Preface

Computer science and biology have a lot to offer each other. My background is in computer science, and like a growing number of researchers I work at the interface of computer science and biology. There is considerable interest at this interface both in using biological systems as inspiration to develop new computational methods (*bio-inspired computation*) and in using computational modeling techniques to assist in the understanding of biological systems (*computational biology*). Although bio-inspired computation does not always assist in the understanding of the biology that inspired it, and computational biology models do not always require new insights on the part of the computer scientist, in some fortuitous cases, work in these fields helps progress both disciplines together. This kind of synergistic result occurs when the biological system or principle being studied is fundamentally an algorithmic or computational process. Some biological processes stand out as obvious candidates for such research. For example, in the science of cognition, the development of artificial neural networks has provided new approaches to solving computational problems and also changed our view of the kind of processes that might go on in the cortex. Another candidate is the science of evolution; but despite notable successes in the use of evolutionary computation techniques for problem solving and design, the underlying principles of evolution by natural selection as most of us understand them today are not significantly different in an *algorithmic* sense, that is, in the sense of a formal step-by-step procedure, from those that Darwin laid out nearly a century and a half ago.

Evolution by natural selection occurs whenever individuals show variation in reproductive success that can be inherited (heritable variation). Variants that reproduce more effectively increase in number, and those that do not, don't. Accordingly, variations of the existing individuals that are *fitter*, in this sense, come to replace those that are less fit. In this manner, although the variations are random, selection acting on them will produce directed change. The underlying algorithmic principle here is *random mutation hill climbing*, the linear, or sequential, accumulation of random changes, each one a fitness improvement over the last.

Darwin believed that each random change must be small and that a progression of "successive slight variations" was required for any adaptation (1859). After both the "Modern Synthesis" of the 1930s (Fisher 1930), when Mendelian genetics (Mendel 1866) was integrated with Darwinian natural selection (Darwin 1859), and the discovery of the molecular structure of DNA in the 1950s (Watson and Crick 1953), it was clear that natural selection could act on whatever heritable variation occurred and that this may come about by spontaneous point mutation introduced by DNA replication errors, or sexual recombination, or any mechanism of genetic variation. These genetic changes may be any size in principle but the process is still fundamentally the linear accumulation of random genetic changes. This is what I shall refer to as the *gradualist framework of evolution*. It includes, in principle, random genetic changes of any size, but given that large random genetic changes seem much more likely than small changes to destroy the proper functioning of a well-adapted organism, it is generally assumed that most adaptive changes will be small ones. Nonetheless, fitness improvements from large genetic changes are not categorically excluded in this framework. It is the linear successive improvement by random variation that defines the framework, not the size of the genetic variations involved. It might be useful to mention straight away that *saltationism* (Goldschmidt 1940), *punctuated equilibria* (Gould and Eldredge 1977), and *neutral theory* (Kimura 1983) are, as I will discuss, all included in this definition of the gradualist framework because they are not different algorithmically—although they differ in opinion about the expected size, rate, and fitness effect of random variations respectively, they all agree that evolution by natural selection is based on the sequential accumulation of random variations. Indeed, the rejection of saltationism results from the implausibility of large random beneficial changes under such an algorithmic process.

Some mechanisms of genetic variation do produce large genetic changes and sometimes these do produce something adaptive. Sexual hybridization of different species sometimes creates successful new species (Rieseberg et al. 2003); lateral or horizontal gene transfer (Mazodier and Davies 1991), widespread in single-celled organisms (Doolittle 2000), has resulted in the rapid adaptation of antibiotic resistance, for example (Ochman, Lawrence and Groisman 2000); and one of the most extraordinary events in evolutionary history was the origin of the intracellular organelles, such as mitochondria and chloroplasts, through the union of genetic material from symbiotic bacteria (Margulis 1970). Are these mechanisms accommodated by the gradualist framework of evolution? Is linear sequential improvement sufficient to describe these processes properly? I argue that they are not.

Notably, mechanisms like sex, lateral gene transfer, and symbiogenesis (the genesis of new species through symbiosis) are not merely a different source of random genetic changes that happen to be large and occasionally adaptive (increasing fitness). These mechanisms involve not the *linear* accumulation of random genetic changes but rather the coming together of genetic material that has been preadapted in different lineages—different individuals, subpopulations, or species. Accordingly, the genetic variations that they produce are not "undirected" random genetic changes—but rather "directed" by prior selection. The more important question is: does it matter whether adaptive genetic variations were selected for in parallel lineages and subsequently brought together, rather than selected serially in a single lineage? The finding of this book is that yes, it does matter. These mechanisms do not introduce anything magical or teleological to the process of evolution by natural selection, but they change the underlying algorithmic principles of evolution, and they cannot be accommodated by the gradualist framework.

In this book I will show that these mechanisms enable *compositional evolution* and that this is algorithmically distinct from the gradualist framework of evolution. One particularly important consequence of this distinction can be seen in how it impacts our understanding of which types of systems are evolvable and which are not. This is illustrated by examining the class of complex systems that compositional evolution can produce but gradual evolution cannot. I show that, in principle, certain kinds of complex adaptations that are pathologically difficult for gradual evolution can be produced easily by compositional evolution. I define an example complex system that has difficult dependencies between different parts of the system that prevent linear incremental improvements in fitness. This system is not only *unevolvable* (not evolvable) under the strict gradualist assumption of small genetic changes, but also unevolvable through the linear accumulation of random genetic changes *of any size*. Nonetheless, a compositional process can exploit the underlying modular structure and evolve systems of this type easily.

The differences between compositional evolution and gradual evolution derive from the fact that they have different underlying algorithmic principles. The algorithmic principle of gradual evolution is simply *hill climbing*, whereas the algorithmic principle underlying compositional evolution is *divide-and-conquer problem decomposition*. The method of solving a problem by decomposing it into more manageable subproblems is a familiar and intuitive concept in design and engineering. But whereas this is usually assumed to require top-down knowledge of how to decompose a problem, in this book I show that it can be applied "bottom up" in an evolutionary process. When this algorithmic distinction is understood, we see that it is no longer appropriate to try to force mechanisms like sex, lateral gene transfer, and symbiogenesis into the linear paradigm of the gradualist framework. Instead we must expand the framework of evolution by natural selection to include a greater range of algorithmic possibilities.

Because evolution by natural selection is fundamentally an algorithmic concept, work in extracting inspiration from biology for computational methods has the potential to feed back into the biology that inspired it. The goals of this book are therefore explicitly interdisciplinary: to elucidate the algorithmic principles underlying evolutionary mechanisms to answer existing questions in computer science, and to provide food for thought about our understanding of the biological processes of evolution.

It turns out that arguments about whether events such as symbiogenesis require us to modify our conception of how evolution works have been going on for nearly a century (Famintsyn 1907; Merezhkovsky 1909; Wallin 1927; Margulis 1970, 1993b). More recently, symbiogenesis is included as one of several fundamental transitions in evolution where entities that were formerly reproductively independent became reproductively dependent (Maynard Smith and Szathmary 1995), forming a new entity at a higher level of organization. But thus far the fact of such events in nature has not been permitted to significantly influence the fundamental ideas about how evolution works. In this book I use the tools of complexity theory and algorithmics to support the claim that these mechanisms should expand our ideas about how evolution works in a fundamental sense.

In evolutionary computation there has been a similarly heated debate about whether the *genetic algorithm*, a computational problem-solving technique based loosely on evolution by natural selection, offers any fundamentally interesting problem-solving capability or whether it is just an unnecessarily complicated form of mutation hill climbing. In particular, the potential role of sexual recombination, or crossover, has been central in this debate. The claims within computer science about the ability of crossover to put together well-adapted subsets of genetic material to form well-adapted individuals not only provide a question worth answering in its own right, but moreover, this issue is algorithmically analogous to the potential of symbiogenesis to put together well-adapted simple entities to form well-adapted entities at a new level of organization. Some of the results I develop in this work build on ideas that were familiar in evolutionary computation but previously not well supported formally; in providing clear resolutions to these questions I also provide a computational framework for understanding sex and symbiogenesis as part of a spectrum of compositional mechanisms.

Following these motives, this book offers: (1) a model of sexual recombination using a genetic algorithm that shows that it is algorithmically distinct from mutation hill-climbing methods; (2) a model of symbiogenesis and the major evolutionary transitions that shows that such processes are also a fundamentally different class of adaptive process from the gradualist evolutionary framework of linear incremental improvement; and (3) a framework for understanding both symbiogenesis and sexual recombination as instances of a general class of mechanisms that I have termed compositional evolution: evolutionary processes involving the combination of systems or subsystems of semi-independently preadapted genetic material. This is contrasted with the gradualist framework of evolution that depends on the linear or sequential accumulation of random modifications.

Readers in several related fields may therefore find this book useful. For researchers in evolutionary computation there are results addressing the benefit of sexual recombination or crossover, the "building-block hypothesis" and which problems are difficult or easy for the genetic algorithm, "cooperative coevolution," the maintenance of diversity in evolving populations, and "linkage learning." For researchers in evolutionary biology the models presented provide tools to think about gradualism in evolution, the benefit of recombination, the evolution of cooperation, serial endosymbiosis theory, and the major evolutionary transitions. For readers interested in combinatorial optimization and machine learning in general, the algorithms developed here provide a heuristic bottom-up divide-and-conquer method that enables automatic problem decomposition and automatic module acquisition in an appropriate problem domain. And researchers in complex systems are provided with models addressing modularity and decomposability in dynamical systems, especially hierarchically modular systems.

The book has three parts. The first part (chapters 1 through 3) provides an overview of the arguments set forth in this book, reviews the gradualist framework of evolution and its impact on our understanding of evolvability, and reviews the compositional mechanisms exhibited in evolutionary biology and the analogous principles used in evolutionary computation. This part of the book provides the background in evolutionary biology and evolutionary computation necessary to see how the research questions in both fields are similar and how they complement one another.

The second part (chapters 4 through 8) provides the computational models that illustrate my argument. Chapter 4 provides a definition of a modular complex system designed to exemplify the different adaptive capabilities of gradual and compositional mechanisms. The following three chapters then examine the evolution of this system under different evolutionary scenarios. Those scenarios use three different variation mechanisms. Chapter 5 discusses spontaneous point mutation—which provides only gradual evolution. Chapter 6 discusses sexual recombination—which can sometimes provide compositional evolution depending on properties of population diversity and genetic architecture. Chapter 7 presents a mechanism of symbiotic encapsulation—which allows less restricted compositional evolution. The last chapter of this part, chapter 8, provides a formal analysis that shows how these compositional mechanisms scale with the size of the system being evolved.

In the third part of the book, chapter 9 discusses how compositional evolution enables the scaling up of evolutionary processes, and the conceptual components behind the compositional models. And chapter 10 discusses the impact it has on our understanding of natural evolution and, similarly, the utility of evolutionary computation methods for problem solving and design.

Throughout the book I have attempted to draw the ideas in evolutionary computation and evolutionary biology closer together to help foster continued exchange between the disciplines. I find it hard to imagine working in one field without being excited by the other, and I hope that some of my enthusiasm for the two rubs off. There is much exciting work to be done to bring evolutionary computation and evolutionary biology closer together so that the full benefits of insights and results in each can inform and stimulate the other. A meeting of evolutionary computation and evolutionary biology, like compositional events in evolutionary history, has the potential to produce many synergistic effects.