1

Introduction

Living things are unimaginably complex, yet they have withstood a withering assault of harmful influences over several billion years. These influences include cataclysmic changes in the environment, as well as a constant barrage of internal mutations. And not only has life survived, it has thrived and radiated into millions of diverse species. Such resilience may be surprising, because complexity suggests fragility. If you have ever built a house of cards, you will know what I mean: The house eventually comes tumbling down. Why is an organism not a molecular house of cards? Why do not slight disturbances (especially genetic disturbances in the form of mutations) cause key organismal functions to fail catastrophically? And is the robustness of organisms to change itself a consequence of past evolution? How does it affect evolvability, the potential for future evolution? These are some of the key questions I will address here.

A biological system is robust if it continues to function in the face of perturbations. This is the working definition of robustness I use in this book. The perturbations can be genetic, that is, mutations, or nongenetic, for example, environmental change. A variety of other names—buffering, canalization, developmental stability, efficiency, homeorhesis, tolerance, etc. (171, 183, 186, 368, 472, 499, 578)—have been used for the same phenomenon, and my choice of one among them is arbitrary. The above definition implies that one can sensibly discuss robustness only if one has clarity about two cardinal questions: What feature of a living thing is robust? And what kind of change is this feature robust to?

With respect to the first of these cardinal questions, it is clear that ultimately robustness of only one organismal feature matters: fitness—the ability to survive and reproduce. However, fitness is hard to define rigorously and even more difficult to measure. In addition, a change in fitness can have many different causes. For instance, a mutation that blocks a chemical reaction in metabolism affects fitness for different reasons than a mutation blocking embryonic development. An examination of fitness and its robustness alone would thus not yield much insight into the opening questions. Instead, it is necessary to analyze, on all levels of organization, the systems that constitute an organism, and that sustain its life. I define such systems loosely as assemblies of parts that carry out well-defined biological functions. Examples include DNA with its nucleotide parts, proteins with their amino acids, metabolic pathways and their enzymes, genetic networks and their genes, and developing organs or embryos with their interacting cells. A good part of this book surveys what we know about the robustness of biological systems on multiple levels of biological organization.

With respect to the second cardinal question, what are organisms robust to, this book has a restricted scope: It focuses on robustness to genetic change. I will call this kind of robustness *genetic robustness* or *mutational robustness*. This focus has three motivations. First, genetic change has more serious consequences than nongenetic change. A genetic change is a permanent alteration in the "wiring" of a biological system, and its effects, or lack thereof, thus deserve special attention. Second, by and large, mostly genetic change is heritable, and thus has much more serious long-term consequences on organismal lineage than nongenetic change. Thirdly, a comprehensive account of robustness against nongenetic change would be daunting. For instance, an exhaustive treatment of robustness to environmental change would have to include just about all homeostatic phenomena in biology. These phenomena include the regulation of osmotic balance, metabolite concentrations, and gene expression, thermoregulation in endotherm organisms, flight stabilization in birds, and on and on. The literature on many of these phenomena is already large and needs no further addition. Robustness to mutations, on the other hand, has not been as comprehensively studied. In addition, it is a well-defined phenomenon where a search for general principles that unify observations on different levels of organization is easier. I will propose some such principles here. All this is, of course, not to say that robustness to nongenetic change is unimportant. In fact, it is associated with mutational robustness and may be very important for the evolution of such robustness, as I argue in chapter 17.

Why Study Robustness?

The first and most important reason to study robustness is already stated in the opening paragraph: Why can unimaginably complex systems withstand so much change? As we shall see, biological systems are indeed robust on all levels of organization. Proteins can tolerate thousands of amino acid changes, metabolic networks continue to sustain life even after removal of important chemical reactions, gene regulation networks continue to function after alteration of key gene interactions, and radical transformations in embryonic development can lead to an essentially unchanged adult organism.

A second reason to study robustness (an evolutionary biologist's reason) derives from the fact that evolution by natural selection requires variation among organisms that reflects genetic variation. Genetic variation is abundant in most species, yet how it translates into phenotypic variation is still unknown. In the second part of the 20th century, a debate about precisely this question dominated evolutionary biology. This debate focused on the role and abundance of neutral mutations, mutations that do not affect the function of a biological system. The more neutral mutations a biological system allows, the greater is its mutational robustness, and mutational robustness thus has an important role to play in this debate. Mutational robustness influences the extent to which genetic variation, the result of past mutations, is translated into phenotypic variation. Even more importantly, if mutational robustness itself is subject to evolutionary change, then the ability to evolve by natural selection evolves, and thus evolvability evolves. For this and other reasons, neutral mutations will play a central role in this book. I will argue that they may play a very important role in promoting evolutionary innovation.

The third reason to study robustness regards engineering principles of robust systems. Is robustness in the living fundamentally similar and different from robustness in engineered systems? Can human engineers learn from robustness in the living? Only an engineer could be the judge, but the many examples scattered throughout the book may help in making this judgment. Although the book is primarily directed toward biological systems, I devote one short chapter to robustness in engineered systems.

How to Study Robustness

Empirical evidence for robustness comes in two different forms. First, one can perturb a part of an organism (a protein), a trait (wing shape), or a capability (amino acid biosynthesis) through mutations. The less the feature's properties change in the face of perturbation, the more robust it is. The second type of evidence relies on naturally occurring perturbations, mutations that occurred in evolutionary history. That is, one can compare closely related species that have the same trait or capability, and examine whether they achieve it by different means. If so, this indicates robustness, because not only can the same feature be designed in different ways, these different ways originated in a recent common ancestor and are thus reachable from each other by mutation or recombination. As with most applications of the comparative method, the results of this second approach are more tentative than the results of systematic perturbations.

Neither kind of evidence is easy to produce. Many biological systems, from macromolecules to genetic networks, have large numbers of parts that can occur in many configurations. To assess their robustness systematically requires many perturbations and subsequent measurements of system properties. For instance, to explore only a few variants at each amino acid positions of a protein, one needs to generate thousands of mutant proteins and measure their activity. The evolutionary approach to study robustness suffers from a related problem. First, to compare different organisms is to analyze only a few end products of many possible paths evolution could have taken. Second, sometimes even that is infeasible. There are preciously few well-studied organisms for which any one biological process above the gene level is well characterized, because such characterization is time-consuming. For instance, it took thousands of man-years to elucidate the structure of the genetic network responsible for segmenting a fruit fly's body. It would be prohibitive to analyze the same network in many related species to determine how much its structure has changed while leaving its function intact.

In sum, the experimental evidence to assess robustness in biological systems is hard to come by. The problem is partly alleviated by modeling of such systems, using both analytical and computational methods. Quantitative models that are based on experimental information can provide accurate predictions about a system's robustness, even when systematic perturbations or evolutionary comparisons are difficult. Many of the case studies below involve a tight integration between experimental evidence and quantitative modeling. Some of the most intriguing questions, such as whether robustness itself can evolve, have been mostly addressed with computational models. The heavy reliance on modeling to understand biological robustness may change as more experimental data accumulates. However, because of the many difficulties of providing such data, quantitative modeling will always play an important role in understanding the robustness of biological systems.

An Emphasis on Mechanism

One can analyze biological systems, their robustness, and its evolution from two very different perspectives. The first of them, exemplified by biochemistry and molecular biology, emphasizes mechanistic understanding, dissection of systems and their parts. Most of this book emphasizes this mechanistic perspective. A second approach is represented by population genetics and, even more so, by quantitative genetics. These disciplines emphasize the statistical effects of genes on fitness rather than the roles of genes in a molecular machinery. Both disciplines provide important perspectives complementary to those of molecular biology (55, 104, 141, 185, 186, 228, 244, 274, 305, 404, 415, 448, 462, 472, 499, 519, 520, 562, 579, 582, 585, 591, 592, 610). Population genetics, for example, identifies the conditions—selection pressures, mutation rates, population sizes, etc.—under which robustness can evolve, which is completely outside the scope of molecular biology. I have included general population genetic insights into the evolution of robustness. Nonetheless, the book contains comparatively little material from population genetics and next to none from quantitative genetics. The main reason is the following.

Population genetics and quantitative genetics have been very successful partly because they have eliminated the mechanistic details of biological systems from their thinking. However, the elimination of such detail and the resulting phenomenological perspective on organisms come at a price: Evolutionary explanations built on a statistical understanding of gene effects may be difficult to interpret. Take a recent example from a growing literature on how to measure robustness with quantitative genetic methods (244). Suppose you had found that during the evolution of an organismal lineage B from some ancestral lineage A, the mutational robustness of some trait, say the length of a fly's wing, has apparently increased. That is, the trait shows less change in response to the same amount of "mutation pressure" in lineage B than in lineage A. Houle pointed out that such apparent differences in robustness among lineages and traits could be caused by differences in the genome target size of these traits (244). The genome target size is the number of genes contributing to a trait. In other words, a trait's robustness may appear increased merely because the number of genes contributing to it decreased. To estimate this genome target size with the methods of quantitative genetics is difficult, partly because many genes with very subtle statistical effects contribute to most traits. Because quantitative genetics has not yet resolved such fundamental problems, I chose to focus here on systems whose inner workings are understood to some extent.

Principles of Robustness

This book could not have been written 15 years ago, because much of the mechanistic information I emphasize here has accumulated only recently. One consequence of this fact is that this field of research is not mature. It is rife with open questions, yes, dominated by open questions, questions that define entire research programs in systems biology. (I summarize some of these questions in the short epilogue.) This observation points to two motivations to write this book now. First, a survey of our knowledge brings our ignorance into sharp relief. Second, the available pieces of the puzzle enable us to see the outline of the whole, and allow us to make some general statements about it. Beyond the presentation of the evidence, you will thus find many informed guesses at the shape of the whole here. Whether or not there will be a unified theory of robustness in biological systems, some unifying principles will emerge once this field has reached maturity. Here is a brief summary of a few such principles (my credo, if you will), principles that later chapters elaborate in much greater detail and with concrete examples.

Most problems the living have solved have an astronomical number of equivalent solutions, which can be thought of as existing in a vast neutral space (chapter 13). A neutral space is a collection of equivalent solutions to the same biological problem. Such solutions are embodied in biological systems that ensure an organism's survival and reproduction. Both direct perturbation studies and indirect comparative studies support the notion that problems with many solutions are the rule rather than the exception. This holds on multiple levels of biological organization. We see it, for example, in the structure of important macromolecules such as proteins and RNA, where there are astronomically many different ways to build a molecule with a given structure and function. We see it also in the architecture of transcriptional regulatory regions, which can change drastically in evolution without any change in function. We see it in the structure of metabolic and genetic networks, where large changes in network structure can have negligible effects on network function in any one environment. We even see hints of it at the highest level of organismal organization, where radically different pathways of embryonic development may lead to essentially unchanged adult organisms.

Biological systems are mutationally robust for two reasons. First, robust systems are easier to find in the blindly groping search of biological evolution, simply because of the large neutral space associated with them (chapter 13). In other words, robust systems are systems with a large associated neutral space of equivalent solutions to a given problem. Such systems are easiest to discover in evolution, because they represent a large proportion of all possible solution. Their robustness results from the structure of neutral spaces itself, and may be independent of the particular circumstances under which an organism or the system evolved, such as population sizes or mutation rates. *Second, natural selection can further increase robustness by incremental evolution of a system within a neutral space (chapters 13, 16, 17).* Neutral spaces are not homogeneous. We know this from studies of the neutral spaces associated with the structure of biological macromolecules, and to a more limited extent from studies of genetic networks and the genetic code. This means that neutral spaces often have regions characterized by greater robustness, where mutations are less likely to change a system's structure or function, and regions of lesser robustness. Regions of lesser robustness are more sparsely populated with systems that perform a given function. Evolution by natural selection can drive an evolving population toward regions of a neutral space with high robustness.

Either mutations or nongenetic change can drive incremental evolution of mutational robustness (chapters 16, 17). It is at first sight obvious that robustness to mutations could be an adaptation to mutations. However, mutations are rare in most organisms. This implies, as I argue in chapter 16, that the conditions under which mutations can cause an increase in robustness are very restrictive. They require large populations or high mutation rates. Systems robust to mutations, however, are also robust to nongenetic change. Thus, mutational robustness can emerge as a by-product of selection for robustness to nongenetic change. This second mechanism for incremental evolution of robustness is much less restrictive, because organisms are constantly exposed to a barrage of nongenetic change.

Both of these explanations rely only on individual-based selection, and not on group, lineage, or species selection. That is, robustness need not be advantageous to a group of cells or organisms to increase in evolution. The main reason to emphasize individual-based selection is not so much that group selection is controversial and that it may occur only under limited conditions. Rather, almost all features of organisms that are hard to explain otherwise—among them altruism, sex, and evolvability itself—are easy to explain using group selection. The real challenge is to explain the evolution of robustness and evolvability through individual-based selection, which we know is ubiquitous

Robustness and neutral mutations are key to evolutionary innovation (chapter 14). Robust biological systems permit many neutral mutations, mutations that do not affect a specific system function. However, these mutations can affect other properties of the system, properties that may be the source of future detriment or benefit, and also the source of evolutionary innovations. Much like there are few mutations that will affect the phenotype under all circumstances—in all environments or genetic backgrounds—there are no mutations that are neutral under all circumstances. As I argue in chapter 14, if we can abandon an essentialist concept of neutrality—once neutral, always neutral—the concept of neutrality will continue to be useful and provide insight into the mechanics of innovation.

Redundancy of a system's parts is a minor mechanistic cause of robustness to mutation. More important is distributed robustness (chapter 15). In distributed robustness, interactions of multiple system parts, each with a different role, can compensate for the effects of mutations. I use the word redundancy here only for two or more system parts that perform the same or similar tasks. Perhaps the best example is gene redundancy. Gene redundancy occurs if one gene has several copies in a genome. Such redundancy can render an organism robust to mutations in one of these copies. I argue in chapter 15 that redundancy plays a role in mutational robustness, but not the predominant role. This holds for systems whose parts are genes, but also for systems on other levels of biological organization, such as biological macromolecules, whose parts are nucleotides or amino acids.

Fragility in a biological system, the opposite of robustness, can have several evolutionary causes (chapter 18). By fragility I mean that a system varies greatly in either structure or function in response to mutations. The first possible cause is that the biological problem to which the system is a solution has only a few alternative solutions and thus a small associated neutral space. Second, variation in the system may be advantageous to the organism. A paradigmatic example of advantageous variation is antibody diversity in the vertebrate immune system. The third possibility is that trade-offs with other aspects of the system's function preclude maximal robustness. Potential examples include enzymes: To catalyze chemical reactions, enzymes need to change their tertiary structure, their folded three-dimensional structure, in subtle ways. Such flexibility is not possible if this tertiary structure is maximally robust to mutations, as I discuss in chapter 5.

Many natural systems below and beyond living organisms show great robustness to changes in their parts. Such robustness can also increase over time, but the cause is usually self-organization instead of natural selection (chapter 19). I will illustrate this principle with one main example, the robustness of ecological communities to species invasions. (Ecological communities, although composed of living things, are strictly speaking not themselves living things.) The example is my only excursion into ecology, where the subject of robustness to various perturbations has been of long-standing interest (362), and has spawned a bewildering array of terminology and different criteria for robustness (212). *Many of the mechanistic principles that underlie robustness in living systems can also be observed in man-made, engineered systems (chapter 20).* I will illustrate this notion with some anecdotal examples from areas such as telecommunications and electrical engineering. A more exhaustive comparison would itself merit a book.

A Word on the History

It has been said that nothing is ever new. To the extent that this is true, it also holds for these and other ideas I emphasize here. Some of them have been expressed previously, in various degrees of clarity, within a restricted field of investigation or with an emphasis different from that on evolution I take here (130, 183, 201, 491). I have become aware of some previous work only late during this writing, and may be completely unaware of other work. Germs of these concepts may go back many decades and may be buried, unbeknownst to me, in one or the other parenthetical observation of many research papers. However, there are two main historic threads of research into mutational robustness. Both go back to the first half of the 20th century. The first of them regards the phenomenon of *dominance*. Dominance means that a phenotypic feature of an organism is robust to elimination of one among two copies of a gene product or to a corresponding 50% change in the concentration of a gene product. The phenomenon has been known for most of the 20th century, and its discovery can be traced back to Gregor Mendel's experiments in the 19th century, which are cornerstones of classical genetics (364). A major protagonist in the history of research on dominance is Ronald Fisher (155, 156), who proposed the first evolutionary explanation of dominance in the 1930s. I discuss this explanation in chapter 8, as well as some reasons why most evolutionary biologists consider it no longer viable.

A second important phenomenon and early line of investigation is that of *canalization*. An organismal feature is canalized if its embryonic development is insensitive to variation in the environment or in genes. The term canalization originated with the embryologist Conrad Waddington (575, 578). He and others studied canalization by trying to disrupt it in organisms such as fruit flies and in mice through environmental stressors or specific mutations. The result of disrupting canalization is a drastic increase in variation that is caused by previously hidden genetic variation. Canalization is thus a specific aspect of robustness in organismal traits. I revisit this phenomenon and some of its history in chapter 11.

Research on dominance and canalization constitute the two main lines of investigation into mutational robustness. However, I note in passing that the broader phenomenon of robustness to any change, whether genetic or nongenetic, has a much longer history. A case in point is the concept of homeostasis, an organism's ability to sustain a physiological state in the face of change. It was coined in 1932 by the American physiologist Walter Cannon. However, its roots can be traced back to the French physiologist Claude Bernard and his book *Introduction to the Study of Experimental Medicine* (44). In essence, Bernard argued there that the constancy of the milieu inside an organism results from regulatory mechanisms inside the body.

Functions and Purpose, Problems and Solutions

As the preceding pages show, I will heavily use functional language in this book. That is, I will speak of biological systems as serving specific functions or purposes inside an organism. In other words, such systems solve problems that organisms face in reproducing and surviving. For instance, enzymes are systems that solve the problem of converting chemical compounds into useful forms; gene circuits in development solve the problem of reliably patterning an embryo to produce a viable adult; the genetic code solves the problem of translating genetic information into a protein's amino acid sequence; and so on.

Such language raises thorny philosophical problems if taken literally (470, 631). Part of the reason is that words like "function" and "problem" insinuate an intelligent agent standing behind a system's design. However, for all we know, the biological systems I examine here emerged from the blindly groping search that characterizes all of biological evolution. That they embody solutions to important biological problems is obvious only in hindsight, after the systems that embody these solutions survive. It should be understood that functional language merely provides a convenient and compact way to describe the endpoint of the convoluted paths evolution takes.

Who Is This Book for?

If you are a specialist who already knows some or most of the literature in this field, much of this book will not be news to you. However, if you are a nonspecialist interested in the questions I pose, this book may be for you. The book presupposes some knowledge of biological principles, particularly genetics and biochemistry, on the level of an introductory course in both subjects. Most of the chapters of parts I and II survey our knowledge in a representative range of examples, and the relevant background material is contained within each chapter. Only a few chapters contain mathematical material, which requires some basic understanding of linear algebra, differential calculus, probability theory, and differential equations. However, even where I found some mathematical treatment necessary, such as in the chapters on metabolism, I took pains to describe the central concepts verbally as well.

The Organization of This Book

The book consists of four parts. The chapters of parts I and II take you on a tour through examples of genetic robustness. Each chapter addresses a different aspect of the cardinal question "what is robust?" In other words, each chapter examines the robustness of a different feature of an organism. The sequence of chapters is a tour through the hierarchy of biological organization, from molecules to whole organism. Some of the central ideas I just mentioned make an appearance in these chapters, but I defer their detailed discussion to part III.

Chapter 2 examines the robustness of the genetic alphabet to replication errors. It is one of the most speculative chapters, because it examines evidence for several possible DNA and RNA chemistries, among which only one is realized in life as we know it. Chapter 3 discusses a large body of work on the robustness of the genetic code to nucleotide changes in individual codons. Chapter 4 examines the robustness of RNA structure to nucleotide changes, and chapter 5 does the same for protein structure and amino acid changes. Chapter 6 focuses on the robustness of proteins to a different kind of genetic change, recombination. It also briefly discusses a phenomenon related to recombination, lateral gene transfer, which has profoundly influenced microbial evolution.

The subsequent chapters constitute part II and discuss higher levels of biological organization. Chapter 7 discusses how gene expression can be robust to changes in promoter organization. Chapters 8 and 9 survey the robustness of metabolic pathways and metabolic networks to changes in enzyme activity. Chapter 10 examines the robustness of genetic networks in embryonic development to changes in regulatory gene interactions. Chapter 11 focuses on the organismal level and on phenotypic characters, such as the wings and eyes of insects. It surveys evidence that the embryonic development of such characters is highly robust to mutations, and how this robustness can be disrupted in laboratory experiments. In addition, the chapter also discusses the breakdown of robustness in genetic diseases, a phenomenon illustrating the importance of robustness for medicine. Chapter 12 uses examples from the evolution of three very different organisms, nematode worms, sea urchins, and parasitic wasps, to illustrate how the formation of whole body plans can be robust to massive developmental changes, which are ultimately caused by genetic change.

The chapters of parts I and II contain a moderate number of illustrative, hand-picked examples. I chose these examples either because they are especially well understood or because they illustrate a general principle well. This means that the material I discuss is representative rather than comprehensive. It contains obvious omissions that I made for one or the other reason. One such example regards neural circuits and the computations they perform. Theoretical work shows that neural circuits can be highly robust to removal of neurons and to removal of interactions between neurons (19). However, we know little about the robustness of biological neural circuits—as opposed to abstract models of such circuits—to mutations.

Part III uses many of the examples in parts I and II as raw material to discuss general principles behind robustness and its evolution. The first chapter in this part, chapter 13, focuses on one of the key concepts emerging from parts I and II, the concept of a neutral space. Chapter 14 explores the relation of robustness to future evolutionary potential—evolvability. It emphasizes the positive role that robustness and neutral mutations play for evolvability. Chapter 15 focuses on the phenomenon of redundancy. The chapter discusses empirical evidence that distributed robustness and not redundancy of parts may be the predominant mechanistic cause of robustness in biological systems. Chapters 16 and 17 focus on the evolution of mutational robustness. Specifically, chapter 16 discusses how robustness can evolve as an adaptation to mutational pressure. Chapter 17 shows that mutational robustness can emerge as a by-product of selection for robustness to nongenetic change. Chapter 18 focuses on systems that are not robust but fragile. It discusses various evolutionary causes of such fragility.

Part IV is an elaborate afterthought that relates mutational robustness in the living to robustness in other systems. Chapter 19 shows how robustness in natural inanimate systems can change over time. It does so through a mechanism fundamentally different from natural selection, a mechanism involving self-organization. Chapter 20, finally, provides a nexus between robustness in the living and in nonliving engineered systems. It highlights some similarities between these kinds of systems, such as the existence of distributed robustness. The book closes with an epilogue that states important open questions about robustness and its evolution.