociation from niche 1 to niche 50=0). In the negative control experiment, there was no increase in environmental dissociation across niches, and complexity remained very low. Case study examples illustrating typical behaviour in each experiment show that there was a clear, system-wide general trend towards increasing complexity of the most complex organisms in the positive control experiment (Figure 34, bottom). Such trends were observed in every repetition of the positive control experiment. In contrast, in all of the negative control experiments (e.g. Figure 34, top), no such general trend of increasing complexity was ever observed.

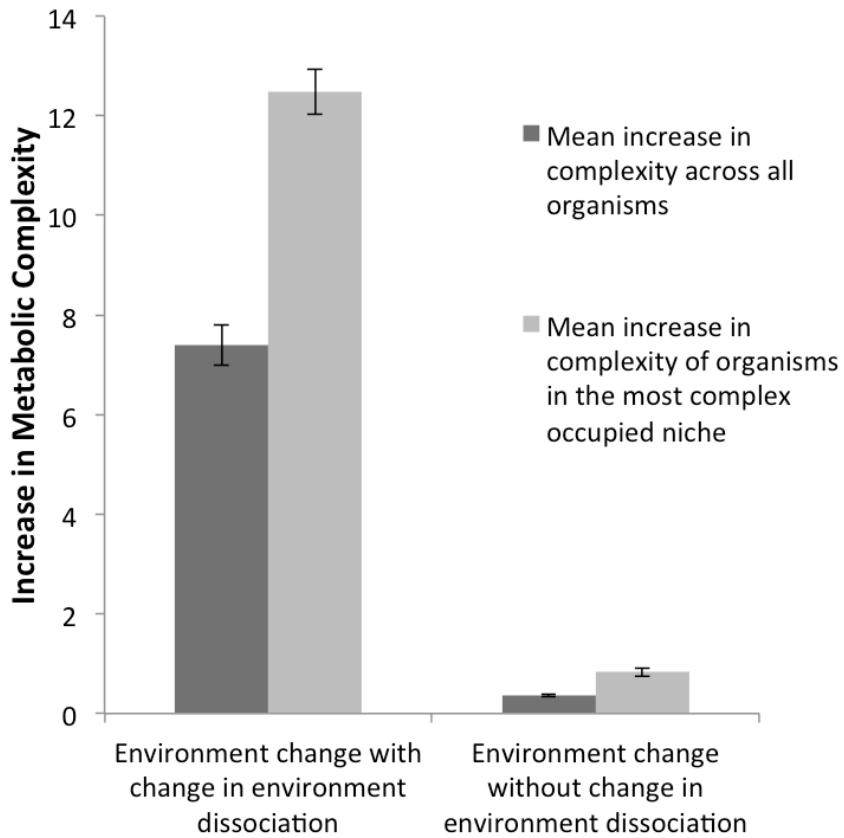


Figure 33. Plot illustrating that environment change alone is not sufficient to generate significant increases in organismal complexity in this model, but that instead environmental dissociation is a key factor controlling complexity. The left columns show statistical results from 100 repetitions of experiments in which environmental dissociation was free to increase (the mean increase was +11.88 over all niches in these repetitions), resulting in a significant increase in mean complexity and the mean complexity of organisms in the most complex occupied niche ($p < 0.001$). In contrast, the right columns show results from 100 identical repetitions where there was no increase in environmental dissociation across any niches, resulting in very little change in complexity.

Chapter 4: Complexity Lower Bounds

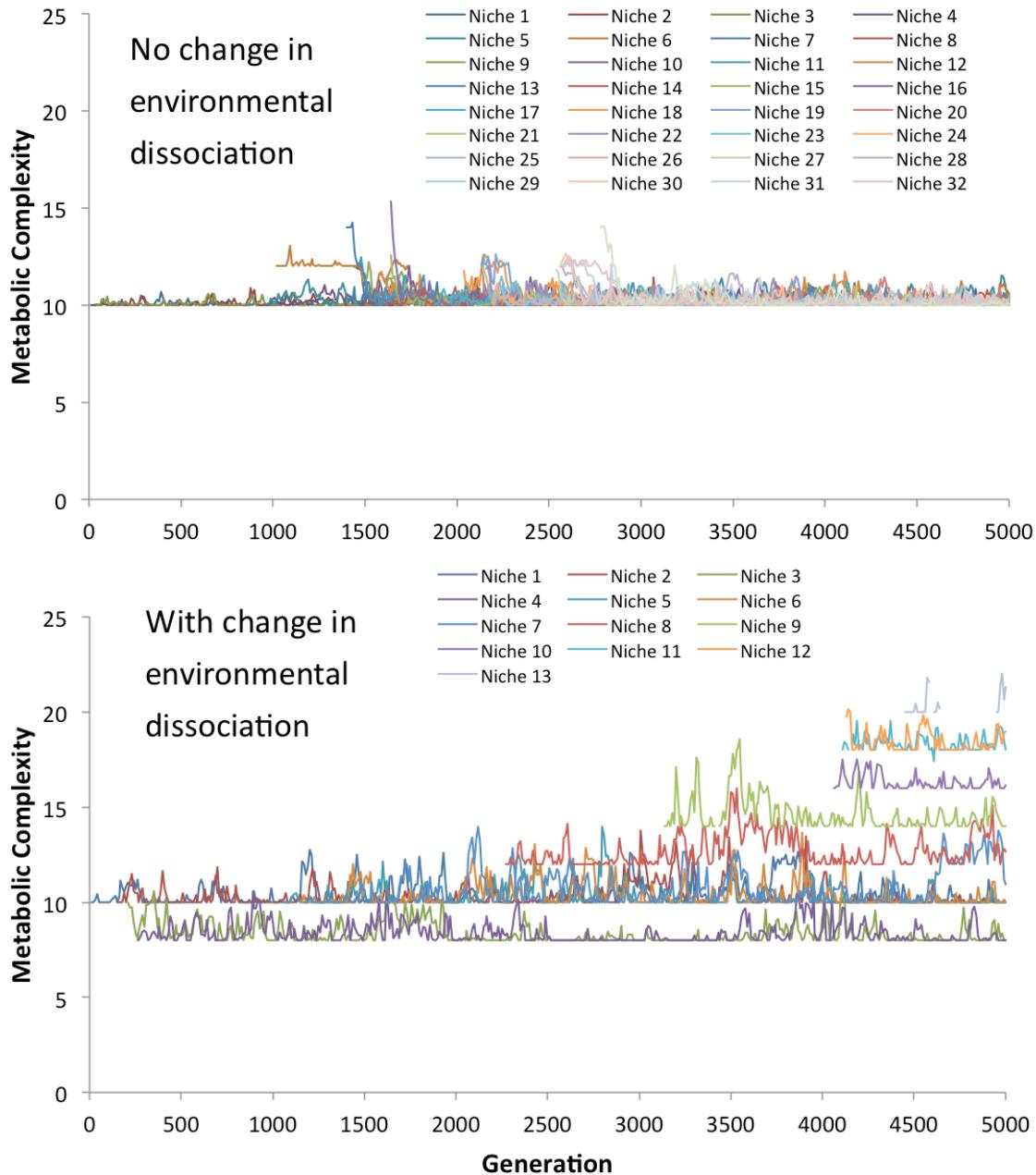


Figure 34. Case study example of the negative control experiment (top) and positive control experiment (bottom) for environmental dissociation. Top: even though environment change occurred between niches (and hence viable organisms were different in each niche), no increase in complexity was necessitated as new niches were colonised because all niches had the same complexity lower bound. Bottom: In contrast, when environmental dissociation was free to vary between niches, trends of increasing and decreasing complexity are observed that correspond to changes in complexity lower bound from one niche to the next.

4.5.2.3 Negative control experiment 2: Punctuated environment change

We next asked how the gradual nature of environment change affected complexity trends in the system. To achieve this we compared a positive control experiment in which the environment changed gradually between niches with a negative control experiment in which environment change was less gradual. Specifically, in the positive control experiment, the difference between neighbouring niches, x , was a single chemical reaction, representing one step on the chemical reaction network ($x=1$). In contrast, for the negative control experiment $x=3$, representing a more punctuated environment. Mean results for 100 repetitions of this experiment are shown in Figure 35, and a case study showing the typical behaviour of a single repetition is shown in Figure 36. The punctuated environment severely limited the capability of evolution for colonising new niches in the negative control experiment, which in turn limited the extent to which complexity could evolve. In the positive control experiment, the mean increase in complexity across all organisms in the system was very significantly higher than observed in the negative control experiment (positive control: mean=10.94, standard error=0.21; negative control: mean=0.31, standard error=0.11; $P<0.001$ Student's t-test). The mean increase in complexity of organisms in the most complex occupied niche was also very significantly higher in the positive control experiment than in the negative control experiment (positive control: mean=19.89, standard error=0.39; negative control mean=0.59, standard error=0.20; $P<0.001$ Student's t-test).

The behaviour of this control experiment can be described by a simple mathematical argument that implies the time taken to colonise a new niche will increase exponentially with x . In a niche with one compound different from the set of compounds required by the organisms' replicator, the probability of producing a viable mutant for that niche = $(m/nR)L$, where m =per locus mutation rate, nR =number of reactions possible in the artificial chemistry (assuming no evolvability / directed variation, and that all possible mutants catalyse some reaction), and L = number of loci. In the case of an environment with two compounds different, the probability is approximately $(m/nR)^2L$, as two separate but correct mutations are required at once. The details are slightly more complex, as both mutations could occur at different generations. However, given the cost of keeping redundant genes, it is unlikely that a single

Chapter 4: Complexity Lower Bounds

one of these mutations (which alone are redundant) would be kept for more than a few generations. The general probability is therefore $p \sim (m/nR)^{nC}L$, where nC is the number of compounds different. Thus the probability of evolving a viable mutant decreases exponentially with the rate of environment change (i.e. the number of compounds different in the new environment). The time to evolve such a mutant increases with the inverse of p - i.e., exponentially.

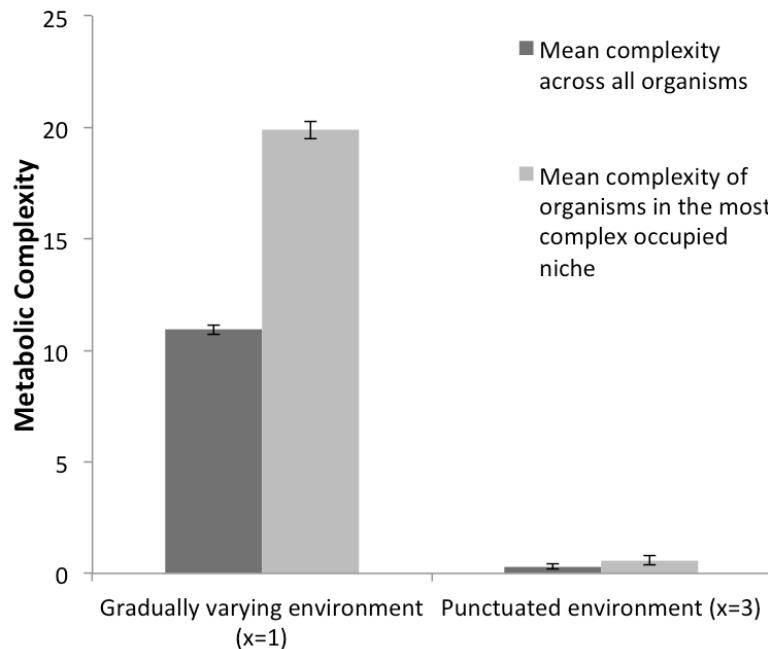


Figure 35. Plot illustrating the effect of gradual environment change on generating complexity in this model. In the positive control experiment (left columns) the amount of environment change, x , was a single chemical reaction between neighbouring niches ($x=1$). In the negative control experiment (right columns) environment change was less gradual ($x=3$). Given an environment that changed in a less gradual manner, evolution was significantly less capable of generating complexity ($p<0.001$).

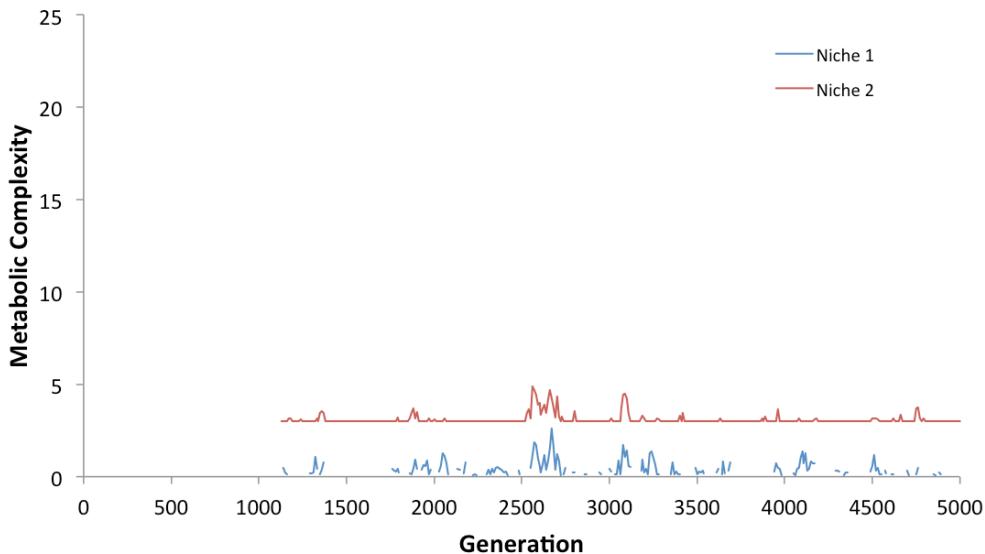


Figure 36. Case study example of the negative control experiment for gradual environment change showing typical behaviour of evolution in this system. Each data line shows the mean metabolic complexity of viable organisms in a given niche. Niches are only plotted once they are colonised. In this experiment, the environment changed in a more punctuated manner, having 3 chemical reactions between each neighbouring niche ($x=3$), compared to only one in the positive control experiment ($x=1$). The results illustrate that given a more punctuated environment, evolution struggles to colonise new niches, which in turn limits the evolution of complexity. Each data series shows the mean metabolic complexity for organisms in a given niche. In this example, only one new niche is colonised (niche 2, colonised at ~ 1200 generations). Here, niche 2 is 3 chemical reactions from niche 1, resulting in an environmental dissociation of 3. Accordingly, the complexity lower bound for organisms in niche 2 is 3, and the complexity of viable organisms in niche 2 never passes below this value.

4.5.2.4 Negative control experiment 3: No cost of resources

Finally, we sought to test the effect of having a cost of resources in the model to act as an inherent selection pressure against complexity in the model.

Chapter 4: Complexity Lower Bounds

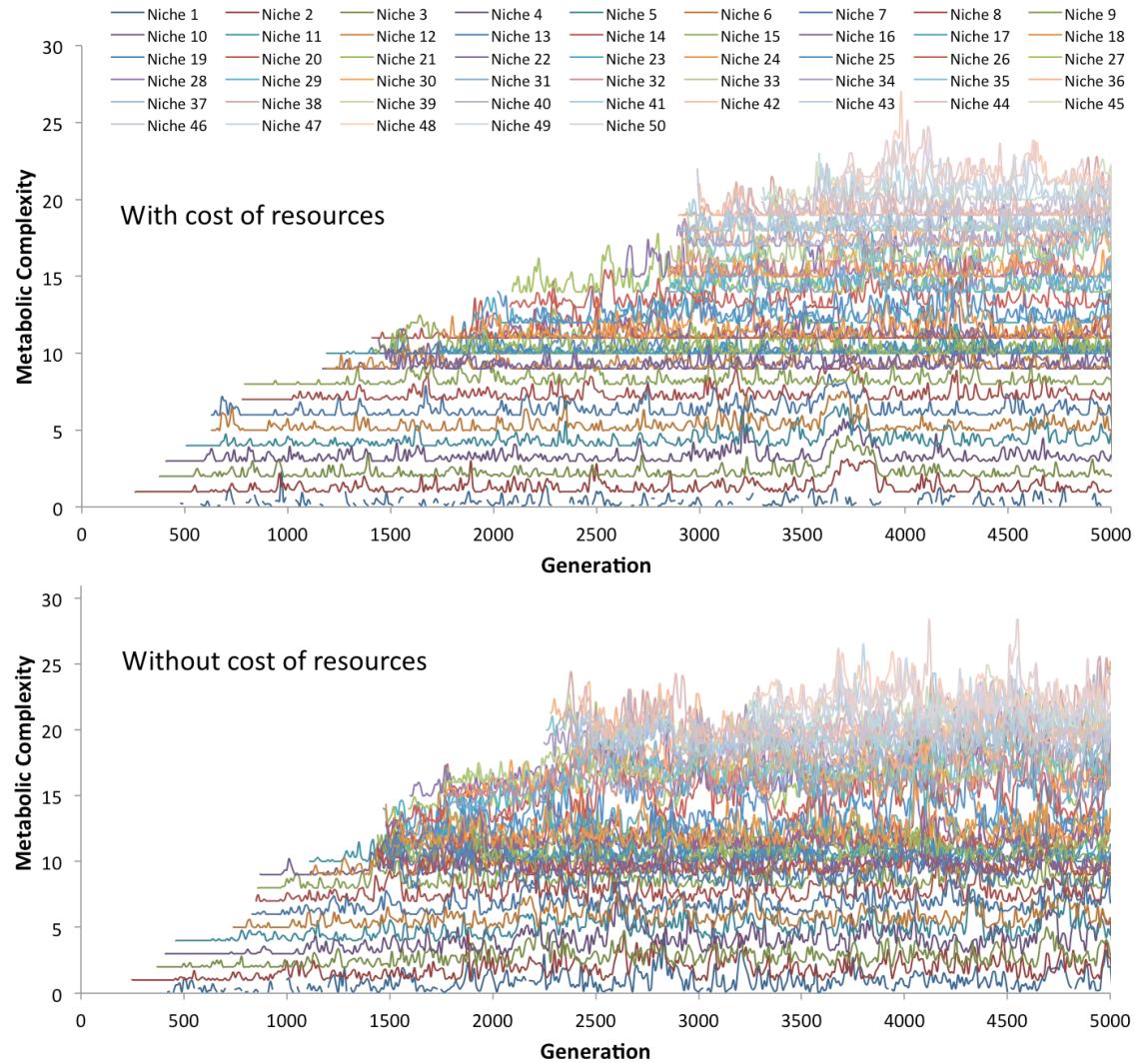


Figure 37. Case study examples of evolution in a system with an explicit pressure against complexity in the form of a cost of resources (top chart) and without a cost of resources (bottom chart). The cost of resources is expressed as a fitness penalty proportional to the number of genes expressed in the phenotype. The case studies shown here represent typical behaviour of the two systems. Although complexity data series for individual niches typically appear to be slightly more variable without a cost of resources, the system does not show the random walk in complexity that might be expected: mean complexity of individual niches appear to generally remain close to their respective complexity lower bound, implying that there is another implicit pressure against complexity in this system.

To achieve this we compared a positive control experiment in which the model contained a cost of resources (specifically, as described in section 3.5.2, a fitness value of 1 was subtracted for the organism's total fitness for each gene expressed in its phenotype), with a negative control experiment in which there was no cost of resources (i.e. no fitness was subtracted regardless of the number of expressed genes). Case study results are shown in Figure 37. It was expected that without a cost of resources, there would be nothing to restrict the metabolic complexity from following a random walk, and potentially increasing significantly. However, although complexity data series for individual niches typically appear to be slightly more variable in the negative control experiment, complexity within individual niches remained close to the complexity lower bound over the course of the simulation. This behaviour was observed in all repetitions of the negative control experiment. This implies that even without a cost of resources in the model, the implicit dynamics of the system still generate an inherent pressure towards complexity minimisation in individual niches.

We suggest two possibilities for this phenomenon:

1. First, longer pathways require more genes, and because mutation occurs with a per-locus probability, longer pathways have a greater chance of undergoing mutation. Because any random mutation to a working metabolic pathway is much more likely to break that pathway than not, then longer pathways are more likely to undergo mutation, and hence be broken, than shorter pathways. Shorter pathways are therefore more robust to the probability of deleterious mutations than longer pathways.
2. Second, even ignoring the increased probability of mutation associated with longer pathways, for any individual mutation, there is still a strong bias towards complexity minimisation inherent within the system. Specifically, imagine we have a 5-step pathway, and that there is one available 4-step pathway, and one available 6-step pathway, both accessible by a single mutation from the current 5-step pathway. Even with an equal number of simpler and more complex pathways as in this thought experiment (and with an equal probability of mutation adding or removing genes, as present in the system), the simpler pathway is much more likely to be evolved. The reason is that a mutation to find

the 6-step pathway must correspond to the right enzyme (from the 120 available) to add to the pathway in the right position. In contrast, the mutation to shorten the pathway only has 5 possible enzymes to choose from (i.e. those already in the pathway), one of which by definition must be able to be removed to find the shorter pathway. The result is that even though in this process, both the more complex and less complex pathways are selectively neutral, due to the inherent dynamics of the system, there is an inherent bias towards shortening any given pathway than lengthening it.

The results of this control experiment suggest that even without a cost of resources, other implicit properties of this type of system cause simplicity to be favoured, resulting in trends of complexity minimisation in individual niches. This implies that it is possible to further generalise the hypothesised mechanism by removing the condition that there must be a cost of resources to generate local trends of complexity minimisation, as other factors within such systems also inherently act to minimise complexity.

In summary, the results of the control experiments described above generally support the environmental dissociation complexity hypothesis, and further suggest that it may be generalised by removing the condition that a cost of resources is necessary to generate trends of complexity minimisation.

4.6 Discussion

The above control experiments provide significant support for the claims of environmental dissociation complexity. The positive control experiment robustly produces simultaneous complexity trends of a system-wide, general trend of increasing complexity of the most complex organisms and local trends of complexity minimisation in individual niches. Meanwhile, the negative control experiment without change in environmental dissociation and the negative control experiment without a sufficiently gradually varying environment both failed to generate similar increases in complexity, supporting the hypothesis that these are necessary components of the complexity generation mechanism. On the other hand, the negative control experiment without a cost of resources still showed an inherent preference for

simplicity in the model, suggesting that such a pressure will generally be present in this type of system even if resources are highly abundant, widening the scope of conditions in which the mechanism applies.

Given the gradually heterogeneous nature of the biosphere (Manahan 2004; McBride 1994; Solomon 2007), and the prevalence and constrained nature of DNA (Reaves et al. 2012; Kornberg and Baker 1992; Lindahl 1993; Grogan 1998; Marmur and Doty 1962), it seems highly likely that the biosphere commonly provides the conditions required in the mechanism necessary to produce a trend of increasing complexity of the most complex organisms. Indeed, in the biosphere such a trend is widely documented (Bedau 2009; McShea 1991). On the other hand, many evolutionary experiments that do not observe such a trend of increasing complexity (e.g. Spiegelman's experiments; Spiegelman et al. 1965) do not provide the conditions stated in the mechanism. As such, these results help to explain these observations. Moreover, they also help to explain how evolution in the biosphere is able to create a trend of increasing complexity in spite of the apparently ubiquitous pressure towards simplicity in evolution created by the cost of resources (and potentially other factors).

Although these results provide support for the hypothesis in this chapter, there are still a number of issues that remain outstanding that we will attempt to address in the rest of this section. We will first discuss in more detail how to link the theoretical description of complexity lower bounds with the chemical model used in this chapter – in particular, we will describe the state space of the model. Next, we will discuss a potential further control experiment that due to model limitations was not feasible to simulate: the effect on complexity trends of having a replicator that is *not* too constrained to change to adapt to environment change. We will then discuss the effect on the mechanism of having an environment that is possible for evolution to change by niche construction. Finally, following this, we will delve into what causes complexity lower bounds in the first place by examining the particular topological property of state space that causes them, and discuss how this relates to biological evolution. We will now discuss the nature of the state space in the metabolic model of this chapter, which will help to provide conceptual support for other topics described in this discussion.

Chapter 4: Complexity Lower Bounds

Although the chemical reaction network defined in the model describes the available state-change operators and their effects on environment state, this chemical reaction network is not the state space of the model. The reason is that a single environment state is defined by the position of multiple compounds in the chemical reaction network. If the environment only contained a single compound, then the reaction network would be identical to the state space. Because the environments in the model contain multiple compounds, then the state space is more complicated. The key difference between the state space of the model and the chemical reaction network is that each node in state space represents an environment – that is, a particular *collection* of chemical compounds. In the model, each environment contains 18 chemical compounds. To define the state space, we would simply create a node for every possible set of 18 compounds, and connect those sets to neighbouring sets by the individual chemical reactions that are required to change from one to another, just as in the Rubik’s cube analogy. As with the Rubik’s cube, we can visualise this state space (Figure 38), and in it define the ‘solved state’ – the small subset of environments in which the organisms’ replicator can reproduce (green bounded region, Figure 38).

How does the system behave in this conceptual model of state space? Given a replicator that is too constrained to change, as in the above simulations, then the set of compounds required by the replicator is fixed over time and cannot change via evolution. The result is that this bounded region in state space cannot move (in state space) over time. Meanwhile, at any time, the current external environment is a given set of chemical compounds, which is simply represented by a single node in state space (Figure 38, blue circle). Here, following environment change, evolution is forced to generate a metabolic pathway that recreates the niche required by the replicator (Figure 38 *b*) – and the minimum complexity of this pathway is defined by the shortest available path length in state space (i.e. the complexity lower bound).

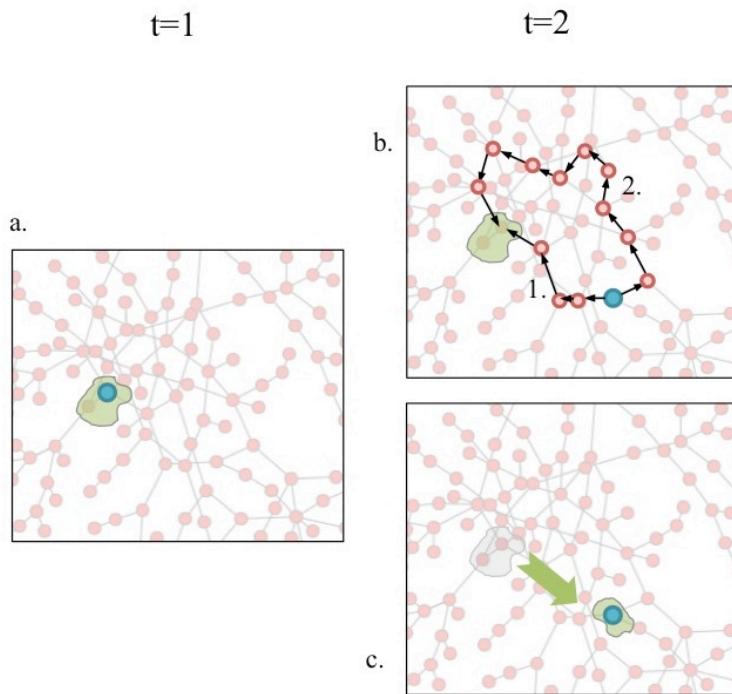


Figure 38. Hypothetical comparison of complexity evolution from a starting state (a) given a core replicator with fixed constraint (b) (similar to DNA) and no constraint (c). (a) The network represents the state space of the model. Nodes are chemical environments and links are chemical reactions. The blue node is the current environment, and the green area is the set of chemical compounds required by the organism's replicator for reproduction. When the environment changes, it does so on the network. (c) If the organism's replicator can be easily changed to cope with new environmental conditions, then the green area simply moves to the new environment. This poses no transformational problem evolution, and hence no increase in complexity is necessary. This would result in a new base replicator of evolution (i.e. other than DNA). (b) If the replicator has fixed constraint, then the replicator will always require the same input environment (the green area is fixed). In this case, environment change necessitates that evolution evolve machinery capable of transforming the new environment into that required by the organism's replicator. State space dictates how this can be achieved; some routes are short (e.g. 1) and hence less complex; others are longer and more complex (e.g. 2).

Chapter 4: Complexity Lower Bounds

Obviously, the complexity lower bound that this shortest path defines by no means dictates what pathway will actually evolve in any given environment, but does provide rigid constraints on what is possible.

With this conceptual model in mind, let us consider what might happen in the case of a control experiment in which organisms' have a replicator that, unlike DNA, can be easily change to be able to replicate given many different sets of environmental conditions; for example, by swapping key molecules within its structure for other molecules, as suggested by Wolfe-Simon et al. (2011), which was subsequently discredited by Reaves et al. (2012). Modelling such a system was beyond the scope of our investigation, and given that there is no evidence of such a capability in nature, was less of a priority than other control experiment simulations. However, given our conceptual description of state space in the metabolic model, we can at least logically analyse what the effect of having such a replicator might be.

In our Rubik's cube analogy of environmental dissociation complexity, having an replicator that, unlike DNA, can easily be changed to reproduce with a different set of environmental inputs effectively allows evolution to move the 'solved state'. Considering this ability in terms of the state space of the metabolic model, the result is that the green bounded region can be moved by evolution, thus removing the requirement to create a complex transformation (Figure 38 c). Changing the replicator in this way causes the environmental dissociation to be zero: no new function is (ever) required, and organismal complexity is not required to increase. The result would be that rather than bother to produce costly metabolic pathways, evolution could just substitute DNA for a similarly capable molecule that could simply operate with whatever chemical inputs the current environment happened to contain.

If this were possible in biological evolution, life could presumably expand into a vast array of niches without having to become more complex. In this case, we would presumably observe a large array of different base replicator molecules, each suited to the particular chemical makeup of its environment. This stands in contrast to reality, in which we observe a near uniform reliance on DNA as the base replicator of life across all niches (Reaves et al. 2012; Ridley 2009; Kornberg and Baker 1992) – and moreover, is commonly surrounded by an array of environmental transformation machinery (i.e. metabolisms) of varying

complexity across the different occupied niches in the biosphere (McShea 1994; Ridley 2009; Horton et al. 1996), similar to the behaviour predicted by the earlier simulations. Of course, this conceptual analysis can only take us so far; it would be useful to carry out further control experiments to test this predicted effect on the resultant complexity trends in the model.

4.6.1 Environmental dissociation and niche construction

We will now move on to briefly discuss how the results of this chapter, which have so far been described in terms of homeogenesis – organisms creating environmental adapters that sit between existing functionality and the external environment – can be integrated with niche construction, which is effectively a similar process but in which environment change occurs in the external environment as a product of behaviours. In short, here we suggest that changing the environment by niche construction will also be subject to complexity lower bounds, in a similar manner to homeogenesis – and therefore, that the results and predictions of this chapter can be similarly applied to systems that allow niche construction.

Because environmental dissociation and complexity lower bounds are inherent to the problem of environment change itself (Papadimitriou 2003) as opposed to being a property of homeogenesis, it follows that the same requisite trends of complexity should be generated by evolution regardless of the location of the environment change solution (e.g. internally within the organism, such as with homeogenesis and the models in this chapter, or externally such as with niche construction). It stands to reason that complexity lower bounds would limit minimum complexity solutions of both homeogenesis and niche construction in a similar manner. If evolution opts to undertake niche construction to adapt to environment change, then the greater the environmental dissociation between the current external environment conditions and the conditions required by the organisms' metabolism, the greater the number of environmental change steps that will be required by any process of niche construction – and in turn, the more complex the niche construction mechanism that will be required. Another way to illustrate the same point is to consider the metabolic model in this chapter. Although the model is described in terms of a evolving metabolism, it is sufficiently abstract that it could be directly interpreted instead as a problem of niche construction,

Chapter 4: Complexity Lower Bounds

in which organisms must evolve their phenotype to generate niche construction behaviours to adapt to changes in the external environment. This interpretation does not affect any of the assumptions in the model, and hence will not change the results.

This generalisation of the mechanism potentially increases the explanatory power of environmental dissociation. In this more general interpretation of the mechanism, as described earlier, trends of increasing complexity of the most complex organisms are generated when evolution is required to perform environment transformations that have inherent complexity lower bounds – only given this generalisation the resulting environment transformation algorithm can either be expressed as a set of internal adapters in the phenotype (e.g. a metabolic network), or as a set of behaviours and behavioural machinery used to transform the external environment (such as limbs, eyes, brains, innate behaviours, etc.).

4.6.2 Average path length

Finally, given that we have identified the magnitude of environmental dissociation, which is a property of two given points in a state space network, as the primary cause of complexity trends in this mechanism, we can ask: what network property of state spaces causes high environmental dissociation? Environmental dissociation is a property describing the shortest path distance between two given nodes in a state network. There are many possible network topologies of state network (e.g. fully connected, small world, ring network, etc.) that a given system could have – and some of these preclude high environmental dissociation and hence preclude high complexity lower bounds, and others may affect their magnitude. For example, given a state space network that was fully connected, then any environment state could be transformed into any other in a single step, precluding the possibility of complexity lower bounds higher than 1 – and hence carrying out evolution in such a system would be unlikely to generate any significant complexity by this mechanism. A key property that describes this network property is *average path length*. Average path length is defined as the average number of steps along the shortest paths for all possible pairs of network nodes (Newman 2009). For example the average path length in a fully connected network is 1, because all nodes are connected directly to all other nodes. In contrast,

chemical reaction networks generally have a much higher average path length (Papachristodoulou and Recht 2007): to get from one compound to another on a chemical reaction network in many cases requires multiple steps (Vogel 1974; Hammett 1970). In short, this reasoning implies that a state space with a high average path length (such as that provided in chemical systems, or spatially distributed systems) may be a further necessary condition to allow such trends of increasing complexity to evolve.

4.7 Evidence for complexity lower bounds in other models

We now move on from discussing the conceptual details of environmental dissociation complexity, and instead search for further evidence for it. To do this, we have carried out modelling to test whether significant environmental dissociation (and hence complexity lower bounds) are present in other evolutionary systems – in particular, a standard NAND gate model of functional evolution. The model system evolves circuits of logic gates to perform predefined calculations, building on the significant body of work in this area (Kashtan and Alon 2005; Kashtan, Noor, and Alon 2007; Milo et al. 2002). The model evolved solutions to the same, large set of arbitrary environment transformations, many times over from many different starting genotypes. We then observed the number of gates in the simplest circuits evolved that successfully completed each given transformation. We sought to test two hypotheses: That complexity lower bounds existed, and that transformations between environments of the same complexity (in this case, size of binary input) could result in complexity lower bounds of different magnitudes.

4.7.1 Methods

The model builds on the NAND logic gate model of Kashtan and Alon (Kashtan, Noor, and Alon 2007; Kashtan and Alon 2005). NAND gates were used because they are computationally universal, meaning that they can be combined to make any other type of logic gate. Circuits consisted of four layers of (8,4,2,1) NAND logic gates, making a total of 15 gates per circuit. There were 8 circuit inputs. Circuits evolved connections between gates, not allowing feedbacks. The goal of each circuit was to logically transform an input environment into

Chapter 4: Complexity Lower Bounds

an output environment using as few gates as possible. This could represent, for example, transforming the external environment into the conditions required by the organism's replicator. There were 32 niches. Each niche had a different external environment (but the same target output environment), representing a different necessary transformation for each niche. Each input environment consisted of 8 binary bit strings of length $L=8$. To be able to reproduce in that environment, the circuit had to convert all 8 of those bit strings, in order, to a specific 8-bit binary output sequence, defined by the target output environment. Because the output of each circuit was a single gate, all 8 of the $L=8$ bit strings per input environment were transformed to a single binary bit. Thus the target output environment was a single $L=8$ bit string. The size of the input environments was the same for every niche (8 $L=8$ bit strings). Circuits had a genotype of length $L=15$. Each gene corresponded to a specific gate. Each gene consisted of two integers $0 < v < 23$, each defining one of the two input locations for that gate. There were 15 gates and 8 circuit inputs, making a total of 23 possible input locations for each gate. (However, feedbacks were not allowed, so depending on the level of the gate, the number of available input locations was reduced.) The output of the single final gate was assumed to be the circuit output.

Although each genotype coded for 15 possible gates, circuits were measured on how many gates were actually used in the transformation ('effective gates'; after Kashtan and Alon, 2005). Gates that were not part of a connected route from circuit input to output were not included in the effective gate count.

Fitness (F) was calculated according to

$$F = \max(0, fs - fe)$$

where $fs=40$ was fitness awarded if the circuit successfully completed the required transformation in its niche, and fe was the number of effective gates used by the given circuit, thus providing a pressure against complexity.

Each of the 32 niches had a different, randomly chosen input environment that had to be converted into the target output environment - each representing a different but equal size (in terms of input and output bits) transformation. Niches were connected in space on a fully connected network. This represented organisms evolving across a series of spatially connected niches. Circuits could access new niches by migration, which occurred to offspring with a probability

$P_m=0.2$. Each niche was equally likely to be the migration location. Each niche had an $n=200$ carrying capacity. During each generation, reproduction occurred by (linear) fitness proportional selection until 200 offspring were produced, unless all circuits in that niche had zero fitness. Only circuits with $F>0$ fitness could reproduce (i.e. those that successfully performed the necessary transformation). Each simulation started with a single organism the first niche. Because this organism had to be able to reproduce (otherwise the initial population would simply go extinct), the first organism's genotype was selected by random search that continued until an organism was found that could successfully perform the transformation in its given niche. Using random search ensured that in each repetition, evolution began in different random starting position.

4.7.2 Results

Results are displayed in Figure 39. The aggregate results are a combination of 1000 repetitions, where each repetition was 10,000 generations. Each repetition used the same 32 niches, and thus necessitated the same 32 transformations. In all of the 1000 repetitions, one-gate solutions (i.e. circuits that achieved the necessary transformation with only one effective gate) were only found in 5 of the 32 niches. Solutions were found to all niches, but the number of gates in the simplest solutions found for each niche varied from 1 to 4.

The point of this exercise was to examine, to as great an extent as possible, the set of solutions available for 32 randomly chosen, equal size transformations in NAND circuit space. The results suggest that transformations of the same size can have a different minimum number of gates with which they can possibly be solved. This supports the existence of complexity lower bounds in logical systems, and also that complexity lower bounds can be different even from transformations that have equally complex inputs and equally complex outputs.

Chapter 4: Complexity Lower Bounds

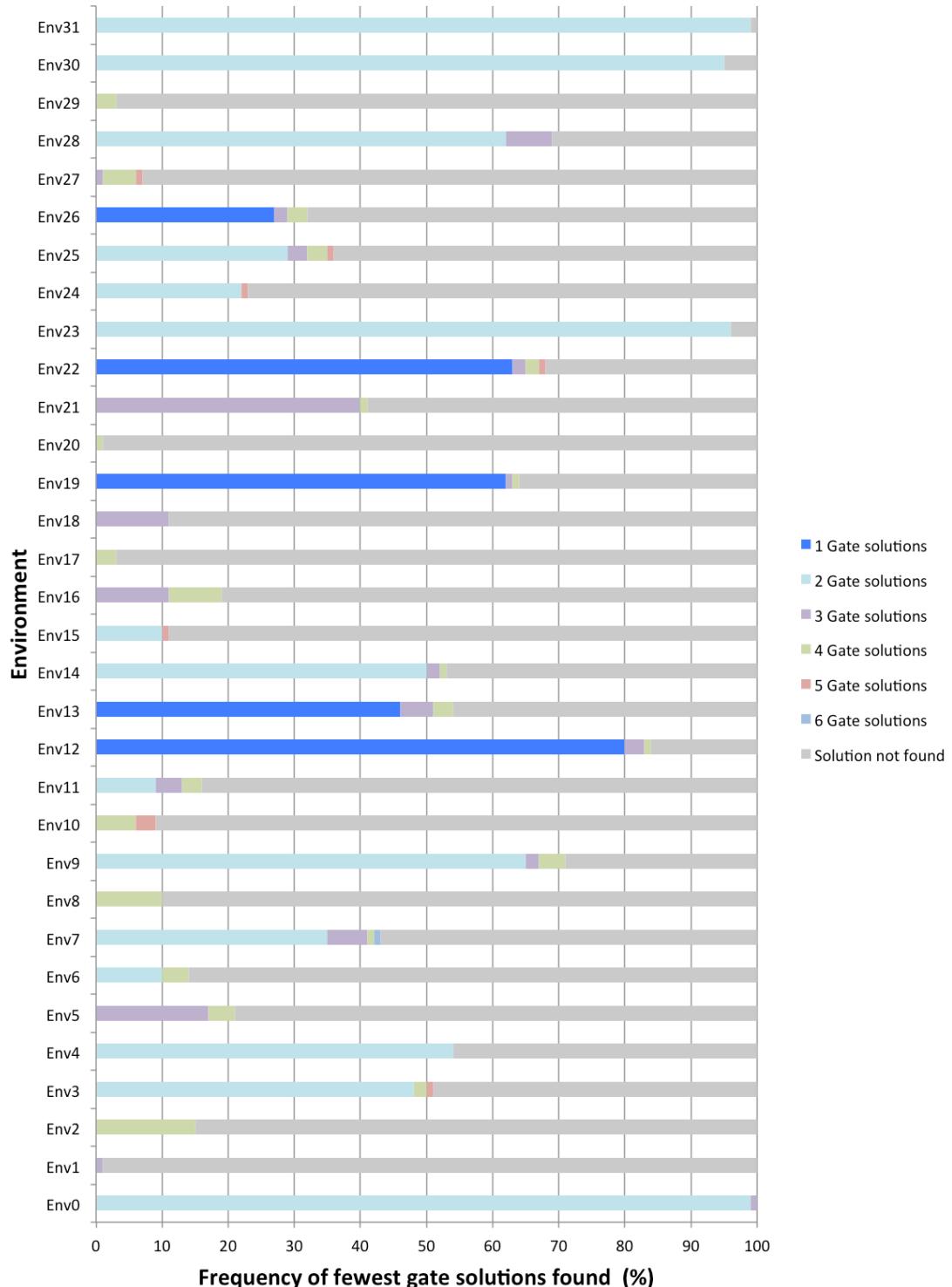


Figure 39. The frequency of simplest circuit sizes (in terms of number of gates) found after each repetition of the NAND gate model. Only 5 of the 32 transformations evolved solutions with a single gate. This suggests that complexity lower bounds exist in this common model of evolutionary function, and that transformations of different sizes can have different sized complexity lower bounds.

In this system, the search space of possible circuits is too large to search exhaustively. Although we carried out many repetitions of evolution for each transformation, we only sampled the space of available solutions. Therefore there remains a possibility that for the transformations for which no one gate solutions were found, such solutions do actually exist, but were not found by evolutionary search: Although this simulation gives us some idea of the likely distribution of complexity lower bounds in this system, it cannot provide us with absolute evidence that complexity lower bounds exist in this system. To rule out this possibility, for each of the 32 transformations in the experiment we exhaustively tried every possible one-gate circuit.

We found that many transformations did not have one-gate solutions – providing support for the results of the simulation. This definitively illustrates that some transformations of equal size inputs and outputs have minimum solutions with different numbers of gates, proving that complexity lower bounds are generally present in this system of NAND logic gates, and also that transformations of the same size can have different size complexity lower bounds.

Evidence from computer science (Papadimitriou 2003), electrical engineering (Hambley 2008) and chemical reaction networks (Vogel 1974; Papachristodoulou and Recht 2007; Hammett 1970) support the conclusions from the above model, and suggests that complexity lower bounds are a widespread phenomenon.

4.8 Conclusions

In this chapter we have introduced a novel complexity-driving mechanism in evolution. This framework shows how, given a difficult to change replicator such as DNA, the amount of environmental dissociation in a given niche dictates the minimum possible complexity of viable organisms in that niche – which given an environment with heterogeneous environmental dissociation, will generally result in a system-wide trend of increasing complexity of the most complex organisms (as new niches with higher environmental dissociation are colonised), coupled with local trends of complexity minimisation in individual niches. This work therefore links computational complexity theory (e.g. Papadimitriou 2003) to biological evolution. We have

Chapter 4: Complexity Lower Bounds

also identified the minimum necessary conditions for this complexity driving process to be expressed in evolution, which are simply a) a replicator whose replication functionality is too difficult to change, and b) a gradually varying environment with heterogeneous environmental dissociation. (We also include c) a selection pressure against complexity, although experiments suggest that this is may be an inescapable inherent part of the dynamics of the system and so may not need to be included in the list of necessary conditions.) Evidence that a) DNA has a very small range of conditions in which it can replicate (Lindahl 1993; Grogan 1998; Reaves et al. 2012; Marmur and Doty 1962), is the base replicator of all known life (Kornberg and Baker 1992; Reaves et al. 2012; Ridley 2009); b) DNA has functionality that is extremely preserved across all life (Kornberg and Baker 1992); c) the biosphere is widely heterogeneous and often changes gradually in time and space (Manahan 2004; McBride 1994; Chester 2009; Solomon 2007), implies that these conditions are routinely met in natural biological systems.

Why does natural evolution display a characteristic trend of increasing complexity of the most complex organisms? The chain of causation that brings about such trends that is suggested by this framework is:

1. evolution is constantly pushed by population pressure and temporal environment change into new environment conditions
2. which, because life on earth is based on a highly constrained replicator, often results in evolution favouring converting the environment back to conditions favoured by its replicator, as opposed to the replicator itself
3. and such environment conversion problems have inherent lower bounds on the complexity of minimum possible solutions, whose magnitude depends on the environmental dissociation of the two environments being converted between
4. which in turn results in requisite minimum organismal complexity for any given niche
5. and combines with the inherent favouring of simplicity within evolution (for example, due to cost of resources, etc.) to produce a general, system-wide trend of increasing complexity of the most complex

organisms, coupled with local trends of complexity minimisation within individual niches.

We suggest the name environmental dissociation complexity for this framework of complexity evolution. Environmental dissociation complexity also enables us to make some general, speculative predictions about complexity trends in general evolutionary systems, such as in life on other planets, or in artificial evolution:

1. There is nothing inherent within ENS, or evolution as a whole to *force* complexity to occur; however, given a constrained replicator, complexity change will be controlled by environmental dissociation as evolution spreads to different environments.
2. Therefore, life based on DNA (or another similarly constrained replicator), given a gradually varying environment with heterogeneous environmental dissociation will generally cause the characteristic types of multifaceted, environmentally mediated complexity trends observed in these simulations, including a system-wide general trend of increasing complexity of the most complex organisms and local trends of complexity minimisation.
3. Therefore, if we were to 'replay the tape of life (Gould 2000)' on earth or on another planet (starting with DNA, or another similarly constrained replicator), we should generally expect to observe a similar overall pattern of complexity trends as we observe in the evolutionary record and in these simulations.
4. However on other planets significant complexity could still be hampered even if life begins with DNA (or a similarly constrained replicator) if the environment is not sufficiently heterogeneous, or if the environment does not provide some path of gradual environmental change to allow gradual increase of environmental dissociation, and hence gradual increase of evolved complexity.

It is important to stress that although the presence of complexity lower bounds imply that in many cases complexity increases may be driven by the particular environment transformation required in that niche, it is clear that other factors

Chapter 4: Complexity Lower Bounds

not included here can affect and control complexity evolution. The existence of other such factors means that we should not necessarily expect to see a clear pattern of repeated homeogenesis as observed in the simple models in this chapter. For example, Lane argues that genome size can also act as a limiting factor on organismal complexity (Lane 2002). Lane argues that genome size was a key limitation that held back the evolution of eukaryotes from prokaryotes. In short, prokaryotes had energy limitations that in turn limited their genome size – and so eukaryotes (necessitating greater complexity and hence longer genomes) could only be evolved once this energy barrier was transcended (Lane 2002). Interestingly though, although this view focuses on genome length as a control on complexity, it may implicitly assume the existence of complexity lower bounds. If there were no complexity lower bounds on the eukaryotic niche, then presumably there should be arbitrarily simple viable solutions to the eukaryotic niche that would, therefore, not be affected by the genome size limitation. It is only if the eukaryotic niche has some lower bound of complexity that a limited genome size would deny evolution access to this niche. Interpreting this example from a general complexity lower bound perspective, in some cases it may be that the transformation required by a neighbouring, unoccupied niche (such as the eukaryote niche) requires an increase in complexity (due to its complexity lower bound) that is simply not possible given the genome size limitations. In this case, evolution would be halted, caught between two limitations: the need for extra complexity to carry out the new environment transformation (i.e. satisfy the complexity lower bound), and the limit on extra complexity due to genome size limitations.

The point of this illustration is to demonstrate that many other physical factors no doubt limit organismal complexity in different ways (i.e. as complexity roadblocks) in combination with complexity lower bounds, resulting in more complicated trends in complexity than produced by the mechanisms described this chapter. In short, although we have described here a mechanism for complexity generation in evolution whose causal factors seem to be significantly widespread, this theory by no means discounts all other existing theories for complexity generation, which may well act alongside, or in concert with the mechanism we describe here.

4.8.1 Key Results

The key claim of this chapter is that

Environmental change motivates evolutionary change, but not necessarily any increase in complexity. However, given

- a. an organism with a replicator that can replicate in some small subset of environmental conditions, and whose replicator cannot feasibly be changed to replicate in conditions outside of this subset;
- b. an environment with heterogeneous environmental dissociation whose conditions change sufficiently gradually;
- c. an inherent selection pressure against complexity such as a cost of resources

then as competition forces evolution to leave the original environment (a), and colonise new environments (b), the magnitude of environmental dissociation of a new environment will dictate the minimum possible complexity of viable organisms in that environment, resulting in a system-wide trend of increasing complexity of the most complex organisms, coupled with local trends of complexity minimisation in individual environments, caused by (c).

This claim is supported by the following results.

- Evidence that an evolutionary system that satisfies (a), (b) and (c) results in a general, system-wide trend of increasing complexity of the most complex organisms coupled with local trends of complexity minimisation in individual niches is provided by:
 - the positive control experiment results (section 4.5.2.1), that show a extremely significant increase in system-wide, mean organismal complexity, and
 - Figure 31 and Figure 32, that illustrate case studies of this general result, and show robust trends of local complexity minimisation in individual niches.
- Evidence that environmental dissociation dictates the minimum amount of complexity in a given niche (and hence that heterogeneous environmental dissociation is a necessary condition to generate the observed complexity trends, and that environmental

Chapter 4: Complexity Lower Bounds

change alone is not necessarily sufficient to generate increases in complexity) is provided by results from negative control experiment 1 (section 4.5.2.2 - Figure 33 and Figure 34).

- Evidence that a sufficiently gradually changing environment is a necessary condition to generate the observed complexity trends is provided by results from negative control experiment 2 (section 4.5.2.3 - Figure 35).
- Evidence that the relative amount of selection pressure against complexity controls the complexity minimisation trends is provided by negative control experiment 3 (section 4.5.2.4). The results of this section suggest that a selection pressure towards complexity minimisation may be an inherent consequence of the internal dynamics of this system, and so could potentially be removed from the necessary conditions.

Chapter 5: Summary and Conclusions

5.1 Research aims

Biology displays a general trend of increasing complexity of the most complex organisms (McShea 1994; McShea 1991; Bedau 2009). However, experimental attempts have generally failed to reproduce this type of open-ended complexity in biological systems or in simulations. As a result, how evolution generates this type of complexity is an open question (Bedau 2009; McShea 2009).

This study set out to determine:

how evolutionary theory can be refined to better explain the apparently conflicting trends of generally increasing maximal complexity observed in nature, and complexity minimisation commonly observed in experiments.

The research was focused in two areas in particular that resulted in two specific research questions. The first of these questions related to evolution by combining functions. Combining functions is a common theme in a range of evolutionary mechanisms and models capable of evolving non-decomposable functions (and non-decomposable functions are potential roadblocks to the evolution of complexity); furthermore, many of these mechanisms also produce transition like behaviour (Watson 2006; Lenski et al. 2003; Maynard Smith and Szathmary 1997). However, despite significant research in this area, there are two important outstanding problems. First, there is no agreed theoretical framework for evolution by combining functions that identifies its central mechanism and connects it with other related theories. Second, existing mechanisms involving combining functions have struggled to carry out evolution open-endedly over multiple levels of organisation. Accordingly, the first more detailed question addressed in this thesis was

what is the underlying evolutionary mechanism of evolution by combining functions, and what enables natural evolution to perform it recursively across multiple scales of organisation?

Chapter 5: Summary and Conclusions

The second more specific research question in this dissertation was related to what, in the absence of any roadblocks, drives evolution to complexity. Many theories have been proposed to account for large-scale trends in biological complexity, but most are not supported by computational or mathematical models and hence lack rigorously defined predictions, and there is no agreement that any sufficiently explain the trends observed (McShea 1991; Bedau et al. 2000; McShea 2009). In particular, there is no consensus on what explains the trend of increasing maximal complexity in evolution. To get to the root of this problem, the second more detailed research question in this dissertation was therefore

what type of environmental changes, and under what evolutionary conditions, necessarily produce adaptations that are more complex rather than merely different?

5.2 Summary of research findings

This study has identified two new contributions to evolutionary theory that can provide potentially important progress in understanding trends of complexity in evolution:

1. A theoretical framework for evolution by combining functions that includes a mechanism capable of recursively transitioning to progressively higher scales of organisation, and
2. A new mechanism for the generation of complexity in evolution that can necessitate changes in complexity, and as a result causes characteristic trends of complexity that resemble the system-wide general trend of increasing maximal complexity observed in the biosphere, and local trends of complexity minimisation observed in many evolutionary experiments.

These two contributions relate to the two more detailed research questions respectively. We will briefly describe the findings of each below. First, the theoretical framework for combining functions unites three major existing theories of non-decomposable function evolution (exaptation, building block mechanisms and tinkering) and shows that they essentially describe two core processes that constitute two separate types of exaptation:

1. *shift exaptation*, where the shift in function occurs at the level of the trait itself (as in the current description of exaptation), and
2. *combinatorial exaptation*, where the shift in function occurs at a organisational level above the individual traits being combined (i.e. at the level of the combination of traits)

Using this understanding, the framework illustrates that combinatorial exaptation is the central mechanism of evolution by combining functions. We also provided a model of combinatorial exaptation, building on previous work and building block models, to explore its properties. First, the model shows that for combinatorial exaptation to feasibly occur, some mechanism of 'encapsulation' is required within the genotype-phenotype map that practically allows whole phenotypic traits to be redeployed as integrated units in the phenotype – and in particular, describes how this problem is caused by increasing 'burden' in the form of 'internal selection'. Moreover, the model shows that to allow combinatorial exaptation to occur potentially open-endedly across multiple scales of organisation, the encapsulation mechanism must also evolve in a similarly open-ended manner. Finally, the model provides a solution to this problem. It shows that if the modular hierarchical structure of the genotype and that of the phenotype are somehow linked, this is sufficient to act as an open-ended system of encapsulation, and hence allow combinatorial exaptation to occur recursively and potentially open-endedly. Moreover, the model shows that physical constraints placed on the genotype-phenotype map by the type of development that occurs in biological organisms can introduce such a link.

The second main contribution of this dissertation is in the form of a theoretical framework and set of models based on a particular type of combinatorial exaptation, that we term homeogenesis. The framework illustrates that homeogenesis is a potentially novel mechanism of adaptation to environment change because, unlike most existing mechanisms, does not change an organism's existing function or its external environment, but instead occurs by organisms evolving an internal environmental 'adapter' that converts the new external environment into conditions expected by its existing functionality.

Chapter 5: Summary and Conclusions

The evolution of C₄ photosynthesis represents a well-studied biological example of this evolutionary mechanism. The computational model in chapter 3 shows that in adapting organisms to environment change, homeogenesis commonly creates internal representations of previously experienced external environment conditions. Thus, we show that homeogenesis can potentially explain the poorly-understood observation that biological organisms commonly contain conditions within their metabolisms that appear to represent the ancient environments in which their ancestors lived (Mulkidjanian et al. 2012) in a more detailed and mechanistic manner than Macallum's 'chemistry conservation principle' (which is commonly referred to and simply states that this occurs because the chemical traits of organisms are more conservative than the changing environment; Macallum 1926; Mulkidjanian et al. 2012).

Most importantly, our simulation results show that in evolution by homeogenesis, some types of environment change essentially necessitate increases in organismal complexity, thus acting as a mechanism that can create robust complexity trends in evolution. Our simulations show that different types of environment change cause characteristic complexity change that resemble some common characteristic patterns in nature, including a system-wide general trend of increasing maximal complexity, and local trends of complexity minimisation. We carried out further analysis to explain these results. The resulting theoretical framework identifies that the key factor that dictates complexity trends in this system is the presence of environmental dissociation, which creates inherent complexity lower bounds on the complexity of possible solutions for survival in any given niche. The framework connects this property to similar lower bounds known in algorithmic complexity theory.

Bringing these results together, we defined the mechanism of environmental dissociation complexity, a theoretical framework surrounding complexity lower bounds that describes how these lower bounds affect evolution. We used this theory to address one of the key motivating observations in this thesis: conflicting observations of a general trend of increasing complexity of the most complex organisms in nature and common trends of complexity minimisation in experiments. Our results show that having a replicator with difficult to change functionality (such as DNA), will often result in it being

easier for evolution to convert the external environment into the conditions required by this replicator, rather than change the functionality of the replicator itself (i.e. undertake homeogenesis or counteractive niche construction). The result is that in any given niche, evolution is required to generate a chemical or physical algorithm that is capable of converting between these two environments. Our simulation models to show that, in agreement with algorithmic complexity theory, each niche therefore introduces complexity lower bounds on the complexity of possible algorithms available, and hence enforces a niche-specific floor on the minimum complexity of viable organisms in that niche.

Crucially, we showed that given a set of conditions that routinely occur in natural evolution, the mechanism of environmental dissociation complexity robustly produces a general, system-wide trend of increasing complexity of the most complex organisms, coupled with local trends of complexity minimisation within individual niches. We identified the set of conditions that produce these trends as (a) organisms' have a replicator that can reproduce in a small subset of environmental conditions and cannot be feasibly changed to reproduce outside of those conditions; (b) evolution occurs in an environment with heterogeneous environment dissociation that varies sufficiently gradually, and (c) the system contains an inherent selection pressure against complexity such as a cost of resources. Given the weight of evidence that DNA satisfies (a) (Reaves et al. 2012; Lindahl 1993; Grogan 1998; Marmur and Doty 1962; Kornberg and Baker 1992), and that natural environments satisfy (b), and that evolution inherently contains selection pressures against complexity (e.g. Lane 2010), we propose that these conditions are routinely met in natural biological systems. This work thereby helps to ease tensions between conflicting observations of general trends of increasing complexity of the most complex organisms in the biosphere and complexity minimisation in experiments.

5.3 Contributions and Implications

Taken together, the results of this study illustrate how the current theory of evolutionary could be refined to better explain the origin and nature of complexity trends observed in natural biological systems and in evolutionary experiments.

Chapter 5: Summary and Conclusions

This work helps to clarify existing knowledge of complexity evolution by providing a new general theoretical framework for combining functions that unites tinkering, building block mechanisms and exaptation, and includes a mechanism capable of joining functions recursively across multiple levels of organisation. This supports existing work showing that evolution by combining functions may be an important biological process, and shows how it can occur in a potentially open-ended manner across multiple organisational levels.

The research presented here also contributes to the understanding of complexity evolution by providing a new mechanism for complexity generation, termed environmental dissociation complexity. This mechanism links evolutionary biology to known causes of complexity in mathematics and computer science. Furthermore, under specific conditions that are likely present in nature, the mechanism produces a general, system-wide trend of increasing complexity of the most complex organisms, coupled with local trends of complexity minimisation within individual niches. This helps to ease the tension between apparently conflicting observations of a general trend of increasing complexity of the most complex organisms in the biosphere, and common observations of apparent complexity minimisation in evolutionary experiments. Environmental dissociation complexity has broad implications. For example, one implication is that the difficulty associated with changing the function of DNA might have played a vital role in generating complexity in evolution on earth (with a more evolvable replicator, functional change might have been more available, potentially allowing adaptation without complexity increase). Environmental dissociation complexity also implies that it may be possible, in theory at least, to predict the complexity lower bound for viable organisms in a given niche, given knowledge of its core replicator and the state space of the surrounding environment.

This dissertation also contributes to organismal biology. First, it provides an expanded theory of exaptation that contains two distinct types of exaptation. This could have important implications for understanding the mechanism of evolution in a range of evolutionary events attributed to exaptation, and provide deeper understanding of the place of exaptation in evolution. This research also contributes to organismal biology by providing, through homeogenesis, a mechanism of adaptation that can potentially explain the observation that organisms often contain internalised versions of previously

experienced environments. This could have a significant impact on current research that uses this observation to deduce environmental conditions of early life, in addition to wider related research (Mulkidjanian et al. 2012; Mulkidjanian and Galperin 2007).

5.4 Limitations

Finally, a number important limitations to this work need to be considered. First, in cases where clear examples from evolutionary biology are not provided, much of this work relies on evidence from models that are abstract representations of biological systems. These models build on biological fact, and illustrate the capabilities of non-teleological adaptive processes. However, in many cases there is much work to be done before their conclusions can be empirically supported in biological systems. Accordingly, such results should only be applied to real biological systems with the appropriate amount of consideration and qualification.

Second, there are numerous limitations of the models used. In both combinatorial exaptation and homeogenesis models, function was not explicitly included. As a result, further assumptions were necessary about the nature of function that could have been avoided if function was included explicitly. In a similar manner, development was not explicitly included in either model; this again requires assumptions about development to be added (especially in the combinatorial exaptation model where developmental constraints cause linkage between genotype-phenotype map and phenotype structure). Again, including development explicitly would have increased the confidence in the simulation results. In a similar manner, in the model of homeogenesis, because the intention was to study its process, neither adaptation by changing the existing function nor niche construction were allowed. This restricts understanding of what conditions cause homeogenesis over other mechanisms of adaptation, and hence how likely it is to occur in natural evolution. Furthermore, both combinatorial exaptation and homeogenesis models had intentionally contrived environments that were chosen to illustrate particular properties of those respective processes. Although this was intentional, this prescribed nature limits understanding of how combinatorial exaptation and homeogenesis would occur in larger, less restricted and more realistic environments.

Chapter 5: Summary and Conclusions

The theoretical frameworks described here also have significant limitations. For example, there are many other mechanisms capable of non-decomposable function evolution described in the literature that could potentially be, but were not incorporated into the analysis that resulted in the combinatorial exaptation framework. In a similar manner, environmental dissociation complexity describes how complexity lower bounds potentially affect evolution, but it is not particularly well connected with other theories of complexity evolution. In particular, environmental dissociation complexity has an obvious omission, which is that it does not take into account energy: Complexity evolution is described as the result of building machinery to convert one environment into another; however, in reality that machinery must also be powered, thus requiring other transformation pathways that glean energy from some environmental source, and transport it to sites within the metabolism as it is needed.

5.5 Further work

Considerably more work will be needed to determine the extent to which the mechanisms described in this dissertation apply in natural evolution. In more detail, these findings provide the following insights for future research:

1. It would be interesting to test whether functions in biological organisms known to have been produced by combining functions (e.g. Alcock et al. 2010; Flicek 2013) were facilitated by gene regulatory networks in the manner predicted in the combinatorial exaptation model, and also the extent to which gene regulatory network hierarchical structure imitates phenotypic structure as the model predicts.
2. It would also be interesting to carry out further modelling of combinatorial exaptation in a less restricted environment, where the selection pressures necessary to evolve a given complex function were not necessarily present in a single niche, but were distributed across a heterogeneous spatiotemporal environment. The resulting spatiotemporal patterns could then be compared with those observed in biological adaptive radiations; moreover, the model could be used to explore links between evolutionary and ecological models and theory.
3. Further research is also warranted on understanding the extent of homeogenesis in nature. The evolution of C₄ photosynthesis is a well-

studied example of evolutionary innovation, and that it apparently occurred by homeogenesis implies that homeogenesis may be a common process in biological evolution. Moreover, the complexity trend results of environmental dissociation complexity theory imply that homeogenesis may be common in the evolutionary record. A comprehensive review would help to elucidate this possibility.

4. Finally, it would be interesting to test the environmental dissociation complexity hypothesis in a real biological system, where simple bacteria were evolved across a range of environments with varying environmental dissociation, while measuring their capability for homeogenesis, the resultant effects on their evolving metabolism, and the resulting trends in metabolic complexity.

5.6 Concluding remarks

In summary, the theory, simulations and analytic results in this dissertation demonstrate (a) how evolution can, when complexity is beneficial, scale to complexity over multiple organisational levels, and (b) the conditions in which complexity is beneficial in evolution. These models describe a set of phenotypic, ontogenetic and environmental conditions that are generally present in biological evolution, in which evolution consistently generates an overall trend of increasing complexity of the most complex organisms.

Chapter 5: Summary and Conclusions

List of References

- Abzhanov, A., M. Protas, B. R. Grant, P. R. Grant, and C. J. Tabin. 2004. "Bmp4 and morphological variation of beaks in Darwin's finches." *Science* 305 (5689): 1462-65.
- Adami, C. 2002. "What is complexity?" *BioEssays* 24 (12): 1085-94.
- Adami, C., C. Ofria, and T. C. Collier. 2000. "Evolution of biological complexity." *Proceedings of the National Academy of Sciences* 97 (9): 4463-68.
- Alcock, F., A. Clements, C. Webb, and T. Lithgow. 2010. "Tinkering inside the organelle." *Science* 327 (5966): 649-50.
- Aldana, M., E. Balleza, S. Kauffman, and O. Resendiz. 2007. "Robustness and evolvability in genetic regulatory networks." *Journal of Theoretical Biology* 245 (3): 433-48.
- Altenberg, L. 1994. "The Evolution of evolvability in genetic programming." In L. Spector, W. B. Langdon, U. O'Reilly and P. J. Angeline (Eds.) *Advances in Genetic Programming* 3: 47-74. MIT Press.
- Arthur, W. B. 1993. "Why do things become more complex." *Scientific American* 268 (5): 144.
- Arthur, W. B., and W. Polak. 2006. "The evolution of technology within a simple computer model." *Complexity* 11 (5): 23-31.
- Auerbach, J. E., and Bongard, J. C. 2014. "Environmental influence on the evolution of morphological complexity in machines." *PLoS Computational Biology* 10 (1): e1003399.
- Badyaev, A. V. 2009. "Evolutionary significance of phenotypic accommodation in novel environments: an empirical test of the Baldwin effect." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364 (1520): 1125-41.
- Barve, A., and A. Wagner. 2013. "a latent capacity for evolutionary innovation through exaptation in metabolic systems." *Nature* 500 (7461): 203-6.
- Bedau, M. A. 2009. "The evolution of complexity." In A. Barberousse, M. Morange, T. Pradeu (eds) *Mapping the Future of Biology*, 111-30. Springer.
- Bedau, M. A., J. S. McCaskill, N. H. Packard, S. Rasmussen, C. Adami, D. G. Green, T. Ikegami, K. Kaneko, and T. S. Ray. 2000. "Open problems in artificial life." *Artificial Life* 6 (4): 363-76.

List of References

- Bedau, M. A., E. Snyder, C. T. Brown, and N. H. Packard. 1997. "A Comparison of evolutionary activity in artificial evolving systems and in the biosphere." In P. Husbands and I. Harvey (eds) *Proceedings of the Fourth European Conference on Artificial Life*, 125-34. MIT Press.
- Behe, M. J. 2009. "Irreducible complexity: Obstacle to Darwinian evolution." *Philosophy of Biology: An Anthology*, 427. Wiley.
- Bejerano, G., C. B. Lowe, N. Ahituv, B. King, A. Siepel, S. R. Salama, E. M. Rubin, W. J. Kent, and D. Haussler. 2006. "A distal enhancer and an ultraconserved exon are derived from a novel retroposon." *Nature* 441 (7089): 87-90.
- Bennett, C. H. 1995. *Logical depth and physical complexity*. Springer Vienna.
- Ben-Or, M.. 1983. "Lower bounds for algebraic computation trees." In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, 80-86. ACM.
- Bernard, C. 1865. *Introduction à l'étude de la médecine expérimentale*. JB Baillière et fils.
- Berner, R. A., J. M. VandenBrooks, and P. D. Ward. 2007. "Oxygen and evolution." *Science* 316 (5824): 557-58.
- Bhagavan, N. V. 2002. *Medical biochemistry*. Elsevier.
- Bird, A. P. 1995. "Gene Number, Noise Reduction and Biological Complexity." *Trends in Genetics* 11 (3): 94-100.
- Blount, Z. D., Borland, C. Z., and Lenski, R. E. 2008. "Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*". *Proceedings of the National Academy of Sciences* 105 (23): 7899-7906.
- Bock, G., and J. Goode. 2007. *Tinkering: The Microevolution of Development* 284-287. Wiley.
- Brosius, J. 1999. "RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements." *Gene* 238 (1): 115-34.
- Budd, G. E. 2006. "On the origin and evolution of major morphological characters." *Biological Reviews* 81 (4): 609-28.
- Cannon, W. B. 1935. "Stresses and strains of homeostasis." *The American Journal of the Medical Sciences* 189 (1): 13-14.

- Cannon, W. B. 1929. "The sympathetic division of the autonomic system in relation to homeostasis." *Archives of Neurology and Psychiatry* 22 (2): 282.
- Carroll, S. B. 2001. "Chance and necessity: The evolution of morphological complexity and diversity." *Nature* 409 (6823): 1102-9.
- . 2005. *Endless forms most beautiful: The new science of Evo Devo and the making of the animal kingdom*. WW Norton & Company.
- Channon, A. 2001. "Passing the ALife test: Activity statistics classify evolution in Geb as unbounded." In J. Kelemen and P. Sosik (Eds.) *Advances in Artificial Life*, 417-426. Springer Berlin Heidelberg.
- Channon, A. D., and Damper, R. I. 2000. "Towards the evolutionary emergence of increasingly complex advantageous behaviours." *International Journal of Systems Science* 31 (7): 843-860.
- Chester, R. 2009. *Marine geochemistry*. John Wiley & Sons.
- Claessens, L. P. A. M., P. M. O'Connor, and D. M. Unwin. 2009. "Respiratory evolution facilitated the origin of Pterosaur flight and aerial gigantism." *PLoS ONE* 4 (2): e4497.
- Clune, J., J. B. Mouret, and H. Lipson. 2013. "The evolutionary origins of modularity." *Proceedings of the Royal Society B: Biological Sciences* 280 (1755).
- Cohn, M. J., and C. Tickle. 1999. "Developmental basis of limblessness and axial patterning in snakes." *Nature* 399 (6735): 474-79.
- Copley, S. D. 2000. "Evolution of a metabolic pathway for degradation of a toxic xenobiotic: The patchwork approach." *Trends in Biochemical Sciences* 25 (6): 261-65.
- Crispo, E.. 2007. "The Baldwin effect and genetic assimilation: Revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity." *Evolution* 61 (11): 2469-79.
- Crutchfield, J. P., and K. Young. 1989. "Inferring statistical complexity." *Physical Review Letters* 63 (2): 105.
- Darwin, C. 1859. *On the origins of species by means of natural selection*. Murray, London.
- Davidson, E. H. 2010. "Emerging properties of animal gene regulatory networks." *Nature* 468 (7326): 911-20.
- Davidson, E. H., and D. H. Erwin. 2006. "Gene regulatory networks and the evolution of animal body plans." *Science Signaling* 311 (5762): 796.

List of References

- Davidson, E., and M. Levin. 2005. "Gene regulatory networks." *Proceedings of the National Academy of Sciences* 102 (14): 4935–4935.
- Dawkins, R.. 1999. *The Extended Phenotype: The Long Reach of the Gene*. Oxford University Press.
- . 2004. "Extended phenotype—but not too extended. A reply to Laland, Turner and Jablonka." *Biology and Philosophy* 19 (3): 377–96.
- De Oliveira, P. P. B. 1994. "Simulation of exaptive behaviour." In *Parallel Problem Solving from Nature—PPSN III*, 354–64. Springer.
- De Visser, J. A., and S. F. Elena. 2007. "The evolution of sex: empirical insights into the roles of epistasis and drift." *Nature Reviews Genetics* 8 (2): 139–49.
- Dobzhansky, T.. 1970. *Genetics of the evolutionary process*. vol. 139. Columbia University Press.
- Draghi, J., and G. P. Wagner. 2008. "Evolution of evolvability in a developmental model." *Evolution* 62 (2): 301–15.
- Dwyer, M. B., S. Elbaum, S. Person, and R. Purandare. 2007. "Parallel randomized state-space search." In *Software Engineering, 2007. ICSE 2007. 29th International Conference on*, 3–12. IEEE.
- Edmonds, B. 1995. "What is complexity? The philosophy of complexity per se with application to some examples in evolution." *The Evolution of Complexity*. Kluwer, Dordrecht.
- Edwards, G., and D. A. Walker. 1983. *C₃, C₄: Mechanisms, and cellular and environmental regulation, of photosynthesis*. Blackwell Scientific Publications.
- Ehleringer, J. R., R. F. Sage, L. B. Flanagan, and R. W. Pearcy. 1991. "Climate change and the evolution of C₄ photosynthesis." *Trends in Ecology & Evolution* 6 (3): 95–99.
- Erwin, D. H., and E. H. Davidson. 2009. "The evolution of hierarchical gene regulatory networks." *Nature Reviews Genetics* 10 (2): 141–48.
- Fedonkin, M. A. 2003. "The origin of the Metazoa in the light of the Proterozoic fossil record." *Paleontological Research* 7 (1): 9–41.
- Fisher, D. C. 1986. "Progress in organismal design." In *Patterns and Processes in the History of Life*, 99–117. Springer.
- Fisher, R. A. 1958. *The genetical theory of natural selection*. Dover Publications (New York).

- Fliceck, P. 2013. "Evolutionary biology: the handiwork of tinkering." *Nature* 500 (7461): 158-59.
- Forrest, S., and M. Mitchell. 1993. "Relative building-block fitness and the building-block hypothesis." In *Foundations of Genetic Algorithms 2*. Kaufmann Publishers.
- Gabora, L., E. O. Scott, and S. Kauffman. 2013. "A quantum model of exaptation: incorporating potentiality into evolutionary theory." *Progress in Biophysics and Molecular Biology* 113 (1): 108-16.
- Gell-Mann, M., and S. Lloyd. 1996. "Information measures, effective complexity, and total information." *Complexity* 2 (1): 44-52.
- Gerhart, J., and M. Kirschner. 2007. "The theory of facilitated variation." *Proceedings of the National Academy of Sciences* 104 (1): 8582-89.
- Godfrey-Smith, P. 2007. "Conditions for evolution by natural selection." *The Journal of Philosophy* 104 (10): 489-516.
- Goldberg, D. E. 1989. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley.
- Goldberg, D. E., and J. H. Holland. 1988. "Genetic algorithms and machine learning." *Machine Learning* 3 (2): 95-99.
- Gould, S. J. 2000. *Wonderful life: The Burgess shale and the nature of history*. Random House.
- Gould, S. J., and N. Eldredge. 1993. "Punctuated equilibrium comes of age." *Nature* 366 (6452): 223-27.
- Gould, S. J., and E. S. Vrba. 1982. "Exaptation-a missing term in the science of form." *Paleobiology* 8 (1): 4-15.
- Graham, L., and F. Oppacher. 2007a. "Speciation through exaptation." In *Artificial Life, 2007. ALIFE'07. IEEE Symposium on*, 433-39. IEEE.
- . 2007b. "A multiple-function toy model of exaptation in a genetic algorithm." In *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*, 4591-98. IEEE.
- Grant, B. S. 1999. "Fine tuning the peppered moth paradigm." *Evolution* 53 (3): 980-84.
- Gregory, T. R. 2008. "The evolution of complex organs." *Evolution: Education and Outreach* 1 (4): 358-89.
- Grogan, D. W. 1998. "Hyperthermophiles and the problem of DNA instability." *Molecular Microbiology* 28 (6): 1043-49.

List of References

- Gross, C. G. 1998. "Claude Bernard and the constancy of the internal environment." *The Neuroscientist* 4 (5): 380-85.
- Haldane, J. B. S. 1990. *The causes of evolution*. Princeton University Press.
- Hambley, A. R. 2008. *Electrical engineering: principles and applications*. Pearson Prentice Hall.
- Hammett, L. P. 1970. *Physical organic chemistry: reaction rates, equilibria, and mechanisms*. McGraw-Hill New York.
- Hastad, J. 1986. "Almost optimal lower bounds for small depth circuits." In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, 6-20. ACM.
- Haszprunar, G. 1988. "On the origin and evolution of major gastropod groups, with special reference to the streptoneura." *Journal of Molluscan Studies* 54 (4): 367-441.
- Hazen, R. M., A. Bekker, D. L. Bish, W. Bleeker, R. T. Downs, J. Farquhar, J. M. Ferry, E. S. Grew, A. H. Knoll, and D. Papineau. 2011. "Needs and opportunities in mineral evolution research." *American Mineralogist* 96 (7): 953-63.
- Heylighen, F. 1999. "The growth of structural and functional complexity during evolution." *The Evolution of Complexity*, 17-44. Kluwer, Dordrecht.
- Horton, H. R., L. A. Moran, K. G. Scrimgeour, M. D. Perry, and J. D. Rawn. 1996. *Principles of biochemistry*. Prentice Hall Upper.
- Houle, D. 1992. "Comparing evolvability and variability of quantitative traits." *Genetics* 130 (1): 195-204.
- Huxley, J. 1942. *Evolution: The Modern Synthesis*. Allen & Unwin.
- Jacob, F. 1977. "Evolution and tinkering." *Science* 196 (4295): 1161-66.
- Kashtan, N., and U. Alon. 2005. "Spontaneous evolution of modularity and network motifs." *Proceedings of the National Academy of Sciences* 102 (39): 13773-78.
- Kashtan, N., A. E. Mayo, T. Kalisky, and U. Alon. 2009. "An analytically solvable model for rapid evolution of modular structure." *PLoS Computational Biology* 5 (4): e1000355.
- Kashtan, N., E. Noor, and U. Alon. 2007. "Varying environments can speed up evolution." *Proceedings of the National Academy of Sciences* 104 (34): 13711-16.
- Katz, M. J. 1987. "Is evolution random." In *Development as an Evolutionary Process*, 285-315. Alan R. Liss, New York, NY.

- Kirschner, M., and J. Gerhart. 1998. "Evolvability." *Proceedings of the National Academy of Sciences* 95 (15): 8420-27.
- Knoll, A. H., and R. K. Bambach. 2000. "Directionality in the history of life: diffusion from the left wall or repeated scaling of the right?" *Paleobiology* 26 (4): 1-14.
- Kolmogorov, A. N. 1965. "Three approaches to the quantitative definition of information'." *Problems of Information Transmission* 1 (1): 1-7.
- Koonin, E. V. 2011. "Are there laws of genome evolution?" *PLoS Computational Biology* 7 (8): e1002173.
- Kornberg, A., and T. A. Baker. 1992. *DNA replication*. Vol. 5. WH Freeman New York.
- Laland, K. N., J. Odling-Smee, and M. W. Feldman. 2000. "Niche construction, biological evolution, and cultural change." *Behavioral and Brain Sciences* 23 (01): 131-46.
- Laland, K. N., J. Odling-Smee, and S. F. Gilbert. 2008. "EvoDevo and Niche Construction: building bridges." *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 310 (7): 549-66.
- Lane, N.. 2002. *Oxygen: The molecule that made the world*. Oxford University Press.
- . 2010. *Life Ascending: The ten great inventions of evolution*. Profile books.
- Langton, C. G. 1984. "Self-reproduction in cellular automata." *Physica D: Nonlinear Phenomena* 10 (1): 135-44.
- Laubichler, M. D. 2007. "Tinkering: a conceptual and historical evaluation." In *Tinkering: The Microevolution of Development*, edited by Gregory Bock and Jamie Goode, 20-34. John Wiley & Sons, Ltd.
- Lavialle, C., G. Cornelis, A. Dupressoir, C. Esnault, O. Heidmann, C. Vernochet, and T. Heidmann. 2013. "Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 368 (1626).
- Lenski, R. E. 2004. "Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*." *Plant Breeding Reviews* 24 (2): 225-266.
- Lenski, R. E., J. E. Barrick, and C. Ofria. 2006. "Balancing robustness and evolvability." *PLoS Biology* 4 (12): e428.

List of References

- Lenski, R. E., C. Ofria, R. T. Pennock, and C. Adami. 2003. "The evolutionary origin of complex features." *Nature* 423 (6936): 139-44.
- Lilius, J. 1998. "Efficient state space search for time petri nets." *Electronic Notes in Theoretical Computer Science* 18: 113-33.
- Lindahl, T. 1993. "Instability and decay of the primary structure of DNA." *Nature* 362 (6422): 709-15.
- Lloyd, S., and H. Pagels. 1988. "Complexity as thermodynamic depth." *Annals of Physics* 188 (1): 186-213.
- Lund, H. H., and D. Parisi. 1995. "Preadaptation in populations of neural networks evolving in a changing environment." *Artificial Life* 2 (2): 179-97.
- Macallum, A. B. 1926. "The paleochemistry of the body fluids and tissues." *Physiological Reviews* 6 (2): 316-57.
- Manahan, S. E. 2004. *Environmental chemistry, eighth edition*. CRC Press.
- Marmur, J., and P. Doty. 1962. "Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature." *Journal of Molecular Biology* 5 (1): 109-18.
- Maynard-Smith, J. 1969. "The status of Neo-Darwinism." In *Towards a Theoretical Biology* 2, edited by C. H. Waddington, 82-89.
- Maynard-Smith, J., and E. Szathmary. 1997. *The Major Transitions in Evolution*. Oxford University Press.
- Maynard-Smith, J. 1972. "Time in the evolutionary process." In J. T. Fraser, F. Haber, and G. Muller (Eds.) *The Study of Time*, 200-205. Springer Berlin Heidelberg.
- McBride, M. B. 1994. *Environmental chemistry of soils*. Oxford University Press.
- McLennan, D. A. 2008. "The concept of co-option: Why evolution often looks miraculous." *Evolution: Education and Outreach* 1 (3): 247-58.
- McMillan, K. L., and D. K. Probst. 1995. "A technique of state space search based on unfolding." *Formal Methods in System Design* 6 (1): 45-65.
- McMullin, B. 2000. "John von neumann and the evolutionary growth of complexity: looking backward, looking forward" *Artificial Life* 6 (4): 347-61.
- McShea, D. W. 1991. "Complexity and evolution: What everybody knows." *Biology and Philosophy* 6 (3): 303-24.
- . 1994. "Mechanisms of large-scale evolutionary trends." *Evolution* 48: 1747-63.

- . 1996. "Metazoan complexity and evolution: is there a trend?" *Evolution* 50: 477-92.
- . 2009. "The evolution of complexity without natural selection, a possible large-scale trend of the fourth kind." *Paleobiology* 31: 146-156
- McShea, D. W., and Robert N. Brandon. 2010. *Biology's first law: The tendency for diversity and complexity to increase in evolutionary systems*. University of Chicago Press.
- Miglino, O., S. Nolfi, and D. Parisi. 1996. "Discontinuity in evolution: how different levels of organization imply preadaptation." In *Artificial Life 2 (Santa Fe Institute Studies In The Sciences Of Complexity-Proceedings)*: 399-416. Addison-Wesley.
- Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. "Network motifs: Simple building blocks of complex networks." *Science Signaling* 298 (5594): 824.
- Mitchell, M. 2009. *Complexity: A guided tour*. Oxford University Press, USA.
- Mouret, J. B., and Stéphane Doncieux. 2009. "Evolving modular neural-networks through exaptation." In S. Yang and X. Yao (Eds.) *Evolutionary Computation, 2009. CEC'09. IEEE Congress on*, 1570-77.
- Mulkidjanian, A. Y., and M. Y Galperin. 2007. "Physico-chemical and evolutionary constraints for the formation and selection of first biopolymers: towards the consensus paradigm of the abiogenic origin of life." *Chemistry & Biodiversity* 4 (9): 2003-15.
- Mulkidjanian, A. Y., A. Yu. D. V. Dibrova, M. Y. Galperin, and E. V. Koonin. 2012. "Origin of first cells at terrestrial, anoxic geothermal fields." *Proceedings of the National Academy of Sciences* 109 (14): E821-E830.
- Müller, G. B. 2007. "Evo-devo: Extending the evolutionary synthesis." *Nature Reviews Genetics* 8 (12): 943-49.
- Newell, N. D. 1949. "Phyletic size increase, an important trend illustrated by fossil invertebrates." *Evolution* 3 (2): 103-24.
- Newman, M. 2009. *Networks: An introduction*. Oxford University Press.
- Odling-Smeel, F., K. N. Laland, and M. W. Feldman. 2013. *Niche construction: The neglected process in evolution*. Princeton University Press.
- Oehlenschläger, F., and M. Eigen. 1997. "30 years later--a new approach to Sol Spiegelman's and Leslie Orgel's in vitro evolutionary studies. Dedicated to Leslie Orgel on the occasion of his 70th birthday." *Origins of Life and*

List of References

- Evolution of the Biosphere: The Journal of the International Society for the Study of the Origin of Life* 27 (5-6): 437-57.
- Orr, H. A. 2005. "The genetic theory of adaptation: A brief history." *Nature Reviews Genetics* 6 (2): 119-27.
- Palmer, A. R. 1979. "Fish predation and the evolution of gastropod shell sculpture: Experimental and geographic evidence." *Evolution* 33: 697-713.
- Palmes, P. P., and S. Usui. 2005. "Robustness, evolvability, and optimality of evolutionary neural networks." *BioSystems* 82 (2): 168-88.
- Papachristodoulou, A., and B. Recht. 2007. "Determining interconnections in chemical reaction networks." In R. Judd and L. G. Bushnell (eds) *Proceedings of the American Control Conference, 2007. ACC'07*, 4872-77.
- Papadimitriou, C. H. 2003. *Computational complexity*. Wiley.
- Parter, M., N. Kashtan, and U. Alon. 2007. "Environmental variability and modularity of bacterial metabolic networks." *BMC Evolutionary Biology* 7 (1): 169.
- . 2008. "Facilitated Variation: How evolution learns from past environments to generalize to new environments." *PLoS Computational Biology* 4 (11): e1000206.
- Pavlicev, M., J. M. Cheverud, and G. P. Wagner. 2011. "Evolution of adaptive phenotypic variation patterns by direct selection for evolvability." *Proceedings of the Royal Society of London B: Biological Sciences* 278 (1713): 1903-12.
- Pigliucci, M. 2007. "Do we need an extended evolutionary synthesis?" *Evolution* 61 (12): 2743-49.
- . 2008. "Is evolvability evolvable?" *Nature Reviews Genetics* 9 (1): 75-82.
- Ray, T. S. 1992. "Evolution, ecology and optimization of digital organisms." Technical Report Working paper 92-08-042, Santa Fe Institute.
- Ray, T. S. 1993. "An evolutionary approach to synthetic biology: Zen and the art of creating life." *Artificial Life* 1 (2): 179-209.
- Razborov, A. A. 1990. "Lower bounds for monotone complexity of boolean functions." *American Mathematical Society Translations* 147: 75-84.
- Reaves, M. L., S. Sinha, J. D. Rabinowitz, L. Kruglyak, and R. J. Redfield. 2012. "Absence of detectable arsenate in dna from arsenate-grown GFAJ-1 cells." *Science* 337 (6093): 470-73.

- Rensch, B. 1966. *Evolution above the species level*. Wiley.
- Ridley, M. 2009. *Evolution*. Wiley.
- Riedl, R., and R. P. S. Jefferies. 1978. *Order in living organisms: A systems analysis of evolution*. Wiley.
- Russell, S. J., P. Norvig, J. F. Canny, J. M. Malik, and D. D. Edwards. 1995. *Artificial Intelligence: A modern approach*. Prentice Hall Englewood Cliffs.
- Saunders, P. T., and M. W. Ho. 1976. "On the increase in complexity in evolution." *Journal of Theoretical Biology* 63 (2): 375-84.
- . 1981. "On the increase in complexity in evolution II. The relativity of complexity and the principle of minimum increase." *Journal of Theoretical Biology* 90 (4): 515-30.
- Sayama, H. 1999. "A new structurally dissolvable self-reproducing loop evolving in a simple cellular automata space." *Artificial Life* 5 (4): 343-65.
- Schwenk, K., and G. P. Wagner. 2001. "Function and the evolution of phenotypic stability: Connecting pattern to process." *American Zoologist* 41 (3): 552-63.
- . 2004. "The relativism of constraints on phenotypic evolution." In *Phenotypic Integration: Studying the ecology and evolution of complex phenotypes*, 390-408. Oxford University Press.
- Shannon, C. E. 1948. "A mathematical theory of communication". *Bell Systems Technical Journal* 27 (3): 379-423.
- Simon, H. A. 1962. "The architecture of complexity." *Proceedings of the American Philosophical Society* 106 (6): 467-82.
- . 1969. "The sciences of the artificial." MIT Press.
- Sims, K. 1994a. "Evolving virtual creatures." In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, 15-22.
- . 1994b. "Evolving 3D morphology and behavior by competition." *Artificial Life* 1 (4): 353-72.
- Smolensky, R.. 1987. "Algebraic methods in the theory of lower bounds for boolean circuit complexity." In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, 77-82. ACM.
- Solomon, S. 2007. *Climate change 2007 - The physical science basis: Working group I contribution to the fourth assessment report of the IPCC*. Vol. 4. Cambridge University Press.

List of References

- Spiegelman, S., I. Haruna, I. B. Holland, G. Beaudreau, and D. Mills. 1965. "The synthesis of a self-propagating and infectious nucleic acid with a purified enzyme." *Proceedings of the National Academy of Sciences* 54 (3): 919-27.
- Stebbins, G. L., and F. J. Ayala. 1981. "Is a new evolutionary synthesis necessary?" *Science* 213 (4511): 967-71.
- Stebbins, G. L. 1969. *The basis of progressive evolution*. University of North Carolina Press.
- Steiner, C. F. 2012. "Environmental noise, genetic diversity and the evolution of evolvability and robustness in model gene networks." *PLoS One* 7 (12): e52204.
- Suzuki, H., N. Ono, and K. Yuta. 2003. "Several necessary conditions for the evolution of complex forms of life in an artificial environment." *Artificial Life* 9 (2): 153-74.
- Tamura, K., S. Yonei-Tamura, T. Yano, H. Yokoyama, and H. Ide. 2008. "The autopod: Its formation during limb development." *Development, Growth & Differentiation* 50 (1): 177-187.
- Thornhill, R. H., and D. W. Ussery. 2000. "A classification of possible routes of darwinian evolution." *Journal of Theoretical Biology* 203 (2): 111-16.
- True, J. R., and S. B. Carroll. 2002. "Gene co-option in physiological and morphological evolution." *Annual Review of Cell and Developmental Biology* 18 (1): 53-80.
- Uchman, A. 2003. "Trends in diversity, frequency and complexity of Graphoglyptid trace fossils: evolutionary and palaeoenvironmental aspects." *Palaeogeography, Palaeoclimatology, Palaeoecology* 192 (1): 123-42.
- Variano, E. A., J. H. McCoy, and H. Lipson. 2004. "Networks, dynamics, and modularity." *Physical Review Letters* 92 (18): 188701.
- Vermeij, G. J., D. E. Schindel, and E. Zipser. 1981. "Predation through geological time: Evidence from Gastropod shell repair." *Science* 214 (4524): 1024-26.
- Vogel, A. I. 1974. *A textbook of practical organic chemistry including qualitative organic analysis*. Longmans, Green and Company.
- Wagner, A. 2005. *Robustness and evolvability in living systems*. Princeton University Press Princeton.

- Wagner, G. P., and L. Altenberg. 1996. "Perspective: Complex adaptations and the evolution of evolvability." *Evolution* 50 (3): 967-76.
- Wagner, G. P., and M. D. Laubichler. 2004. "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* 302 (1): 92-102.
- Wagner, G. P., J. Mezey, and R. Cakabretta. 2001. "Natural selection and the origin of modules." In *Modularity: Understanding the Development and Evolution of Complex Natural Systems*. MIT Press.
- Wagner, G. P., M. Pavlicev, and J. M. Cheverud. 2007. "The road to modularity." *Nature Reviews Genetics* 8 (12): 921-31.
- Wagner, G. P., and K. Schwenk. 2000. "Evolutionarily stable configurations: Functional integration and the evolution of phenotypic stability." In *Evolutionary Biology*, 155-217. Springer.
- Wald, G. 1964. "The origins of life." *Proceedings of the National Academy of Sciences* 52 (2): 595.
- Watson, R. A. 2006. *Compositional evolution: The impact of sex, symbiosis and modularity on the gradualist framework of evolution*. MIT Press.
- . 2012. "Is evolution by natural selection the algorithm of biological evolution?" In C. Adami, D. M Bryson, C. Ofria and R. T. Pennock (Eds.) *Artificial Life 13: Proceedings of the Thirteenth International Conference on the Simulation and Synthesis of Living Systems*, 121-28. MIT Press.
- Watson, R. A., and J. B. Pollack. 2005. "Modular interdependency in complex dynamical systems." *Artificial Life* 11 (4): 445-57.
- Watson, R. A., G. P. Wagner, M. Pavlicev, D. M. Weinreich, and R. Mills. 2014. "The evolution of phenotypic correlations and 'developmental memory.'" *Evolution* 68 (4): 1124-38.
- Watson, R. A., C. L. Buckley, R. Mills, and A. Davies. 2010. "Associative memory in gene regulation networks." In H. Fellermann, M. Dörr, M. M. Hanczy, L. L. Laursen, S. Maurer, D. Merkle, P. A. Monnard, K. Støy and S. Rasmussen (Eds.) *Artificial Life XII: Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems*, 659-666. MIT Press.
- Watson, R. A., R. Mills, C. L. Buckley, A. Penn, A. Davies, J. Noble, and S. Bullock. 2010. "Adaptation without natural selection." In H. Fellermann, M. Dörr, M. M. Hanczy, L. L. Laursen, S. Maurer, D. Merkle, P. A.

List of References

- Monnard, K. Støy and S. Rasmussen (Eds.) *Artificial Life XII: Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems*, 80–81. MIT Press.
- Whitacre, J., and A. Bender. 2010. "Degeneracy: A design principle for achieving robustness and evolvability." *Journal of Theoretical Biology* 263 (1): 143–53.
- Wimsatt, W. C. 1986. "Developmental constraints, generative entrenchment, and the innate-acquired distinction." In *Integrating Scientific Disciplines*, 185–208. Springer.
- . 2001. "Generative Entrenchment and the developmental systems approach to evolutionary processes." In S. Oyama, R. D. Gray and P. E. Griffiths (Eds.) *Cycles of Contingency: Developmental Systems and Evolution*, 219–37. MIT Press.
- Wolfe-Simon, F., J. S. Blum, T. R. Kulp, G. W. Gordon, S. E. Hoeft, J. Pett-Ridge, J. F. Stolz, S. M. Webb, P. K. Weber, and P. C. W. Davies. 2011. "A bacterium that can grow by using arsenic instead of phosphorus." *Science* 332 (6034): 1163–66.
- Wright, S. 1931. "Evolution in Mendelian populations." *Genetics* 16 (2): 97.
- Xu, P., I. Woo, H. Her, D. R. Beier, and R. L. Maas. 1997. "Mouse eya homologues of the Drosophila Eyes Absent gene require Pax6 for expression in lens and nasal placode." *Development* 124 (1): 219–31.
- Yaeger, L. 1994. "Polyworld: Life in a new context". In C. G. Langton (Ed.) *Artificial Life 3: Proceedings of the Third International Conference on the Synthesis and Simulation of Living Systems*, 263–264. Westview Press.