1 Introduction

If it could be demonstrated that any complex organ existed which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down.
—Charles Darwin, 1859

Conventionally, evolution by natural selection is almost inseparable from the notion of accumulating successive slight variations. Notions of *gradualism* are deeply ingrained in evolutionary thought and strongly influence our ideas about what kinds of systems are evolvable and which are not. The very idea of large adaptive genetic changes is considered *unevolutionary*. Indeed, attacking the plausibility of a linear succession of incrementally improving protosystems is a common form of criticism of evolutionary theory from Paley (1802) to Behe (1996), while asserting the plausibility of such a succession is a common form of defense (Dawkins 1996).

In natural evolution there are a number of mechanisms that provide exceptional forms of genetic change. For example, the eukaryotic cell (from which all plants and animals are descended) has organelles (such as mitochondria and chloroplasts) that were formed from the genetic integration of previously separate species of bacteria. In these events, groups of organisms, formerly living in close symbiotic relationship, were thus transformed into a new individual at a new level of organization (Margulis 1970). It was suggested as early as 1909 (by Merezhkovsky) that such symbiotic mechanisms, by bringing together the genetic material of two genetically distinct organisms, provide an alternative to gradual evolutionary change. Similarly, on a less dramatic but more frequent scale, sexual recombination has the potential to bring together sets of coadapted genes from genetically distinct parents providing a large genetic change that would be extremely unlikely from spontaneous point mutations applied to either parent alone.

There has been disagreement about what impact such mechanisms should have on our understanding of evolutionary processes. Some suggest that the genetic integration of symbionts cannot be very important in understanding evolutionary processes in general simply because such events are rare. Also, sex is usually restricted to populations that are genetically similar so the opportunity for radical genetic change seems low. Moreover, for both sexual and symbiotic integration, it still seems likely that any large genetic change would more likely be disastrous than beneficial. And finally, even if it were the case that a particular adaptation resulted from genetic integration of different individuals, this might not be a substantive challenge to gradualism since the genetic material provided by each of the component individuals was (presumably) evolved gradually.

The biological fact of the genetic mechanisms is not disputed, and neither is their role in enabling significant innovation such as the eukaryote cell. It is also clear that large genetic changes could be selected for by natural selection if they were beneficial, and mechanisms such as these provide large genetic changes that have evidently, at least sometimes, been beneficial. Accordingly, it might seem that these mechanisms do not require any fundamental change to the gradualist framework of evolution by natural selection. However, a change is required.

In these mechanisms, genetic material is adapted in parallel in different individuals, subpopulations, or species, and subsequently brought together. Thus the genetic changes produced by such events are quite unlike the random genetic changes of, for example, those produced by spontaneous genetic mutations. This parallel adaptation enables these mechanisms to evolve systems that cannot be evolved through the linear accumulation of random genetic changes. Such a process does not introduce anything teleological into evolution, and it still sits squarely within the framework of evolution by natural selection. But it causes us to separate evolution by natural selection from the assumption of gradual change—not merely from the assumption of strict gradualism, involving only "successive slight variations," but from the linear accumulation of random changes *of any size*.

This book uses *evolutionary computation* models to aid in the understanding of these processes. Since the 1970s, considerable interest has developed within computer science in computational optimization methods based loosely on Darwinian natural selection. Work in *evolutionary algorithms*, EAs, is a highly active and rapidly growing field—the most well-known variety being *genetic algorithms*, GAs. Within this field there is considerable interest in the possibility of methods that encapsulate primitive functional units into higher-order components that can be reused in subsequent evolution, which has obvious parallels with these biological mechanisms. There is also significant controversy over the supposed utility of sexual recombination, or *crossover*, in such algorithms; whether it affords any special problem-solving utility or whether spontaneous point mutation alone is just as good. In this book, by providing an answer to these computational questions we also gain an increased understanding of the adaptive capacity of sex and other compositional mechanisms in nature. A computer science approach is useful, not just for providing a modeling tool, but moreover in providing theoretical tools. Algorithmics and complexity theory provide the means to properly formalize our questions and identify differences in adaptive processes in a fundamental sense. In particular, this enables us to pose and answer a formal version of the following question: What kind of complex systems (if any) can be evolved with compositional mechanisms that cannot be evolved through gradual evolution?

The answer to this question depends on *modularity*.[1] I identify a particular form of modular complexity, where subsystems can be evolved semi-independently despite important intermodule dependencies. Systems with this kind of modularity cannot be evolved by gradual evolution because resolving the interdependencies between modules may require making specific large genetic changes. In fact, this class of system typifies characteristics that are pathological for gradual evolution: the corresponding fitness landscape is highly rugged, has many local optima creating broad fitness saddles, and includes "irreducibly complex" adaptations that cannot be reached by a linear succession of randomly modified protosystems. However, an evolving process with compositional mechanisms does not necessarily have the same restrictions. With compositional mechanisms that enable preadapted genetic material from different individuals to be combined, genetic material from another individual can provide specific large genetic changes, enabling continued adaptation. I provide results of simulations to show that this class of modular system is easily evolvable under compositional evolution and not evolvable under gradual evolution. Thus, certain kinds of complex systems, considered unevolvable under the gradualist framework of evolution, are easily evolvable under compositional evolution.

The fact that compositional evolution can provide evolutionary adaptation that is not possible in the gradualist framework causes us to rethink our notions of evolvability (how easy or difficult it is for systems to evolve) and reassess the potential of sex and symbiosis in evolutionary innovation. The simulations and analytic results show us that there are fundamental differences in the adaptive capacities of these mechanisms, and help us to understand the conditions under which compositional mechanisms provide an adaptive advantage.

This book has two main objectives: first, to present and develop the above argument that sex, symbiosis, and other compositional mechanisms move beyond the normal "linear accumulation of random changes" dogma in evolutionary theory, giving new results and theoretical work to support this view; and second and more generally, to foster interdisciplinary exchange between evolutionary biology and evolutionary computation, showing how some of the deepest questions in each of these disciplines can be approached with the tools and insights of the other.

1.1 Gradual and Compositional Evolution

I propose the term *compositional evolution* to refer to evolutionary processes involving the combination of systems or subsystems of semi-independently preadapted genetic material. A family of such mechanisms exists including sex, hybridization, lateral gene transfer, allopolyploidy, and mechanisms of "symbiotic encapsulation" such as endosymbiosis, the genetic integration of symbionts (symbiogenesis), or other mechanisms that encapsulate a group of simple entities into a complex entity at

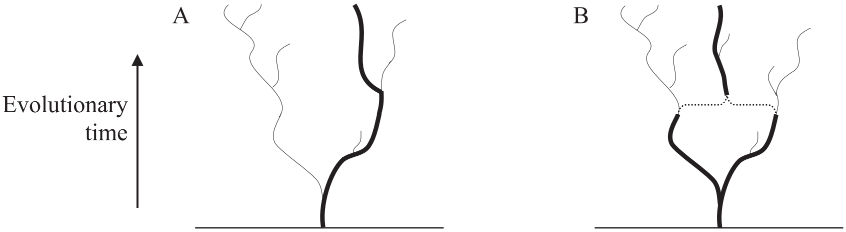


Figure 1.1
Gradual and compositional evolutionary processes. (a) The linear accumulation of random changes under the normal framework of gradual evolutionary adaptation produces the familiar branching ancestral tree. (b) In compositional evolution, in addition to the normal branching process of gradual evolution, there is also the occasional creation of a new entity from the combination of preadapted genetic material from different individuals. This creates a new entity that may be genetically distant from both donor individuals.

a higher level of organization, as exhibited in several of the major transitions in evolution.

In contrast, the familiar gradualist framework of evolution refers to mechanisms that sequentially accumulate random variations in genetic material, that is, the new genetic material introduced by such changes has not been preadapted elsewhere as a set. Thus gradual evolution is driven predominantly by small modifications, that is, "successive slight modifications" (Darwin 1859), though not exclusively so, in principle. This forms the basis of the normal understanding of evolutionary change. Chapter 2 details the background of the gradualist framework of evolution and the established notions of evolutionary difficulty that, in fact, depend on this framework.

Figure 1.1 shows a schematic comparison of ancestral relationships in gradual evolution and compositional evolution. Gradual evolution produces the familiar ancestral "tree" phylogeny (Haeckel 1876). Note that for any extant entity produced by gradual evolution there is a single linear path (shown in bold) linking it to the root of the tree via a succession of (small) random changes. In contrast, individuals descended from a compositional event cannot be traced to the root of the tree via a single linear path of gradual changes. Instead it is correct to respect the true multiple-inheritance origin of this new lineage producing a disconnected ancestral graph where the links to the root of new subtrees (shown by dashed lines) are of a different kind to the rest of the graph. Specifically, not only are they two-to-one edges (connecting two nodes to a new third) but more significantly they do not indicate a succession of gradually changing intermediates.

It is useful to compare this depiction with other nontree diagrams of ancestral relationships that have been offered previously for various events at different scales from horizontal gene transfer to symbiogenesis (e.g., Doolittle 2000; Rivera and

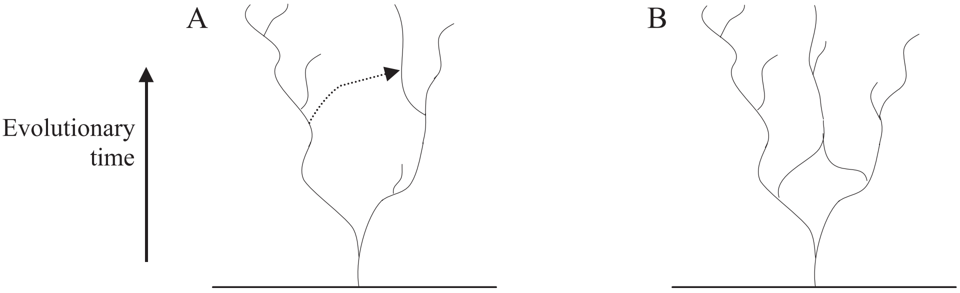


Figure 1.2
Other depictions of compositional events. (a) An ancestral tree with "exceptions." (b) A nontree graph.

Lake 2004; Margulis 1970). Some of these diagrams depict all entities as fitting within a tree and then add "exceptional" cross-links that indicate the exchange of genetic material from one branch across to another (e.g., Doolittle 2000; Kim and Salisbury 2001)—see figure 1.2a. This depiction obscures the fact that the entities thus created are not a small modification of either ancestral donor even though they may be (very much) more similar to one than the other. Others depict the connections that bring lineages together as smooth transitions (of the same type that draw branches apart in differentiation) (e.g., Margulis 1970; Rivera and Lake 2004)—see figure 1.2b. However, if the horizontal dimension of this diagram is supposed to represent genetic (or phenotypic) diversity then this depiction is misleading since it is not the case that the donors of genetic material in a compositional event evolve toward one another gradually becoming genetically or phenotypically similar until they become indistinct and form a single lineage. Even Doolittle, who proposes "uprooting the tree of life" in light of evidence for horizontal gene transfer, uses a depiction of ancestral relationships near the base of the graph using a uniform style of link in the form of figure 1.2b (despite challenging the adherence to a single common ancestor at the root of the tree). Even though the unusual mechanisms of genetic variation are included in these figures, these seemingly subtle issues in fact reveal a strong adherence to trees with linear branching (Kim and Salisbury 2001) and thus gradualist assumptions. Of course, a "median tree" is a useful construct for many purposes, especially when the fraction of genetic material laterally transferred or composed is small, but a strict tree discards information relevant to compositional mechanisms, and a tree with added cross-links is somewhat inconsistent in what the edges of such a graph represent.

The genetic variation mechanisms underlying gradual and compositional evolution are shown in figure 1.3 and figure 1.4, respectively. Note that the genetic changes

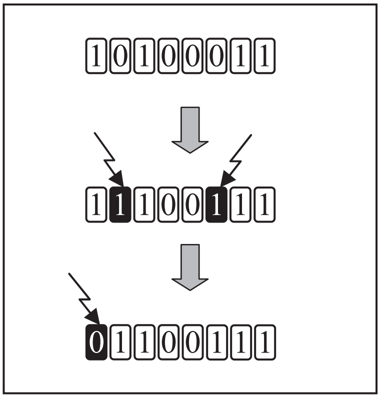
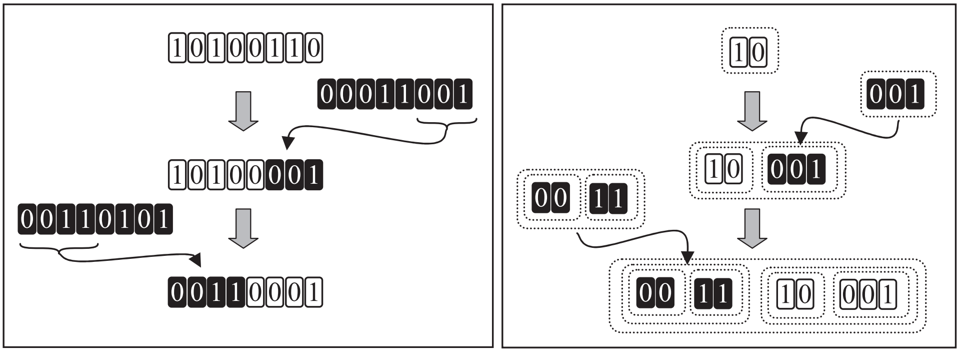


Figure 1.3
Gradual evolution via spontaneous point mutation. Each row represents an individual; each box represents a gene; the value (0 or 1) in each box represents the different forms, or *alleles*, each gene may have. The position of each gene is referred to as a *locus*, and the combination of alleles at all loci defines an organism's *genotype*. When the genetic material is copied during reproduction, errors may be spontaneously introduced—these are genetic *mutations*. Thus the descendents (shown in successive rows) have genotypes that differ in one or more alleles. Under mutational variation, newly introduced genetic material (shown in black) is random.



Sexual recombination Symbiotic encapsulation

Figure 1.4
Mechanisms of compositional evolution. Left: sexual recombination: sections of genetic material may be transferred as a set from one individual to another. Right: hierarchical encapsulation of symbiotic groups: preadapted sets of genetic material may be assembled by endosymbiosis. In both mechanisms, newly introduced genetic material has already been subject to selection. An interesting difference between these mechanisms that is expanded in the models that follow is that whereas the size of individuals remains constant under sexual recombination, the size of individuals grows under symbiotic encapsulation.

produced by spontaneous point mutation may be large, affecting numerous loci simultaneously, if mutation rates are high. And presumably, there is some nonzero chance that such large changes may be beneficial. But since the genetic changes from mutation are random, beneficial large genetic changes are likely to be rare and evolution is driven predominantly by the accumulation of small modifications. The primary form of gradual variation mechanism in nature is spontaneous genetic mutation, but other forms of genetic variation acting on one individual (or involving genetically similar individuals) may also produce genetic changes that can be accommodated in the gradualist framework.

Of course, spontaneous point mutation as depicted in figure 1.3 is also involved, if not required, for creating the diverse individuals undergoing the compositional change shown in figure 1.4. But the fact remains that compositional mechanisms enable large genetic changes that (unlike changes from spontaneous point mutation, whether they be large or small) are not arbitrary changes in the genetic material of the recipient—rather, the changes that are thus enabled are informed by prior adaptation of the donor individual. The results of such processes are not formed by the gradual accumulation of small changes, or random changes of any size, to either of the ancestral types.

Compositional mechanisms, such as those depicted in the previous figure, form part of a general evolutionary process of subdivision and integration. Abstractly, a process of subdivision and integration exhibits the following components. Evolving entities of different types are evolved in parallel lineages—that is, the population of evolving entities is *subdivided*. Subsequently, some mechanism of *integration* composes genetic material from different lineages together into a new entity. After integration, integrated genetic material reproduces together. Figure 1.5 shows a schematic of this process.

Mechanisms of subdivision and integration are seen at different scales in natural systems. Specifically, at the single-species scale, the separate lineages involved may be subpopulations of a given species, semi-isolated perhaps by virtue of spatial location, and the mechanism of integration may be provided by migration and sexual recombination. At the multispecies scale, the separate lineages may be reproductively isolated species, and integration may occur through horizontal gene transfer or endosymbiosis. Several of the major transitions in evolution share the characteristic that “entities that were capable of independent replication before the transition can replicate only as part of a larger whole after it” (Maynard Smith and Szathmary 1995); and this notion is important in providing the idea that processes of composition may occur through several increases in scale.

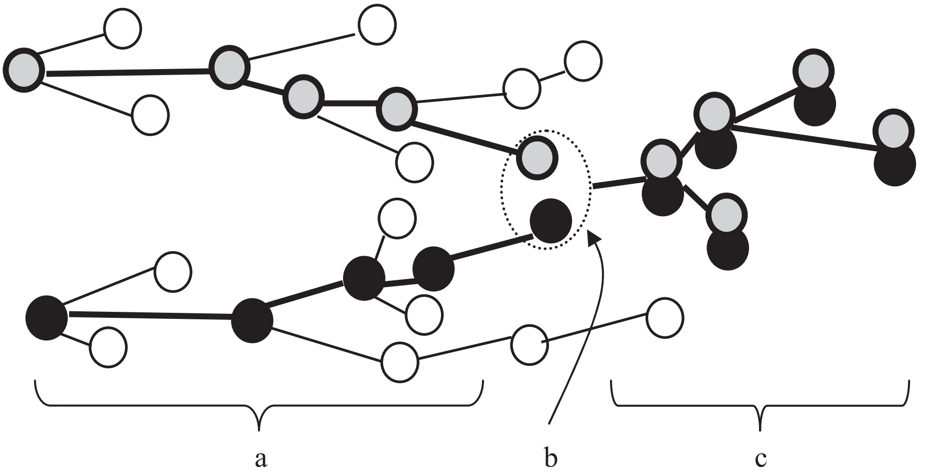


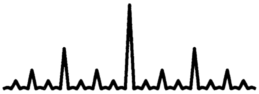
Figure 1.5
Subdivision and integration in compositional evolution. A number of reproductive entities, shown by nodes, in different lineages. Lines in the graph represent ancestral relationships with evolutionary time progressing from left to right: (a) two particular lineages (light shading and dark shading) progress independently in a subdivided population; (b) some mechanism of integration, such as sexual recombination or symbiotic encapsulation, composes together preadapted subsets of genetic material from the previously separate lineages; (c) the integrated genetic material is now replicated together. The integration of genetic material from subdivided populations will occur repeatedly in compositional evolution.

In evolutionary biology, the single-species scale is relevant to models such as shifting balance theory (Wright 1977), and the multispecies scale is relevant to models such as serial endosymbiosis theory (Margulis 1993a). In evolutionary computation, the single-species scale is relevant to issues such as diversity maintenance methods, the utility of crossover, and the building-block hypothesis (Holland 1975, 2000; Goldberg 1989)—and the multispecies scale is relevant to models such as cooperative coevolution (Potter 1997) and the explicit encapsulation of primitive components into reusable composites in models such as automatic module acquisition (Angeline and Pollack 1993).

1.2 The Algorithmic Paradigms of Evolution

I claim a fundamental distinction between gradual and compositional evolution based on the likelihood of evolving certain kinds of complex systems. The basic intuitions here are provided by analogy with well-known classes of optimization algorithms in computer science. In optimization, different kinds of optimization algo-

Table 1.1
Algorithmic classes and their evolutionary analogues.

Dependency of variables	Arbitrary interdependencies	Modular interdependencies	Few/weak interdependencies
Landscape			
Algorithmic paradigm	exhaustive search, random search	divide-and-conquer problem decomposition	hill-climbing—accumulation of small variations
Complexity	K^N	N^K	KN
Evolutionary analogue	"impossible"/ "intelligent design"	compositional evolution	gradual evolution

rithms are suitable for different kinds of problems according to the assumptions we have about the dependencies between the problem variables being optimized. A dependency between variables occurs when the optimal setting for one variable is dependent on the setting of other variables. Table 1.1 outlines three main classes of assumptions and the algorithm types that are appropriate for each of them. Starting with the leftmost column, when we have no knowledge of the nature or structure of interdependencies between the problem variables we might assume that the dependencies are arbitrarily difficult. Such dependencies create a random, highly rugged fitness surface, and the appropriate algorithmic paradigms for this class of problem are exhaustive search or random search. Neither of these methods uses any assumptions about the problem domain to guide or reduce the search necessary and therefore the expected time to find high-fitness configurations of variables with these methods is proportionate to the size of the entire search space. The total number of possible configurations for a problem that has N variables, each being able to take one of k values, is k^N. This is an exponential function of N. It means that the size of the search space grows very rapidly with the number of problem variables and a search method that has an operating time proportional to this is feasible only when the number of problem variables is small. In computer science, we say that the computational complexity of a method like this is exponential. This is a very important categorization because it means that the method is infeasible for large problems.

At the other extreme, addressing the rightmost column, we may in other cases hold different assumptions about the interdependencies of variables—in this case, that interdependencies between variables are weak or few. This creates a relatively smooth and generally monomodal fitness landscape, where the appropriate algorithmic paradigms are "greedy" algorithms making incremental improvements, and the simple class of "hill-climbing" algorithms which climb the local gradient in the fitness landscape. When the assumption of weak interdependencies between variables is correct, the expected time to find high-fitness configurations of variables is at worst linear, that is, simply proportionate to the number of variables in the problem. This is a very different class of computational complexity. It means that as N increases, the difficulty of solving the problem increases proportionately—a problem with $2N$ variables is only twice as difficult as problem with N variables. Unlike exponential complexity, this kind of increase is quite manageable and is feasible even for large N.

These two cases—on the one hand, arbitrary interdependencies and exponential time complexity, and on the other hand, weak interdependencies and linear time complexity—represent the naive extremes of the possibilities that might occur in a problem domain. These extremes correspond closely to extreme positions prevalent in evolutionary thought: specifically, if one assumes that the interdependencies between components of a natural system are arbitrarily difficult and complex then it is inconceivable that that system evolved (or at least, the size of such systems that can be achieved with "uninformed" processes is limited); in contrast, if one assumes that the interdependencies between the components of a natural system are weak or few then such systems are easily explained by the normal gradualist model of evolutionary change. Note, however, that if a search method is applied to an inappropriate problem type it will not succeed. For example, when a hill-climber is applied to a problem with arbitrarily difficult interdependencies it will not find high-fitness solutions in less than exponential time (and will generally perform worse than random search because it will follow local gradients that, in this case, will be uninformative). This is why large complex problems with arbitrary interdependencies are unevolvable under the gradualist framework of evolution.

In between these two extremes there are many other possibilities for the number, structure, and nature of variable interdependencies, and in optimization there are important algorithmic paradigms that address this range. One possibility for the interdependencies of variables is that there may be strong and numerous interdependencies but that these dependencies may have a particular structure—for example, a modular structure. It is not so obvious how to represent the high-dimensional landscape of a modular problem domain in the simple one-dimensional caricatures used here, but as we will see later, although they may appear to be as rugged and unforgiving as landscapes with arbitrarily complex interdependencies, the properties of semi-independence among problem variables create significant structure that can be exploited. The algorithmic paradigm that is appropriate for problems of this class broadly includes those known as "divide-and-conquer" techniques. Divide-and-conquer optimization is a form of problem decomposition that utilizes the modular structure of a problem by breaking it into smaller subproblems that are easier to solve, and using solutions to these subproblems to compose a solution to the problem as a whole. For example, if a problem can be divided into two equal halves then each requires time exponential in $N/2$ (i.e., half the number of problem variables), at worst. If each subproblem can be further decomposed, and so on, the time complexity of the whole process can be reduced to polynomial time—that is, the time to optimize the entire problem is some polynomial function of the number of variables in the problem. This means that although large problems are not as easy to solve as when the complexity is linear, they are still a lot easier than when the complexity is exponential, and they remain feasible for large N. Consider how the values $2N$ (linear), N^2 (polynomial), and 2^N (exponential) grow with increasing N. In computational complexity theory, the most important distinction is whether the complexity of a method is exponential or less than exponential—if a method takes exponential time it is simply infeasible for large problems, but if it is polynomial (or less) it is feasible for large problems. So, when applied to problems of a suitable type, hill-climbing and divide-and-conquer methods are feasible for large N.

Following these broad algorithmic classes, the basis of the central claim in this book is that compositional evolutionary mechanisms can provide adaptation akin to divide-and-conquer problem decomposition under certain circumstances. This means that some kinds of large complex systems that could not feasibly be evolved under the gradualist framework are nonetheless evolvable. As is the case for divide-and-conquer methods in general, the adaptive advantage of compositional mechanisms will depend on the structure of interdependencies in the problem domain—in particular, the presence of a modular interdependency structure. If such a structure is present, then compositional mechanisms may provide a form of adaptation that lies in between, yet is fundamentally distinct from, either of the extreme scenarios mentioned above.

The possibility of compositional evolution, and the fundamental distinction from gradual evolution, shares underlying intuitions with some previous work (e.g., Bermudes and Margulis 1985; Margulis 1993a). But the association of these intuitive ideas with this algorithmic class has not been previously developed, nor have formal analyses of the relevant combinatorics been developed in evolutionary theory. In computer science, the underling intuition here is basically that which underlies the building-block hypothesis, which has been present in evolutionary computation, more or less explicitly, since its inception (Holland 1975, 2000). But despite several attempts to demonstrate the validity of the hypothesis (e.g., Mitchell, Holland, and Forrest 1995; Forrest and Mitchell 1993b), this has proved difficult. Chapter 3 details the relevant background on concepts relevant to compositional evolution that appear under various guises through both the evolutionary biology and evolutionary computation literature.

In large part, the difficulty in formalizing the possible impact of compositional evolution in natural evolution, and the difficulty of demonstrating optimization of a compositional style in artificial evolution, stems from an inadequate understanding of modularity. In the next subsection, I introduce a class of complex systems that is sufficient to properly exemplify the difference in the adaptive capacities of gradual and compositional mechanisms.

1.3 Complex Systems with Modular Interdependency and Their (Un)evolvability

A system can be understood as *modular* if it can be described in terms of subsets of variables that are semi-independent (Callebaut 2005, p. 6). Here we are interested in whether modular subsystems have a semi-independent effect on fitness. If this is the case then it will be possible for them to evolve semi-independently. But usually the utility of a module is not independent of context; it may depend strongly on the state of the rest of the system. Accordingly, a module that is fit in one context may be very different from a variant of that module that is fit in another context, and getting the right modules together may be critically important for fitness. We understand a system to be modular not because it has weak dependencies between subsystems, but because the modules encapsulate the internal complexity of subsystems and thereby reduce the dimensionality of the system. I call the property of having modules with significant intermodule dependencies *modular interdependency*.

In this kind of system, changing a module to resolve intermodule dependencies may require changing many genes simultaneously. Accordingly, gradual evolution is unable to resolve the interdependencies between modules and is thus unable to find high-fitness configurations of the system. In contrast, compositional mechanisms can in principle “swap in” a module that has been preadapted in another individual. Compositional mechanisms that are able to discover and manipulate modules effectively are thereby able to search combinations of modules and satisfy intermodule dependencies more effectively.

Chapter 4 details a particular class of system exhibiting modular interdependency and builds on this to illustrate the distinction between gradual and compositional evolution. I start with a basic pairwise relationship between variables: specifically, I

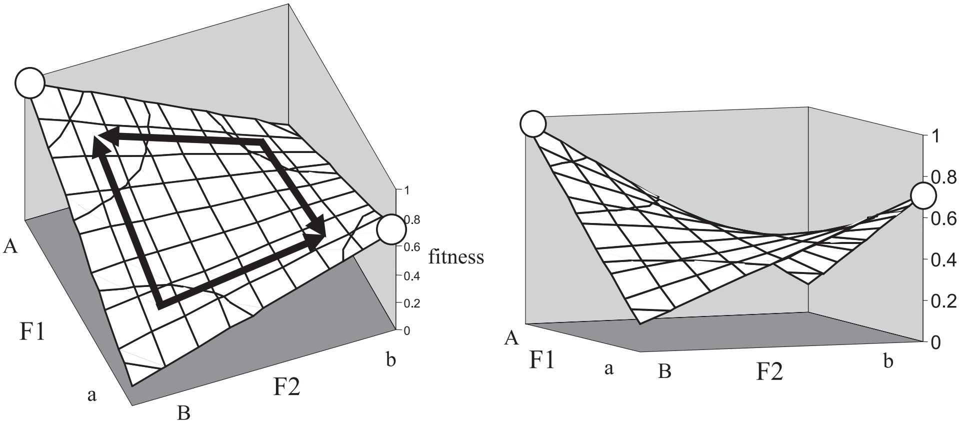


Figure 1.6
A fitness saddle created by interdependency between two variables. A system of two variables, or features, F1 and F2, showing nonadditive fitness interaction or epistasis. Left: The surface shows a fitness landscape and the overlaid arrows show the paths of increasing fitness that can be followed by changing one feature at a time. Right: This particular kind of interaction creates a fitness saddle that is shown more clearly on the rotated view of the same surface. This fitness saddle creates two optima indicated by the large dots located at AB and ab. An algorithm that arrives at ab cannot cross the fitness saddle to AB without changing both features at once.

define the fitness interactions between two variables such that a “fitness saddle” is created. This is illustrated in figure 1.6.

I then use this pairwise interdependency over a larger system of variables to define a system with modular interdependency. Figure 1.7 illustrates an example system exhibiting modular interdependency. The highly regular and symmetric construction used in this particular example enables us to see clearly the underlying structure that is important (though the regularities of this example are not, in fact, required). The dependency matrix on the left of the figure shows the strength of pairwise dependencies between eight problem variables, and the graph on the right depicts the same relationships. Sets of dependencies with equal strengths are indicated by appropriate borders in the table, and by the number of edges and (approximately) by the spatial proximity of nodes in the graph. The fitness of a given configuration for the system of eight variables can be calculated from the sum of pairwise interactions (as per figure 1.6) weighted by the corresponding entry in the dependency matrix.

The salient feature of this structure is that there are subsets of variables in which the variables are more strongly dependent on other variables within their own subset than on variables in other subsets. This creates a dependency matrix where the large

	1	2	3	4	5	6	7	8
1	-	16	4	4	1	1	1	1
2	16	-	4	4	1	1	1	1
3	4	4	-	16	1	1	1	1
4	4	4	16	-	1	1	1	1
5	1	1	1	1	-	16	4	4
6	1	1	1	1	16	-	4	4
7	1	1	1	1	4	4	-	16
8	1	1	1	1	4	4	16	-

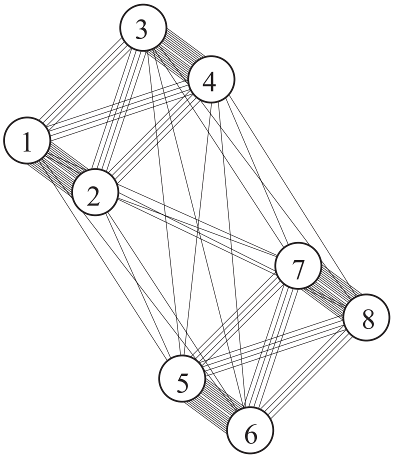


Figure 1.7
A system of variables exhibiting modular interdependency.

values are diagonalizable. But significantly, the values of the matrix that are off the diagonal, representing the strength of dependencies of variables across modules, are nonzero. Clearly, the structure of dependencies in this example system are not just modular but also hierarchically modular—that is, there are clusters and subclusters of more strongly interdependent variables. This potentially allows an adaptive mechanism that is capable of exploiting modularity to decompose the system recursively, or equivalently, to compose together subsolutions repeatedly. This kind of hierarchical decomposability is closely related to the notion of nearly decomposable systems discussed by Simon (1969).

This kind of system may be interpreted in different ways as appropriate for different domains. For example, we may interpret the graph on the right of figure 1.7 as the network structure of a dynamical system where the future state of each variable is dependent both on itself and the state of connected nodes (in the style of "random Boolean networks," Kauffman 1993). A biological example using a dynamical interpretation equates each variable with a gene, and interdependencies between variables may represent transcription factor interactions that "upregulate" or "downregulate" one another. In this interpretation, the fitness of different configurations corresponds to the stability of attractors in the gene expression dynamics. An alternate interpretation is that subsets of genes interact during development to control the attributes of higher-level features, and sets of these interact to create still higher-level features, and so on. If the interaction between genes in one subset and genes in another subset can

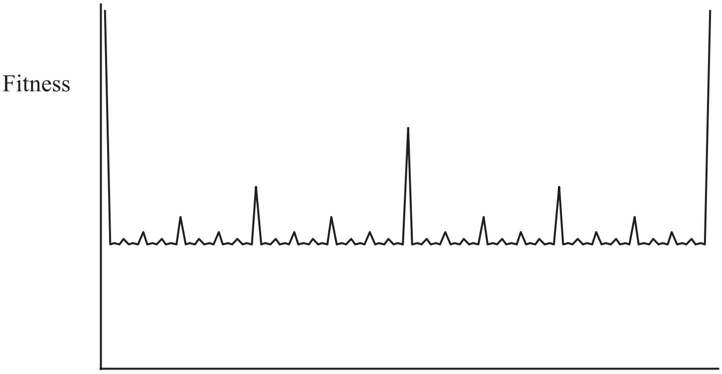


Figure 1.8
A particular cross section through the fitness landscape of a system with modular interdependency. See section 5.1 for details.

be understood through the interaction of their respective higher-level features, then the system will exhibit modular interdependency.

This system of modular interdependency is particularly suited to supporting the general point—that certain kinds of complex systems that are considered unevolvable under the gradualist framework of evolution are easily evolvable under compositional evolution. Its clear modular structure makes it very easy for a compositional mechanism to adapt; however, it also exhibits all the properties usually associated with evolutionary difficulty. Specifically, it creates a highly rugged fitness landscape with an exponential number of local fitness peaks; the peaks are separated by broad fitness saddles (in fact the higher the fitness of a peak the further it is to the next peak of higher fitness); high-fitness configurations of the system are irreducibly complex (in the sense that any small change to the configuration causes a catastrophic decrease in fitness); and finally, there is no guaranteed path of monotonically increasing fitness leading to high-fitness configurations in this system. Figure 1.8 shows a cross section through the fitness landscape created by a complex system with modular interdependency that illustrates most of these features. However, all of these concepts of evolutionary difficulty are dependent on the assumption of gradual evolution. Chapter 5 shows that this kind of system cannot be evolved by a gradual evolutionary process, regardless of the mutation rate that is used. In contrast, when such a system is evolved under compositional evolution these properties of the system do not create evolutionary difficulty. This result depends on the presence of appropriate modularity, a property that gradual evolution cannot exploit but compositional evolution can.

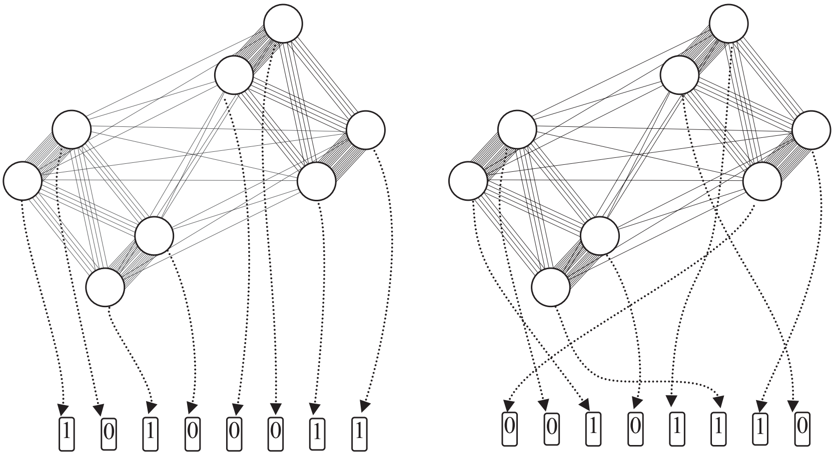
1.4 Compositional Mechanisms

This book provides two highly abstract computational models to illustrate a sufficient set of mechanisms and conditions for compositional evolution to produce systems with hierarchical modular interdependency. The first model is based on sexual recombination (see figure 1.4, left); the second is based on symbiotic encapsulation—the hierarchical encapsulation of symbiotic groups inspired by endosymbiosis and the major evolutionary transitions (see figure 1.4, right).

Composition Based on Sexual Recombination

The simulation models in chapter 6 show that evolutionary processes with appropriate population diversity and variation provided by sexual recombination are able to provide compositional evolution. Specifically, a form of genetic algorithm using sexual recombination, is able to quickly and reliably find the global optima in this problem class. Different individuals in the population discover different modules, and sexual recombination is able to "swap in" modules from one individual to another to discover better combinations of modules that resolve the dependencies between these modules. This provides our first successful example of compositional evolution.

However, the ability of sexual recombination to search combinations of modules is strongly dependent on assumptions about the "genetic map," the placement or ordering of genes on the chromosome (see figure 1.9). In linear chromosomes, a system of genes must be mapped to positions on the chromosome, and the distance between genes in this ordering affects the likelihood that they will co-occur in offspring produced by recombining with other chromosomes. Ideally, features (nodes in the graph of figure 1.7) which are functionally interdependent (shown by numerous interaction lines) will be next to each other on the chromosome (figure 1.9, left). In this case, sexual recombination (in some circumstances) will be able to recombine subsystems effectively. However, without such a favorable situation, functionally interdependent subsets of genes may be arbitrarily positioned on the chromosome (figure 1.9, right). I'll call this a "random genetic map." In this case, sexual recombination is unable to recombine subsystems effectively—even though different well-adapted modules may have been discovered in different members of the population, they cannot be extracted as a unit and recombined effectively. In biological systems, the position of genes might be adaptively reordered, but in engineering domains where the interdependencies of variables are not known *a priori*, we cannot assume a favorable ordering of genes on the chromosome.



“Tight” genetic map

Random genetic map

Figure 1.9
The correspondence (or not) of the genetic map with epistatic dependencies. The graph of nodes represents the strength of fitness dependencies of genes (subsets of genetic material) in the evolving system—this is modular in both cases. In a “tight” genetic map the modularity of fitness dependencies corresponds well with the strength of physical linkage between genes. However, if the position of genes in the genetic map is arbitrary (random genetic map) the physical linkage of genetic material does not correspond well with the fitness modularity of genetic material.

Composition Based on Symbiotic Encapsulation

Despite the limitations of composition based on sex, the ability of compositional mechanisms to exploit the structure of problems with modular interdependency is not necessarily dependent on assumptions about the favorable positioning of genes on the chromosome. In chapter 7, I provide a novel model of evolutionary adaptation based on the encapsulation of coadapted simple entities into a new whole. This symbiotic encapsulation model is derived from a mechanism of symbiogenesis or endosymbiosis rather than sexual recombination, and enables compositional evolution in more general circumstances than that provided by sexual recombination.

The salient distinction between sexual recombination and symbiotic encapsulation for our purposes is that whereas sexual recombination produces an offspring by taking, on average, half the genetic material from one parent and half the genetic material from a second parent, symbiotic encapsulation may simply take the union of genetic material from both donor individuals. Naturally, this is most productive when the expression of genetic material from the two parents is not mutually exclusive. Accordingly, the model of this mechanism allows for the possibility of entities that specify complementary characteristics rather than overlapping or mutually exclusive genetic material. This allows whole individuals to represent a functional module that can be assembled with another whole individual, and precludes the need for a mechanism like sex to take a part of one individual and a part of another where it is necessary to use the genetic map to bias which parts to take.

The model is also *coevolutionary*—involving the evolution of entities that are evaluated in the context of one another and thereby coadapt. In contrast to the common adversarial relationships seen in many coevolutionary models (e.g., Hillis 1992; Reynolds 1994; Pollack and Blair 1998), the coevolutionary relationships we are most interested in here occur between entities having complementary adaptive abilities (see Sigmund 1998).

The idea is that initially different specialist entities will evolve to cover different complementary parts of the problem domain. Over time, the dependencies between specialists will be resolved by forming unions between specialists to create new reproductive entities at a higher level of organization. This process repeats, discovering larger alliances until a complete solution is found—a generalist that solves all modules and resolves all intermodule dependencies in the hierarchical problem structure.

This model must accommodate entities of increasing size, as entities of many levels of organization are modeled together, and thus requires a means to prevent large suboptimal generalists from competitively excluding optimal specialists. This proved to be a critical aspect of the model and required methods that step outside frameworks previously familiar in evolutionary computation. The model developed, the "symbiogenic evolutionary adaptation model," or SEAM (to invoke the notion of joins), uses a selection scheme that explicitly respects the *context sensitivity* of an entity's fitness. That is, a proposed solution to a module may be good in one context and not in another, in the same way that the fitness of an organism is strongly dependent on environment. Selection in a varied environment may prevent competitive exclusion of one type of entity by another if the niches they inhabit are distinct, but in the case where one entity is fitter than another in all environments we may assume that the former will competitively exclude the latter. This is modeled explicitly in SEAM by retaining multidimensional measures of fitness coming from the performance of a specialist in different contexts, and applying selection conservatively using the "Pareto dominance" criterion. This provides an abstract form of automatic niching in an ecosystem of different species and replaces the normal single-dimensional fitness selection familiar in single-species models.

SEAM is able to properly identify and optimize the modules in this class of problem automatically and assemble them together to find larger modules repeatedly. It is thus able to quickly and reliably find globally optimal configurations of the problem variables in this class of hierarchical modular interdependency systems. The behavior of SEAM is entirely insensitive to the ordering of genes on the chromosome. Thus I show that the algorithmic advantage of compositional mechanisms in this class of problem is not dependent on assumptions of favorable gene ordering.

In summary, the abstract models of sexual recombination and symbiotic encapsulation illustrate sufficient sets of conditions under which the adaptive capacity of compositional mechanisms is fundamentally different from that of gradual mechanisms. Specifically, I show that these mechanisms are able to exploit modular interdependency structure and provide efficient adaptation in cases where gradual mechanisms are not able to do so. The models used illustrate that the adaptive potential of sexual crossover is sensitive to gene ordering on the chromosome, but this is not a principled limitation of compositional mechanisms since a model based on symbiotic encapsulation is insensitive to gene ordering.

Simulation results from these models and algebraic analysis provide an existence proof for the central claim: certain kinds of complex systems, considered unevolvable under the gradualist framework of evolution, are easily evolvable under compositional mechanisms. More formally, chapter 8 shows that although the expected time for a gradual mechanism to evolve systems with modular interdependency is exponential in the size of the system, the expected time for the compositional mechanisms is polynomial. This is a formal way of saying that systems of this kind quickly become unevolvable to gradual mechanisms when they are of any significant size, but that even large systems are easily evolvable via compositional mechanisms.

1.5 The Impact on Gradualism

These models and results significantly affect the kinds of systems we consider to be evolvable and unevolvable. For example, systems of many complex interdependent parts can appear unevolvable—especially when any small change in the system causes the system to cease functioning (Behe 1996). In general, when faced with the existence of a particular complex system in nature, the normal approach to explain how it may have evolved is to show that there exists a succession of protosystems that are gradually increasing in function where each one is a small modification of the last. Accordingly, a straightforward approach to explain the plausible evolution of a system where it is proposed that any small change causes it to cease functioning is to show that this is not in fact the case—that there does after all exist a succession of gradually changing protosystems approaching the system in question (Dawkins 1996).

More generally, we are interested in affirming the existence of a linear succession of protosystems where each one is a plausible random genetic change from the last—this tends to imply that changes will be small, but in principle, they may be changes of any size so long as they are random. If such a succession can be found, or reasonable hypotheses about the plausible existence of such a path can be upheld, then all well and good. But the adherence to this framework, the linear accumulation of random genetic changes, derives from assumptions about the underlying algorithmic principles.

Specifically, the simulations and analyses show that the existence of a succession of gradually changing protosystems is not, in principle, essential to explain the evolution of a complex adaptation. Although some intuitions to this effect have been in circulation, this book enables us to clarify the fundamental algorithmic distinction between compositional and gradual mechanisms; to define a class of complex systems that distinguishes their capabilities; to analyze the properties of such systems that make them problematic for gradual evolution; and to provide a sufficient set of circumstances under which they are easily evolvable via compositional mechanisms. The contribution of these results to evolutionary biology is not, of course, to merely point out that compositional variation mechanisms exist, but rather that compositional evolution belongs in a different algorithmic paradigm from gradual evolution, and that accordingly it can evolve things that gradual evolution cannot.

In chapter 8, I provide formal analyses of what "can evolve" and "cannot evolve" mean in a rigorous sense. This depends on showing that one process cannot be guaranteed to find high fitness points in less than time exponential in the size of the system, and that another is guaranteed to find high fitness points in time that is at worst a polynomial function of the size of the system. Basically these match with a biologist's intuitions: if a fitness improvement requires many genetic sites to be changed simultaneously to a specific new configuration, then the expected waiting time for this when the changes are made randomly is exponential in the number of sites that need to be changed. It just is not plausible for more than a small number of sites. In chapter 8 I provide formal analyses of the compositional models to see how long they take to make such changes, and show that (given some necessary assumptions) there is always a relatively easy genetic change available to find fitness improvements in these compositional models.

In chapter 9, I connect the features of the compositional models back to the hierarchical structure of the modularity and discuss the mechanisms and principles necessary for an evolutionary process to "scale up" evolution. Whereas mutational processes start out and end up manipulating individual nucleotide sites, compositional mechanisms begin this way but, through evolutionary time, scale up the variational mechanism to manipulate larger, more meaningful components: genes, suites of genes, and so on. This enables compositional processes to escape from the inherent tension of variability and heritability that is inevitable in random-mutation-based processes, for example.

I have introduced several important conceptual distinctions in the previous discussion, so let me recap. The grandest distinction is between exponential complexity and nonexponential complexity; in evolution these correspond to not-evolvable and evolvable, respectively. Included within nonexponential complexity is linear complexity and polynomial complexity. When applicable, hill-climbing is linear complexity and divide-and-conquer problem decomposition is polynomial complexity; so both of these are nonexponential. Gradual evolution is analogous to hill-climbing and compositional evolution is analogous to divide-and-conquer, so both of these make the evolution of large systems feasible when applicable. However, the types of system to which they can be applied are different and some types of systems that are feasible with compositional evolution (e.g., modular ones) are infeasible with gradual evolution. Gradual evolution can be further divided into strict gradualism (small changes only) and nonstrict, which may include large changes in principle. And finally, compositional evolution can be implemented with different mechanisms including sexual recombination and symbiotic encapsulation. These can both evolve systems that gradual evolution cannot, but again they have different capabilities and sexual recombination is limited in some ways that symbiotic encapsulation is not.

Is Gradualism Sufficient?

The possibility of large adaptive changes, and the possible contrast with gradualism, is already recognized in the evolutionary biology literature (see, e.g., *Symbiosis as a Source of Evolutionary Innovation*, Margulis and Fester 1991). However, there is disagreement about whether symbiotic encapsulation contradicts the supposed sufficiency of gradualism to explain evolutionary change.

Ridley (1985), having specifically mentioned symbiosis as a possible source of increases in complexity, states that "The Darwinian denies (as he must) that the complex co-adaptation arose in a single chance event" (p. 35)—and "in all these cases, including symbiosis, complex organs have been built up in small stages" (p. 41). So each adaptation "appears at first impossible to build up in a series of small but advantageous stages. But in fact they probably were" (p. 41). And thus Ridley concludes that "[c]omplex adaptations must have evolved by the natural selection of a large number of small mutations over a long period of time" (p. 42).

It is correct that the Darwinian must deny that a complex adaptation arose in a single *chance* event. But we must not therefore conclude that linear incremental improvement is sufficient to explain all adaptations. If Ridley is referring to the components that were assembled, then maybe (presumably) they evolved gradually (excluding the possibility of repeated endosymbiosis, for the moment). But the resultant system containing both the component parts did not evolve gradually. It did not evolve exclusively via the accumulation of a large number of small mutations because, in addition to this, there was a large inclusion of nonrandom genetic material that was preadapted in a different lineage. And this means, as figure 1.1 shows, that the resultant system is not derived from an unbroken line of small genetic changes. But does this contradict Darwinian gradualism or not?

On the subject of the apparent contradiction with Darwinian change, Maynard Smith (1991) states: "Symbiosis may give rise rather suddenly to evolutionary novelty; it is therefore seen as presenting a challenge to Darwinian gradualism. I think this is to misunderstand the reason why Darwin was a gradualist: essentially, it was because the origin of a complex adaptation would be miraculous." This does not really support the idea that Darwinian gradualism is correct; in fact, it says that Darwinian gradualism is incorrect, but that something deeper is still correct. Specifically, the deeper thing that is correct is the theory of evolution by natural selection—but this needs to be separated from gradualism.

Maynard Smith then suggests that in light of mechanisms that exchange genetic material across lineages, we see that a complex adaptation acquired in such a manner is not "miraculous," and—"There is, therefore, no contradiction between Darwin's belief that complex adaptations arise by the natural selection of numerous intermediates, and the possibility that new evolutionary potentialities may arise suddenly if genetic material that has been programmed by selection in different ancestral lineages is brought together by symbiosis" (ibid., p. 37).

Although it is easy to see that there is nothing miraculous, we should be careful with what we conclude from this. Of course, if a complex adaptation arises by symbiosis it does not come out of thin air but arises from the assembly of "intermediates." But if we assume that intermediates must form a single, unbroken, linear succession of small changes, then the possibility that a complex adaptation arises suddenly by symbiosis *does* present a contradiction to this. We could give Maynard Smith the benefit of the doubt here, but I think it would be fair to say that his statement strongly suggests that it is fine to carry on as though compositional events do not change anything important in our understanding of evolution, and that he has overlooked the possibility that intermediates from parallel lineages may provide different affordances than intermediates in a single lineage.

So, why is it important to make these seemingly nitpicky distinctions? We are all agreed that evolution by natural selection explains these observations. Nothing miraculous has occurred. There are no large random mutations that are beneficial. So what point is there in emphasizing that there is no linear succession of small random changes that connects the result of a symbiotic event to its ancestral lineages? The point is that by adhering to the framework of linear incremental improvement, we miss the fact that compositional events radically alter the underlying algorithmic principles and thereby the space of systems that can evolve. Compositional evolution enables the evolution of systems that would be otherwise unevolvable, and if we adhere to linear gradual evolution thinking, we make incorrect assumptions about which kind of system is evolvable and which is unevolvable.

What does this say about the necessity and sufficiency of gradualism in evolutionary change? Although there are no *miraculous* large changes, there *are* large changes. And moreover, it is not possible to see where these (nonmiraculous) large beneficial changes come from if we view evolution as a linear hill-climbing process. If the *components* that are brought together in a compositional event evolved gradually then gradualism is necessary in order to explain the evolution of the new entity. But it should be clear that even if the components evolved gradually, *the new entity* was not evolved gradually. Specifically, there is no path of small changes of monotonically increasing fitness approaching the configuration of features exhibited in this new entity. Thus the evolution of this entity is not accommodated by the model of linear incremental improvement.

A more fundamental question then arises from these considerations: if some entity AB was created by the composition of entities A and B, each of which evolved gradually, does it follow that AB *could* have evolved gradually (even if it did not in this particular historical scenario)? I will show that this does not necessarily follow. Specifically, I will show that it may be the case that although A and B are each independently evolvable gradually, AB is not evolvable gradually as a whole—nor are the two entities evolvable one after the other in a single individual. This will be the case when A and B correspond to *interdependent modules* in the sense that I define in chapter 4. As I will show, in systems with modular interdependency, each module is independently evolvable but the two modules cannot evolve together gradually because of dependencies acting between the modules.[2] Put simply, when there are interdependencies between A and B, it may be possible for A to evolve by linear incremental improvement, but if it does it becomes essentially impossible for B to evolve without causing dramatic fitness decreases by disrupting dependencies with A, and vice versa. Nonetheless, if A and B each evolve in parallel in different individuals then, even though not all variants of A work well with all variants of B (because of the dependencies between them), searching combinations of different variants is a relatively easy evolutionary process through compositional mechanisms.

As stated, the strongest evidence supporting the claim that it *does* matter whether genetic adaptations are accumulated linearly, or in parallel and then composed together, is that it changes our perception of what is evolvable and what is not. This is why I am able to show that a complex system that has all the characteristics usually associated with evolutionary difficulty (an exponential number of local fitness peaks, wide fitness saddles separating a local peak from the nearest point of higher fitness, configurations that are irreducibly complex, and fitness optima that cannot be reached through a linear succession of incremental improvements) is nonetheless easily evolvable through compositional mechanisms. In the course of this book, I will explain why this is possible by showing that compositional evolution belongs to a different algorithmic class from simple mutation hill climbing.

1.6 Some Related Issues

There are numerous related issues in both evolutionary biology and evolutionary computation that are raised by these models and results. In this section I briefly acknowledge a few important ones.

Noncompositional Mechanisms

I need to distinguish compositional mechanisms from a few mechanisms that may look similar superficially. Gould's concept of *punctuated equilibria* (Gould and Eldredge 1977) does not concern large genetic changes but merely the rate (the unevenness of the rate) at which small changes accumulate—compositional mechanisms concern large genetic changes. The notion of evolutionary jumps known as "saltations" (Goldschmidt 1940) concerns large genetic mutations (see also *Mutation Theory*, DeVries 1909) but these are undirected random mutations, whereas the point of the compositional models is to show that composition affords directed genetic changes that cannot be achieved with a higher mutation rate, for example. Mechanisms such as genetic inversions and gene duplication (Ohno 1970) provide large genetic mutations that introduce nonrandom genetic material that is considerably less likely to be disruptive than a random genetic change of the same scale (because they are a copy of genetic material that has already been subject to selection)—but this genetic material has been subject to selection in the *same* individual as the rest of the genetic material of the organism, rather than preadapted in a different genetic context as it is in compositional mechanisms. Such duplications of existing genetic material do not provide an increase in complexity in the same way that compositional mechanisms do (Ridley 2001, p. 43). These mechanisms potentially exploit a different kind of modularity, the repeated structures kind of modularity (like the fingers of a hand, or cells of a tissue), rather than the functional independence kind of modularity (like metabolizing nutrients and detecting toxins) addressed in this book.

This also leads us to mention small genetic changes that can produce large phenotypic changes (such as homeotic genes, Lewis 1978). Such changes potentially enable large changes in phenotype that are better than random but this is a different issue from the large *genetic* changes addressed here.

Selection on Groups of Genes

Sewell Wright's shifting balance theory, which describes evolutionary mechanisms whereby a population may escape local optima on a fitness landscape, is related to some of the models presented in this book. The models share some of the underlying motives and thinking but there are significant differences in the mechanisms (see section 9.1). For example, genetic drift does not feature in my models for escaping local optima.

The idea of compositional evolution based on symbiotic encapsulation is basically the idea that subsets of genetic material that have evolved semi-independently become part of a single individual at a new level of organization. This might sound similar to group selection, where individuals are selected for their utility to the benefit of the group rather than to themselves. And since this principle is applied recursively through subsequent hierarchical levels, it might further appear to be a model of hierarchical selection (Keller 1999; Gould 1982a, 2002). But this is not the case. Here we are careful to consider the selfish interests of the parties involved in a symbiotic encapsulation, as required by a microscopic, "selfish-gene"-style model (Dawkins 1976).

Nonetheless, the notion of hierarchical selection makes good sense from an algorithmic point of view. The principle that affords efficient adaptation in these models is that first searching combinations of small numbers of variables finds good small modules, then searching combinations of modules finds bigger modules, and so on. If the search were conducted solely at the lower level then suboptimal solutions would result. And if the search were conducted solely at the higher level then the combinatorics of the search space would be prohibitive. But by implicitly moving incrementally from lower levels to higher levels we gain the algorithmic advantage of a divide-and-conquer approach and the opportunity, in appropriate problem domains, to find good global solutions efficiently. (See discussion section 9.1.)

Modularity

An important concept of modularity in the genetic algorithm literature is Holland's notion of "building blocks" (1975). This is closely related to the notion of modules we define in chapter 4, but it is distinct in several important respects. The modules that the modified GA processes in the compositional model of chapter 6 are indeed building blocks (but not necessarily small ones). In contrast, the modules used by symbiotic encapsulation in chapter 7 are not building blocks: They are represented by subsets of variables but there is no dependence between the position of genes on the chromosome and the ability of this method to exploit modularity. Additionally, it should not be assumed that the notion of modules implies a problem that can be decomposed into *separable* subproblems. This is a mistake that has caused widespread confusion in the controversy around the building-block hypothesis and the operation of the GA. There may be strong nonlinear interactions between modules—that is, modularity does not imply that a problem is separable.

The notions of hierarchical modularity, and hierarchical decomposability, relate strongly to Herbert Simon's (1969) ideas and the wealth of design-engineering and social systems he refers to. However, chapter 4 shows that a system of hierarchical modular interdependency does not depend on the idea of modules with weak inter-module dependencies.

Finally, one of the best-known models of genetic dependencies is Kauffman's "NK landscapes" (1993). When the number of variables involved in each dependency K is large, they produce rugged landscapes that are very difficult for gradual evolution. However, these landscapes have no structural modularity and are therefore also difficult for compositional evolution. We will see that two landscapes with the same number of dependencies can be very different in difficulty because of differences in the structure of those dependencies.

"No Free Lunch" and the Building-Block Hypothesis

The "no free lunch" theorem (Wolpert and Macready 1997) basically says that "you cannot get something for nothing"—specifically, you cannot have an algorithm that is better than random search unless you also restrict the class of problems to which it is applied. One of the primary evolutionary computation aims of this book is to clarify whether the EA can in some circumstances provide an algorithmic advantage akin to divide-and-conquer problem decomposition as the intuition of the building-block hypothesis suggests. As Wolpert and Macready rightly point out, for any algorithm, the algorithmic advantage of the method is dependent on the class of problem to which it is applied. Accordingly, it is not my intent to show that a GA, standard or otherwise, is advantageous in all problem classes. On the contrary, I invest a good deal of this book identifying and illustrating the properties of a particular problem domain—specifically, the class of problems with modular interdependency. Thus, any failure of the building-block hypothesis to explain the operation of the "simple GA" in other problem domains is simply not relevant to my conclusions. But having said this, I find the intuition of the building-block hypothesis to be quite instructive. In contrast, the "schema" theorem does not turn out to be useful in these models. The schema theorem (Holland 1975) concerns the increase in copies of schemata with above-average fitness in a population, and historically, it is considered responsible for the supply of building blocks for the building-block hypothesis. In the models developed in this book, the notion of schemata, and schema combination, is very useful, but we have no need to invoke or defend the schema theorem since the selection schemes are quite different.

1.7 Contributions

The main motives and goals of this book are: to foster and capitalize on two-way interdisciplinary exchange between evolutionary biology and evolutionary computation, to provide a conceptual framework that incorporates a broader view of evolutionary mechanisms than the gradualist framework, to understand the adaptive role (if any) of symbiosis and symbiogenesis, and to better understand the competence of genetic algorithms, and in particular the utility (if any) of crossover and the building-block hypothesis.

The specific contributions for evolutionary computation include: 1. Clarifying the structure of dependencies between variables in modular problem domains, in particular, hierarchically modular problems. (These are related to, but resolve important limitations of, building-block problems in GA theory.) 2. Demonstrating the conceptual contrast of hill climbing and divide-and-conquer problem decomposition mechanisms in evolutionary algorithms. (This is related to, but more general than, the notions of the building-block hypothesis.) 3. Providing a concrete illustration of mechanisms that permit automatic module acquisition and assembly in evolutionary algorithms. This provides a conceptual framework that accommodates cooperative coevolution and function optimization, addressing fundamental issues such as representational recoding and open-ended scalability in evolutionary search. 4. Providing specific models for algorithmic methods such as methods for maintaining appropriate schema diversity, and the use of partially specified genotypes to represent functional or fitness dependencies.

The specific contributions for evolutionary biology closely parallel these: 1. Clarifying the structure of functional or fitness dependencies that produce different kinds of evolutionary difficulty. 2. Demonstrating the conceptual contrasts of gradual evolution and compositional evolution, thereby broadening our interpretation of evolutionary processes. 3. Providing a conceptual framework for evolution that begins to accommodate adaptive roles for symbiosis and symbiogenesis. 4. Providing specific models for evolutionary processes and the role of issues such as the genetic map, population-ecosystem diversity, population subdivision, and shifting balance theory.

The models themselves are quite simple: the fitness landscape can be defined in a simple equation, and each algorithm can be defined in a few lines of pseudocode. However, since the purpose of the models is an illustration of deeper conceptual issues, they require a fair amount of setup in order to convey why the models are the way they are, and what the results do and do not mean.

It is important to be clear about the scope and boundaries of this book and these claims. I provide these models as an illustration of the important potential that these mechanisms afford *in principle*, and to encourage analysis of evolutionary processes such as these that go beyond the usual gradualist conception of evolutionary change. This follows a style of cross-disciplinary research such as, for example, connectionism, which did not claim to necessarily be an accurate model of the way that the cortex worked in a detailed sense, but was nonetheless informative by demonstrating an alternate possibility that presented a challenge to established cognitive science. Moreover, the mechanisms I explore are not arbitrary excursions into possible biologies. First, abstract computational models such as these are relevant to natural evolution at least in the sense that we demonstrate that it is possible for nonteleological adaptive processes, that is, processes based only on fitness feedback, to behave in this manner. Second, mechanisms of sexual recombination, horizontal gene transfer, and symbiogenesis, for example, are a biological fact. Third, some of the other concepts that are explored have a rich, albeit controversial, background in biological thought—and throughout the book effort is made to draw the relevant connections to the biological literature and provide appropriate biological interpretations of models and results.

Similarly, the computational contributions of this book are theoretic and conceptual. The studies provide proof of concept for the algorithmic techniques and address issues of principle, not empirical observations about engineering domains or processes. But likewise, the computational models build on an established set of research questions, algorithms, and conceptual ideas. Highly abstract models like those presented here require further work to connect properly with detailed biological knowledge and validate the many assumptions theoretical models require. But in the meantime the fundamental algorithmic distinctions of the models prompt us to expand the way we think about natural and artificial evolution to include new possibilities.