8

Human Civilization I:
Defining Complexity

Conceptual Outline

8.1 Our ultimate objective is to consider the relationship of a human being to human civilization, where human civilization is considered as a complex system. We use this problem to motivate our study of the definition of complexity.

8.2 The mathematical definition of the complexity of character strings follows from information theory. This theory is generalized by algorithmic complexity to allow all possible algorithms that can compress the strings. The complexity of a string is defined as the length of the shortest binary input to a universal Turing machine, such that the output is the string.

8.3 The use of mappings from strings onto system states allows us to apply the concepts of algorithmic complexity to physical systems. However, the complexity of describing a microstate of the system is not really what we mean by system complexity. We define and study the complexity profile, which is the complexity of a system observed with a certain precision in space and time.

8.4 We estimate the complexity of various systems, focusing on the complexity of a human being. Our final estimate is based upon a combination of the length of descriptions in human language, genetic information in DNA, and component counting.

8.1 Motivation

8.1.1 *Human civilization as a complex system*

The subject of this and the next chapter is human civilization—the collection of all human beings on earth. Our long-term objective is to understand whether and how we can treat human civilization as a complex system and,more particularly, as a complex organism. In biology, collections of interacting biological organisms acting together are called superorganisms. At times, we will adopt this convention and refer to civilization as the human superorganism. Much of what we discuss is in early stages of development and is designed to promote further research.

This subject is distinct from the others we have considered. The primary distinction is that we have only one example of human civilization. This is not true about the systems we have discussed in earlier chapters, with the exception of evolution considered globally. The uniqueness of the human superorganism presents us with questions of fundamental interest in science, related to how much we can know about an individual system. When there are many instances, we can use information provided by various examples and the statistics of their properties. When there is only one system, to understand its properties or predict its behavior we must apply fundamental principles that are valid for all complex systems. Since the field of complex systems is dedicated to uncovering such principles, the subject of the human superorganism should be considered a premiere area for application of complex systems research. Central questions are: How can we characterize this complex system? How can we determine its properties? What can we tell about its dynamics—its past and future? We note that as individuals we are elements of the human superorganism, thus our spatial and temporal experience may very well be more limited than that appropriate for analyzing the human superorganism.

The study of human civilization is guided by historical records and contemporary news. In contrast to protein folding, neural networks, evolution and developmental biology there are few reproducible laboratory experiments. Because of the irreproducibility of historical or contemporary events,these sources of information are properly not considered part of conventional science. While this can be a limitation, it is also apparent that there is a large amount of information available.Our task is to develop systematic methods for considering this kind of information that will enable us to approach questions about the nature of human civilization as a complex system. Various aspects of these problems have been studied by historians, anthropologists and sociologists.

Why consider human civilization as a single complex system? The recently discussed concept of a global economy, and earlier the concept of a global village, suggest that we should consider the collective economic behavior of human beings and possibly the global social behavior as a single system. Considering civilization as a single entity we are motivated to ask various questions about it. These questions relate to all of the topics we have covered in the earlier chapters: spatial and temporal structure, evolution and development. We would also like to understand the interaction of human civilization with its environment.

In developing an understanding of human civilization, we recognize that a widespread view of human civilization as a single entity is relatively new and driven by contemporary developments. At least superficially, the historical epoch described by the dominance of nation-states appears to be quite different from the present global economy. While recent events appear to be of particular significance to the global view, our questions must be addressed in a historical context. Thus we should include a discussion of the transition to a global economy. We postpone this historical discussion to the next chapter because of the groundwork that we would like to build in order to target a particular objective for our analysis—that of complexity classification.

We are motivated to understand complexity in the context of our effort to understand the nature of the human superorganism, or the nature of the global economy. We would like to identify the type of complex system it is—to classify it. The first distinction that we might make is between a complex material or a complex organism (see Section 1.3.6). Could part of the global system be modified without affecting the whole? From historical evidence discussed in the next chapter, the answer appears to be no. This indicates that human civilization is a complex organism. The next question we would like to ask is: What kind of complex organism is it? By analogy we could ask: Is it like a protein, a cell, a plant, an insect, a frog, a human being? What do we mean by using such analogies? At least in part the problem is to describe the complexity of an entity's behavior. Intuitively an insect is a simpler organism than a human being, and this is of qualitative importance for our understanding of their differences. The degree of complexity should provide a scale that can distinguish between the many different complex systems we are familiar with.

Our objective in this chapter is to develop a quantitative definition of complexity and behavioral complexity. We then apply the definition to various complex systems. The focus will be on the complexity of an individual human being. Once we have established our complexity scale we will be in a position to apply it to human civilization. We will understand formally why a collection of complex systems (human beings) may be, but need not be, complex. Beyond recognizing human civilization as a complex system,it is far more significant to identify the degree of its complexity. In the following brief sections we establish some additional context for the importance of measuring complexity using both unconventional and conventional examples of organisms whose complexity should be evaluated.

8.1.2 *Scenario: alien encounter*

The possibility of encountering alien life has been debated within the scientific community. In popular literature, such encounters have been portrayed in various forms ranging from benevolent to catastrophic. The scientific debate has focused thus far on topics such as the statistics of planet formation and the likelihood that planets contain life. The presence of organic molecules in meteorites and interstellar gasses has been interpreted as suggesting that alien life is likely to exist.Efforts have been made to listen for signs of alien life in radio communications and to transmit information to aliens using the Voyager spacecraft, which is leaving the solar system marked with information about human beings. Thus far there has been no scientifically confirmed evidence for the existence of alien life. Even a single encounter would change the human perspective on humanity's place in the universe.

Let us consider one possible scenario for an encounter. An object that flashes light intermittently is found in orbit around one of the planets of the solar system. The humans encountering this object are faced with the question of determining whether the object is: (a) a signal device—specifically a recording, (b) a communication device, or (c) a living organism. The central problem can be seen to revolve around determining whether, and in what way, the device is responsive to external phenomena. Do the flashes of light occur without regard to the external environment

in a predetermined sequence? Are they random? If the flashes are sensitive to the environment,then what are they sensitive to? We will see that these questions are equivalent to the question of determining the complexity of the object's behavior.

The concept of life in biology is often defined, or better yet, characterized, in terms of consumption, excretion and reproduction. As a definition, these characteristics are well known to be incomplete, since there are life-forms that do not reproduce, such as the mule. Furthermore, a particular individual is still considered alive even if it/he/she does not reproduce. Moreover, there are various physical systems such as crystals and fire that have all these characteristics in one form or another. Moreover, there does not appear to be a direct connection between these biological characteristics and other characteristics of life such as sentience and self-awareness. When considering behavior, the biological perspective emphasizes the survival instinct as characteristic of life. There are exceptions to this,since there exist life-forms that are at times suicidal, either individually or collectively. The question of whether an organism actively seeks life or death does not appear to be a characterization of life but rather of life-forms that are likely to survive. In our discussions, we may be developing an additional characterization of life in terms of behavioral complexity. Definitions of life are often considered in speculating about the rights of and treatment of real or imagined organisms—injured or unconscious humans, robots, or aliens. The degree of behavioral complexity is a characterization of life-forms that may ultimately play a role in informing our ethical decisions with respect to various biological life-forms, whether terrestrial or (if found) alien, and artificial life-forms that we create.

8.1.3 *Scenario: blood cells*

One of the areas briefly touched upon in Chapter 6, which is at the forefront of complex systems research, is the study of the immune system. Blood cells,unlike other cells in the body, are mobile on a length scale that is large compared to their size. In this characteristic they are more similar to independent organisms than to the other cells of the body. By their migration they might be said to "choose" to associate with other cells of the body, or with foreign chemicals and cells. It is fair to say that our understanding of the behavior of immune cells remains primitive. In particular, the variety of possible chemical interactions between cells has only begun to be mapped out. These interactions involve a variety of chemical messengers. More direct cell-to-cell interactions where parts of the membrane or cellular fluid are transferred are also possible.

One of the interesting questions that can be asked is whether, or at what level of complexity, the interactions become identifiable as a form of language. It is not difficult to imagine, for example, that a chemical communication originating from one cell might be transferred through a chain of cell interactions to a number of other cells. In the context of the discussion in Section 2.4.5, the question of existence of a language might be formulated as a question about the possibility of messages with a grammar—a combinatorial composition of parts that are categorized like parts of speech. Such combinatorial mechanisms are known to exist even at the molecular level in the DNA coding of antibody receptors that are a composite of different parts

of the genome. It remains to be seen whether intercellular communication is also generated in this fashion.

In the context of this chapter we can reduce the questions about the immune cells to a single one—What is the degree of complexity of the behavior of the immune cells? By its very nature this question can only be answered once a complete understanding of immune cell behavior is reached. A limited understanding establishes a lower bound for the complexity of the behavior. It should also be understood that different types of cells will most likely have quite different levels of behavioral complexity, just as different animals and man have differing levels of complexity. Our objective in this chapter is to show that it is possible to quantify the concept of complexity in a way that is both natural and useful. The practical application of these definitions is a central challenge for the field of complex systems.

8.1.4 *Complexity*

Mathematical definitions of the complexity of systems are based upon the theories of information and computation discussed in Sections 1.8 and 1.9. In Section 8.2 they will be used to treat complexity in the context of mathematical objects such as character strings. To develop our understanding of the complexity of physical systems requires that we relate these concepts to those of thermodynamics (Section 1.3) and various extensions (e.g., Section 1.4) that enable the treatment of nonequilibrium systems. In Section 8.3 we discuss relevant concepts and tools that may be used for this purpose. In Section 8.4 we use several semiquantitative approaches to estimate the value of the complexity of specific systems.

Our use of the word "complexity" is specified as an answer to the question, How complex is it? We say, Its complexity is <number><units>. Intuitively, we can make a connection between complexity and understanding. When we encounter something new, whether personally or in a scientific context, our objective is to understand it. The understanding enables us to use, modify, control or appreciate it. We achieve understanding in a number of ways, through classification, description and ultimately through the ability to predict behavior. Complexity is a measure of the inherent difficulty to achieve the desired understanding. Simply stated, *the complexity of a system is the amount of information necessary to describe it.*

This is descriptive complexity. For dynamic systems the description includes the changes in the system over time. We will also discuss the response of a dynamic system to its environment. The amount of information necessary to describe this response is a system's behavioral complexity. To use these definitions of complexity we will introduce mathematical expressions based upon the theory of information.

The quantitative definition of information (Section 1.8) is relatively abstract. However, it can be measured in familiar terms such as by the number of characters in a text. As a preliminary exercise in the discussion of complexity, the reader is invited to exercise intuition to estimate the complexity of a number of systems. Question 8.1.1 includes a list of systems that are designed to stimulate some thought about complexity as a quantitative measure of the behavior of a system. The reader should devote some thought to this question before proceeding with the rest of the text.

Question 8.1.1 Estimate the complexity of some of the systems in the following list. For this question use an intuitive definition of complexity—the amount of information that would be required to describe the system or its behavior. We use units of bits to measure information. However, to make it easier to visualize, you may use other convenient units such as words or pages of text. So, we can paraphrase the question as, How much would you have to write to describe the system behavior? A rough conversion factor of 1 bit per character can be used to convert these estimates to bits. It is not necessary to estimate the complexity of all the systems on the list. Considering even a few of them is sufficient to develop an understanding of some of the issues that arise. Indeed, for some of these systems a rough estimate is far from trivial. Answers to this question will be given in the text in the remainder of this chapter.

Hint You may find that you would use different amounts of information depending on what aspects of the system you are describing. In such cases try to give more than one estimate or a range of values.

Physical Systems:

Ideal gas (1 mole at T = 0°K, P = 1atm)
Water in a glass
Chemical reaction
Brownian particle
Turbulent flow
Protein
Virus
Bacterium
Immune system cell
Fish
Frog
Ant
Rabbit
Cow
Human being
Radio
Car
IBM 360
Personal Computer (PC/Macintosh)
The papers on your desk
A book

A library
Weather
The biosphere
Nature

Mathematical and Model Systems:

A number
Iterative maps (growth, bifurcation to chaos)
1-D random walk
    short time
    long time
Ising model (ferromagnet)
Turing machine
Fractals
    Sierpinski gasket
    3-D random walk
Attractor neural network
Feedforward neural network
Subdivided attractor neural network ∎

8.2 Complexity of Mathematical Models

Complexity is a property of the relationship between a system and various representations of the system.Our objective is to understand the complexity of systems composed of physical entities such as atoms,molecules or cells. Abstract representations of such systems are described in terms of characters or numbers. It is helpful to preface our discussion of physical systems with a discussion of the complexity of the characters or numbers that we use to represent them.

8.2.1 *Information, computation and algorithmic complexity*

The discussion of Shannon information theory in Section 1.8 was based on strings of characters that were generated by a source. The source generates each string, s, by selecting it from an ensemble. The information from a particular string was defined as

$$I = -\log(P(s)) \tag{8.2.1}$$

where $P(s)$ is the probability of the string in the ensemble. If all strings have equal probability then this is the logarithm of the number of distinct strings. The source itself (or the ensemble) was characterized by the average information of a large number of strings

$$< I > = -\sum_s P(s)\log(P(s)) \tag{8.2.2}$$

It was also possible to consider a more general source that selected characters to form a Markov chain. The probabilistic coupling between sequential characters reduced the information content of the string. It was possible to compress the string using a reversible coding algorithm (computation) that would enable the same information to be represented in a more compact form. The length of the shortest binary compact form is equal to the average information in a string.

Information theory suggests that we can define the complexity of a string of characters by the information content of the string. The information content is the same as the length of the shortest binary encoding of the string. This is intuitive—since the original string can be obtained from its shortest representation, the same information must be present in both. Within standard information theory, the encodings would be limited to compression using a Markov chain model. However, more generally, we could use any possible algorithm for encoding (compressing) the string. Questions about all possible algorithms are precisely the domain of computation theory. The definition of Kolmogorov (algorithmic) complexity of a string makes use of computation theory to describe what we mean by "any possible algorithm." Allowing all algorithms is the same as allowing more general models for the string than a Markov chain. Our objective in this section is to develop an understanding of algorithmic complexity beginning from the theory of computation.

Computation theory (Section 1.9) describes the operations of logic and computation on symbols. All the operations are deterministic and are expressible in terms of a few elementary operations. The concept of universality of computation is based on the understanding that a particular type of conceptual machine/computer—the universal Turing machine (UTM)—can perform all possible computations if the instructions are properly encoded as a finite string of characters serving as the UTM input. Since we have no absolute definition of computation, there is no complete proof. The existing proof shows that the UTM can perform all computations that can be done by a much larger class of machines—the Turing machines (TM). Other models for computation have been shown to be essentially equivalent to these TM. A TM is defined by a table of elementary operations that act on the input string. The word "program" can be used either to refer to the TM table or to its input and so its use is best avoided in this context.

We would like to define the algorithmic complexity of a string, s, as the length of the shortest possible binary TM input, such that the output is s. The relationship of this to the encoding and decoding of Shannon should be apparent. In order to use this as a definition, there are several matters that must be cleared up. To summarize: There are actually two sources of information when we use a TM, the input string and the table. We need to take both of them into account to define the complexity. There are many ways to define complexity; however, we can prove that any two definitions of complexity differ by no more than a constant. We will also show that no matter what definition we use, most strings cannot be compressed.

In order to motivate the logic of the following discussion, it is helpful to think about how we might approach compressing various strings of characters. The short-

est compression should then be the complexity of the string. One string might be formed out of a long substring of zeros followed by a long substring of ones. This is convenient to write by indicating how many zeros followed by how many ones: N_0N_1. We would make a binary string notation for N_0N_1 and write a program that would read this input and then output the original string. Another string might be a representation of the Fibonacci numbers (1,1,2,3,5,8,…), starting from the N_0st number and ending at the N_1st number. We could write this using a similar notation as the previous one, but the program that we would write to generate the string is quite different. Both programs would be quite simple. Now imagine that we want to communicate one of the original strings to someone else. If we want to communicate it in compressed form, we would have to send the program as well as the input. If there were many strings, we might be clever and send the programs only once. The problem is that with only the input string, the recipient would not know which program to apply to obtain the original string. We need to send an additional piece of information that indicates which program to apply. The simplest way to do this is to assign numbers to each of the programs and preface the program input with the program number. Once we do this, the string that we send uniquely determines the string we wish to communicate. This is necessary, because if the interpretation of the transmitted string is not unique,then it would be impossible to guarantee a correct interpretation. We now develop these thoughts using a more formal notation.

In what follows, the operation of a TM or a UTM will be indicated by functional notation. The string that results from its application to a tape is indicated by $U(s)$ where s is the nonblank portion of the tape (input string), U is the identifier of the TM,and the initial position of the TM head is assumed to be at the leftmost nonblank character.

In order to define the complexity of a string, we identify a particular UTM U. Then the complexity $C_U(s)$ of the string s is defined as the length of the shortest string r such that $U(r) = s$. We call an input string r to U that generates s a representation of s. Thus the length of the shortest representation is $C_U(s)$. The central theorem of algorithmic complexity relates the complexity according to one UTM U and another UTM U'. Before we state and prove the theorem, we discuss several incidental matters.

We first ask whether we need to use a UTM and not just any TM in the definition. The answer is that the use of a UTM is convenient,and we cannot significantly improve the ability to compress strings by allowing the larger class of TM to be used in the definition. Let us say that we have a UTM U and a TM V, we define a new UTM W by:

$$\begin{aligned} W(0s) &= V(s) \\ W(1s) &= U(s) \end{aligned} \tag{8.2.3}$$

—the first character indicates whether to use the TM V or the UTM U on the rest of the input. Since the complexity according to the UTM W is at most one more than the

complexity according to the TM V, $C_W(s) \leq C_V(s) + 1$, we see that using the larger class of TM to define complexities can not improve our results for any particular string by more than one bit, which is not significant for long complex strings.

We may be disturbed that the definition of complexity does not indicate that the complexity of an incompressible string is the same as the string length itself. Indeed the definition does not require it. However, if we wanted to impose this as an auxiliary condition, we could define the complexity of a string using a slightly different construction. Given a UTM U, we define a new UTM V such that

$$\begin{aligned} V(0s') &= s' \\ V(1s') &= U(s') \end{aligned} \tag{8.2.4}$$

—the first character indicates whether the string is compressed. We then define the complexity $C_U(s)$ of any string s as one less than the length of the shortest string r such that $V(r) = s$. This is not quite a fair definition, because if we wanted to communicate the string s we would have to indicate all of r, including its first bit. This means that we should define the complexity as the length of r, which would be a sacrifice of at most one bit for incompressible strings. Limiting the complexity of a string to be no longer than the string itself might seem a natural idea. However, we note that the Shannon information, Eq. (8.2.1), is related only to the probability of a string, and may be larger than the original string length for a particular string.

Returning to our basic definition of complexity, we have described the existence of a shortest possible representation of any string s, and a single machine U that can reconstruct each s from this representation. The key theorem that we need to prove relates the complexity defined using one UTM U to the complexity defined using another UTM U'. The theorem is: the complexity C_U based on U and the complexity $C_{U'}$ based on U' satisfy:

$$C_U(s) \leq C_{U'}(s) + C_U(U') \tag{8.2.5}$$

where $C_U(U')$ is independent of the string s. The proof of this expression results from the ability of the UTM U to simulate U'. To prove this we must improve slightly our definition of complexity, or equivalently, we have to limit the UTM that are allowed. This is discussed in Questions 8.2.1–8.2.3. It is shown there that we can preface binary strings input to the UTM U' with a prefix that will make them generate the same output when input to U. We might call this prefix $r_{U,U'}$ a translation program,it satisfies the property that for any string r, $U(r_{U,U'}r) = U'(r)$. Let $r_{U'}$ be a minimal representation for U' of the string s. Then $r_{U,U'}r_{U'}$ is a representation for U of the string s. The length of this string must be greater than or equal to the length of the minimum string r_U necessary to produce the same output:

$$C_U(s) = |r_U| \leq |r_{U,U'}r_{U'}| = |r_{U'}| + |r_{U,U'}| = C_{U'}(s) + C_U(U') \tag{8.2.6}$$

$C_U(U') = |r_{U,U'}|$ is the length of the translation program. We have proven the inequality in Eq. (8.2.5).

Question 8.2.1 Show that there exists a UTM U_0 such that for any TM U that accepts binary input, there is a string r_U so that for all s and r satisfying $s = U(r)$, we have that $s = U_0(r_U r)$.

Hint One way to do this is to use a modified form of the construction given in Section 1.9. The new construction requires modifying the nature of the UTM—i.e., a trick.

Solution 8.2.1 We call the UTM described in Section 1.9, $\tilde{U}_0$. We can simulate the UTM U using $\tilde{U}_0$; however, the form of the input string would not quite satisfy the conditions of this theorem. $\tilde{U}_0$ has an input that looks like $r_U r_t(r)$, where the right part is only a function of the input string r and the left part is only a function of the UTM U. However, the tape part of the representation $r_t(r)$ uses a doubled binary form for characters and markers between them so that it is not the same as the original tape. We must replace the tape part of the representation with the original string in order to have an input string of the form $r_U r$.

Both $\tilde{U}_0$ and U have binary input strings. This means that we might try to use the tape of U without modification in the tape part of the representation given in Section 1.9. Then there would be no delimiters between characters and no doubled binary representation. There is, however, one difficulty. The UTM U_0 must keep track of where the current position of the UTM U would be during the same calculation. This was accomplished in Section 1.9 by converting one of the M_1 markers to M_6 at the current location of the UTM U. There are a number of ways to overcome this problem, but all require us to introduce something new. We will do this by allowing the UTM U_0 to have a counter that can keep track of the current position of the UTM U. There are two ways to argue this.One is to allow, by proclamation, a counter that can reach arbitrarily high numbers. The other is to recognize that the longest string we might conceivably encounter is smaller than the number of particles in the known universe, or very roughly $10^{90} = 2^{300}$. This means that we can use an internal memory of 300 bits to represent such a counter. This counter is initialized to 0 and set to the current location of the UTM U at every step of the calculation. This construction gives us the desired UTM U_0. ∎

Question 8.2.2 Using the result of Question 8.2.1, prove Eq.(8.2.5). See the text for a hint.

Solution 8.2.2 The problem is that Eq.(8.2.5) is not actually correct for all UTM (see Question 8.2.3) so we need to modify our conditions. In a sense, the modification is minor because we only improve the definition slightly. We do this by defining the complexity $C_U(s)$ for an arbitrary UTM as the minimum length of r such that $W(r) = s$ where W is defined by:

$$\begin{aligned} W(0s) &= U_0(s) \\ W(1s) &= U(s) \end{aligned} \tag{8.2.7}$$

—the first bit specifies whether to use U or the special UTM U_0 constructed in Question 8.2.1. $C_U(s)$ defined this way is at most one bit more than our previous definition, for any particular string. It might be significantly

smaller. This should not be a problem, because our objective is to find short representations of strings. By using our special UTM U_0 in this definition, we guarantee that for any two UTM U and U', whose complexity is defined in terms of W and W' by Eq.(8.2.7), we can write $W(r_{WW'}r_{W'}) = W'(r_W)$. This is possible because W inherits the properties of U_0 when the first character of its input string is 0. ∎

Question 8.2.3 Show that some form of qualification of Eq. (8.2.5) is necessary by demonstrating that there exists a UTM that does not satisfy this inequality. Therefore, Eq. (8.2.5) cannot be extended to all UTM.

Solution 8.2.3 One possibility is to have a UTM that uses only certain characters in its input string. Specifically, define a UTM U that acts the same as a UTM U' but uses only every other character in its input string: $U(r) = U'(r')$ if r is any string whose odd characters are the characters of r'. The complexity of a string according to U is twice the complexity according to U' and therefore Eq. (8.2.5) is invalid in this case. With the modified definition of complexity given in Question 8.2.2 this is no longer a problem. ∎

Switching U and U' in Eq. (8.2.5) gives a similar inequality with a constant $C_{U'}(U)$. Defining the larger of the two translation program lengths to be

$$C_{U,U'} = \max(C_U(U'),C_{U'}(\mathrm{U})) \tag{8.2.8}$$

we have proven that complexities defined by the UTM differ by no more than $C_{\mathrm{U},U'}$:

$$|C_U(s) - C_{U'}(s)| \le C_{U,U'} \tag{8.2.9}$$

Since this constant is independent of the complexity of the string s, it becomes insignificant for large enough complexities. Thus, for strings that are complex enough, it doesn't matter which UTM we use to define its complexity. The complexity defined by one UTM is the same as the complexity defined by another UTM. This consistency—universality—in the complexity of a string is essential in order for it to be well defined. We will use a few examples to illustrate the nature of universality provided by this definition.

The first example illustrates the relationship of algorithmic complexity to string compression. Given a string s we can ask what methods of compression are useful for the string. A useful compression algorithm corresponds to a pattern in the characters of the string. A string might have many repetitive digits, or cyclically repeating digits. Alternatively, it might be a sequence that can be generated using simple mathematical operations such as the Fibonacci series, or the digits of π. There are many such patterns that are relevant to the compression of strings. We can choose a finite set of N algorithms $\{V_i\}$, where each one is represented by a TM that reconstructs a string s from a shorter string r by taking advantage of properties of the pattern. We now construct a new TM U which is defined by:

$$U(r_i r') = V_i(r') \tag{8.2.10}$$

where r_i is a binary representation of the number i, having $\log(N)$ bits. This is a UTM if any of the V_i is a UTM or it can be made into a UTM by Eq. (8.2.3). We use U to define the complexity $C_U(s)$ of any string as described above. This complexity includes both the length of r' and the number of bits ($\log(N)$) in r_i that together constitute the length of the input r to U. Once it is defined,this complexity is a measure of the complexity of all strings. We do not use different TM to define the complexity of each string; one UTM is used to define the complexity of all strings.

Despite the message of the last example,let us assume that we are evaluating the complexity of a particular string s. We define a new UTM U_s by:

$$\begin{aligned} U_s(0s') &= s \\ U_s(1s') &= U(s) \end{aligned} \tag{8.2.11}$$

—the first character tells U_s if the string is s. We can use this new UTM to define the complexity of all strings and for this definition the complexity of s is one. How does this relate to our theorem about the universality of complexity? The point is that in this case the translation program between U and U_s contains the complete information about s and therefore must be at least as long as $C_U(s)$. What we have done is to take the particular string s and insert it into the table of U_s. We see in this example how universality is tied to an assumption that the complexities that are discussed are longer than the TM translation programs or, equivalently, the information in their tables. Conceptually, we would say that universality of complexity is tied to an assumption of lack of specific knowledge on the part of the recipient (represented by the UTM) of the information itself. The choice of a particular UTM might be dictated by an implicit understanding of the set of strings that we would like to represent, even though the complexity of a string is defined without reference to an ensemble of strings. However, this apparent relativism of the complexity is limited by our basic theorem that relates the complexity of distinct UTM,and by additional results about the impossibility of compressing most strings discussed in the following paragraphs.

We have gained an additional result from the construction of a single UTM that generates all strings from their compressed forms. This is that a representation r only represents one string s. We can now prove that the probability that a string of length N can be compressed is very small. The proof proceeds from the observation that the number of possible strings decreases very rapidly with decreasing string length. A string s of length $|s| = N$ compressed by k bits is represented by a particular string r of length $|r| = C(s) = N - k$. Since there are only 2^{N-k} strings of length $N - k$, at most 2^{N-k} strings of length 2^N can be compressed by k bits. The fractional compression is k/N. For example,among all strings of length 10^6 bits,at most 1 string in $2^{100} = 10^{30}$ can be compressed by 100 bits or .01% of the string length. This is not a very significant compression. Even so, this estimate of the average number of strings that can be compressed is much too large, because strings that are not of length N, e.g., strings of length $N - 1$ $N - 2$, …, $N - k$, would also be represented by strings of length $N - k$. Thus most strings are incompressible. Moreover, selecting a string at random will yield an incompressible string.

Question 8.2.4 Calculate a strict lower bound for the average complexity of strings of length N.

Solution 8.2.4 We assume that strings of length N are compressed so that they are represented by all of the shortest strings. One string is represented by the null string (length 0), two strings are represented by a single bit (length 1), and so on. The relationship:

$$2^N = \sum_{l=0}^{N-1} 2^l + 1 \tag{8.2.12}$$

means that we will fill all of the possible strings up to length $N-1$ and then have one string left of length N. The average representation length for any complexity measure must then satisfy:

$$< C(s) > \geq \frac{1}{2^N} \left(\sum_{l=0}^{N-1} l 2^l + N \right) \tag{8.2.13}$$

The sum can be evaluated using a table of sums or:

$$\sum_{l=0}^{N-1} l 2^l = \frac{1}{\ln(2)} \frac{d}{d\alpha} \sum_{l=0}^{N-1} 2^{\alpha l} \Bigg|_{\alpha=1} = \frac{1}{\ln(2)} \frac{d}{d\alpha} \frac{2^{\alpha N} - 1}{2^\alpha - 1} \Bigg|_{\alpha=1} = N 2^N - 2(2^N - 1) \tag{8.2.14}$$

giving:

$$< C(s) > \geq (N-2) + \frac{1}{2^N}(N+2) > N - 2 \tag{8.2.15}$$

Thus the average complexity of strings of length N cannot be reduced by more than two bits. This strict lower bound applies to all measures of complexity. ∎

We can also interpret this discussion to mean that the best UTMs to use to define complexity are those that are invertible—they have a one-to-one mapping of strings to representations. In this case we have a mapping $r(s)$ which gives the unique representation of a string. The reason that such UTM are better is that there are only a limited number of representations shorter than N; if we use up more than one of them for a particular string, then we will have fewer representations to use for others. Such UTM are closely analogous to our understanding of encoding and decoding as described in information theory. The UTM is the decoder and the mapping of the string onto its representation is the encoding.

Because most strings are incompressible, we can also prove that if we have an ensemble of strings defined by the probability $P(s)$, then the average algorithmic complexity of these strings is essentially the same as the Shannon information. In particular, the ensemble of all of the strings of length N have a Shannon information of N bits and an average algorithmic complexity which is the same. The catch is recogniz-

ing that to specify $P(s)$ itself requires an algorithm whose complexity must enter into the discussion. The proof follows from the discussion in Section 1.8. An ensemble defined by a probability $P(s)$ can be encoded in such a way that the average string length is given by the Shannon information. We now realize that to define the string complexity we must include the description of the decoding operation:

$$\sum_s P(s)C(s) = \sum_s P(s)I_s + C(P) \tag{8.2.16}$$

where the expression $C(P)$ represents the complexity of the decoding operation for the universal computer U for the ensemble given by $P(s)$. $C(P)$ depends in part on the algorithm used to specify the ensemble probability $P(s)$. For the average ensemble complexity to be essentially equal to the average Shannon information, the specification of the ensemble must itself be simple.

For Markov chains a similar result applies—the Shannon information of a string representing a Markov chain is the same as the algorithmic complexity of the same string, as long as the algorithm specifying the Markov chain is simple.

A general consequence of the definition of algorithmic complexity is a limitation on what TM can do. No TM can generate a string more complex than the input string that it is provided with, plus the information in its table—otherwise we would have redefined the complexity of the output string to take this into consideration. This is a key limitation of TM: TM (and computers that are realizations of this model) cannot generate new information. They can only process information they are given. As discussed briefly in Section 1.9.7, this limitation can be overcome by a TM that is given a string of random bits as input. The infinitely complex input means the limitation does not apply. It remains to be demonstrated what tasks such a TM can perform that are not possible for a conventional TM. If such tasks are identified, there will be important implications for computer design. In this context, it may also be suggested that some forms of creativity might be linked to the availability of randomness (see Section 1.9.7). We will return to this issue at the end of the chapter.

While the definition of complexity using UTM is appealing, there is a profound difficulty with this proof. It is nonconstructive. No method is given to determine the complexity of a particular string. Indeed, it can be proven that this is a fundamentally difficult task—the time necessary for a TM to determine $C(s)$ grows exponentially with the length of s. At least this is true when there is a bound on the complexity, e.g., by Eq. (8.2.4). Otherwise the complexity is noncomputable. We find the complexity of a string by trying all input strings in the UTM to see which one gives the necessary output. If the complexity is not bounded, then the halting problem implies that we cannot tell if the UTM will halt on a particular input, thus it is noncomputable. If the complexity of the string is bounded, then we only try strings up to this bound, and it is possible to determine if the UTM will halt for members of this bounded set of strings. Nevertheless, trying each string requires a time that grows exponentially with the bound, and therefore is not practical except for a few very simple strings. The process of finding the complexity of a string is akin to a process of trying models for the string. A model is a TM that might, when given the

proper input, generate the string. It is possible to try many models. However, to determine the actual compressed string may not be practical in any reasonable time. With any particular set of models, we can, however, find an upper bound on the complexity of a string. One of the possible models is that of a Markov chain as used by Shannon information theory. Algorithmic complexity allows more general TM models. However, by our discussion it is improbable that a randomly chosen string will be compressible by any algorithm.

In summary, the universality of complexity is a statement that the use of different UTMs in the definition of complexity affects the result by no more than a constant. This constant is the length of the program that translates the input of one UTM to the other. Significantly, the more complex the string is, the more universal is the value of its complexity. This follows because the length of translation programs becomes less and less relevant for longer and longer descriptions/representations. Since we are interested in properties of complex systems whose descriptions are long, we can, with caution, rely on the universality of their complexity. This is not the case with simple systems whose descriptions and therefore complexities are "subjective"—they depend on the conventions for description. These conventions, in our mathematical definition,are represented by the choice of UTM used to define complexity. We also showed that most strings are not compressible and that the Shannon information measure is the same as the average algorithmic complexity for all concisely describable ensembles. In what follows,unless otherwise mentioned, we assume a particular definition of complexity $C(s)$ using the UTM U.

8.2.2 *Mathematical systems: numbers and functions*

One of the difficulties in discussing complexity is that many elementary mathematical constructs have unusual properties when considered from the point of view of complexity. Philosophers have been troubled by these points,and they have been extensively debated over the centuries. Most of the problems revolve around various forms of infinity. Unlimited numbers and infinite precision often simplify symbolic mathematical discussions;however, they are not well behaved from the point of view of complexity measures. There appears to be a paradox here that will be clarified when we distinguish between the complexity of a set of numbers and the complexity of an element of the set.

Let us consider the complexity of specifying a single integer. The difficulty with integers is that there are infinitely many of them. Using an information theory point of view, assigning equal probability to all integers would imply that any particular integer would have no probability of occurring. If I ask you to give me a positive integer, from 1 to infinity with equal probability, there is no chance that you will give me an integer below any particular cutoff value,say N. This means that you will need arbitrarily many digits to specify the integer, and there is no limit to the information required. Thus the complexity of specifying a single integer is infinite. However, if we allow only integers between 1 and a large positive number—say $N = 10^{90}$, roughly the number of elementary particles in the known universe—the complexity of specifying one of the integers is only $\log(N)$,about 300 bits. The drastic difference between the

complexity of specifying an arbitrary integer (infinite) and the complexity of an enormously large number of integers (300 bits) suggests that systems that are easy to define may be highly complex. The whole field of number theory has shown that integers are not as simple as they first appear. The measure of complexity of specifying a single integer may appear to be far from more abstract discussions like those of the halting problem or Gödel's theorem (Section 1.9.5),however, they are related. This is apparent since these theorems do not apply to finite sets.

In what sense are integers simple? We can consider the length of a UTM input string that can generate all the positive integers. As discussed in the last section, this is similar to the definition of their Kolmogorov or algorithmic complexity. The program would, starting from zero and keeping a list, progressively add one to the preceding integer. The problem is that such a program never halts, and the task is not complete. We can generalize our definition of a Turing machine to allow for this case by saying that, by definition, this simple program is generating all integers. Then the algorithmic complexity of the integers is quite small. Another way to do this is to consider the complexity of recognizing an integer—the recognition complexity. Recognizing an integer is trivial if we are considering only binary strings, because all of them represent integers. The point,however, is that we can expand the space of possible characters to include various symbols:letters,punctuation, mathematical operations, etc. The mathematical operations might act upon integers. We then ask how long is a TM program that can recognize any integer that appears as a combination of such characters. The length of such a program is also small.

We see that we must distinguish between the complexity of elements of a set and the set itself. A program that recognizes integers is concerned with the attributes of the integers required to define them as a set, rather than the specification of a particular integer. The algorithmic complexity of the set of all integers is small even though the information contained in a single integer can be arbitrarily large. This distinction between the information contained in an element of a set and the information necessary to define the set will also be important when we consider the complexity of physical systems.

The complexity of a single real number is also infinite. Specifying an arbitrary real number requires infinitely many digits. However, if we confine ourselves to any reasonable precision, the complexity becomes very manageable. For example, the most accurately known fundamental constant in science is the electron magnetic moment in Bohr magnetons

$$\mu_e/\mu_B = 1.001159652193(10) \tag{8.2.17}$$

where the parenthesis indicates the error estimate, corresponding to 11 accurate decimal digits or 37 binary digits. If we consider $1 - \mu_e/\mu_B$ we immediately lose 3 decimal digits. Thus, similar to integers, the practical complexity of a real number is not very large.

The discussion of integers and reals suggests that under practical circumstances a single number is not a highly complex object.Generally, the complexity of a system arises because of the presence of a large number of parameters that must be specified.

However, there is only reason to consider them collectively as a system if they are coupled to each other.

The next category of mathematical objects that we consider are functions. To specify a function $f(s)$ we must either describe its operation by a formula or specify its action on each possible argument. We consider Boolean functions (functions with binary output, see Section 1.9.2), $f(s) = \pm 1$, of a binary string, $s = (s_1 s_2 \ldots s_{N_e})$. The number of arguments of the function—input bits—is N_e. There are 2^{N_e} possible values of the input string. For each of these there are two possible outcomes (output values). All Boolean functions may be specified by listing the binary output for each possible input state. Each possible output is independent. The number of different Boolean functions is the number of possible sets of outputs which is $2^{2^{N_e}}$. Assuming that all of the possible Boolean functions are equally likely, the complexity of a Boolean function (the amount of information necessary to specify it) is the logarithm of this number or $C(f) = 2^{N_e}$. The representation of a Boolean function in terms of $C(f)$ binary variables can also be made explicit as a string representing the presence or absence of terms in the disjunctive normal form described in Section 1.9.2.

A binary function with N_a outputs is the same as N_a independent Boolean functions. If we assume that all possible combinations of Boolean functions are equally likely, then the total complexity is the sum of the complexity of each, or

$$C(f) = N_a 2^{N_e} \tag{8.2.18}$$

The asymmetry between input and output is a fundamental one. It arises because we need to specify for each possible input which of the possible outputs is output. Specifying "which" is a logarithmic operation in the number of possibilities, and therefore the influence of the output space on the complexity is logarithmic compared to the influence of the input. This discussion will be generalized later to consider a physical system that acts in response to its environment. The environment will be specified by a number of binary variables (environmental complexity) N_e, and its actions will be specified by a number of binary variables (action complexity) N_a.

8.3 Complexity of Physical Systems

In order to apply our understanding of the complexity of mathematical constructs to physical systems, we must develop a fundamental understanding of representations. The complexity of a physical system is to be defined as the length of the shortest string s that can represent its properties—the results of possible measurements/observations. In Section 8.3.1 we discuss the relationship between thermodynamics and information theory. This will enable us to define the complexity of ergodic and nonergodic systems. The resulting information measure is essentially that of Shannon information theory. When we consider algorithmic complexity, we can ask whether this is the smallest amount of information that might be used. This is discussed in Section 8.3.2. Section 8.3.3 introduces the complexity profile, which measures the complexity as a function of the scale of observation. Implications of the time scale of observation, for chaotic dynamics, are discussed in Section 8.3.4. Section 8.3.5

discusses examples and properties of the complexity profile. Sections 8.3.1 through 8.3.5 are based upon descriptive complexity. To better account for the behavior of a system in response to its environment we consider behavioral complexity in Section 8.3.6. This turns out to be closely related to descriptive complexity. Other issues related to the role of the observer are discussed in Section 8.3.7.

8.3.1 *Entropy and the complexity of physical systems*

The definition of complexity of a system requires us to develop an understanding of the relationship of information to the physical properties of a system. The most direct relationship is the relationship of entropy and information. At the outset, it should be understood that these are very different concepts. Entropy is a specific physical property of systems that are in equilibrium, or are in well-defined ensembles. Information is not a unique physical property. Instead it is related to representations of digits. Information can be a property of a time sequence or any other set of degrees of freedom. For example, the information content of a set of characters written on a piece of paper can be given. The entropy, however, would be largely a property of the paper or the ink. The entropy of paper is difficult to determine precisely, but simpler substances have entropies that have been determined and are tabulated at specific temperatures and pressures. We also know that entropy is conserved in reversible adiabatic processes and increases in irreversible ones.

Despite the significant conceptual difference between information and entropy, the formal definition of information discussed in Section 1.8 appears very similar to the definition of entropy discussed in Section 1.3. Thus, it makes sense that the two are related when we develop an understanding of complexity. It is helpful to review the definitions. The entropy was defined first for the microcanonical ensemble, which specifies the macroscopic energy U, number of particles N, and volume V, of the system. We assume that all states (microstates) of the system with this energy, number of particles and volume are equally likely in the ensemble. The entropy was written as

$$S = k \ln \Omega(U, N, V) \tag{8.3.1}$$

where $\Omega(U, N, V)$ is the number of such states. The coefficient k is defined so that the units of entropy are consistent with units of energy and temperature for the thermodynamic relationship $T = dU/dS$.

Information was defined for a string of characters. Given the probability of the string of characters, the information is defined by Eq. (8.2.1). The logarithm is taken to be base 2 so that the information is measured in units of bits. We see that the information content is related to selecting a single state out of an ensemble of possibilities.

We can relate the two definitions in a mathematically direct but conceptually significant way. If we want to specify a particular microstate of a thermodynamic system, we must select this microstate from the whole ensemble. The probability of this particular state is given in the microcanonical ensemble by $P = 1/\Omega$. If we think about the state of the system as a message containing information, we can use Eq.(8.2.1) to give the amount of information as:

$$I(\{x,p\}|(U,N,V)) = S(U,N,V)/(k \ln 2) \tag{8.3.2}$$

This expression should be understood as the amount of information contained in a microstate $\{x,p\}$, when the system is in the macrostate specified by U,N,V—it is also the information necessary to describe precisely the microstate. This is the fundamental relationship we are looking for. We review its meaning in terms of the description of a particular idealized physical system.

If we want to describe the microstate of a system, like a gas of particles in a box, classically we must specify all of the positions and momenta of the particles $\{x_i,p_i\}$. If N is the number of particles, then there are $6N$ coordinates, 3 position and 3 momentum coordinates for each particle. To specify exactly the position of each particle appears to require arbitrary precision in these coordinates. If we had to specify even a single position exactly, it would take an infinite number of binary digits. However, quantum mechanics is inherently granular, thus there is a smallest distance Δx within which we do not need to specify one position coordinate of a particle. The particle location is uniquely given once it is within a region Δx. More correctly, the particle must be located within a region of position and momentum of $\Delta x \Delta p = h$, where h is Planck's constant. The granularity defines the precision necessary to specify the positions and momenta, and thus also the amount of information (number of bits) needed in order to describe completely the microstate. The definition of the entropy takes this into account, otherwise the counting of possible microstates of the system would be infinite. The complete calculation of the entropy (which also takes into account the indistinguishability of the particles) is given in Question 1.3.2. We now recognize that the calculation of the entropy is precisely a calculation of the information necessary to describe the microstate.

There is another way to think about the relationship of entropy and information. It follows from the recognition that the number of states of a string of $I(\{x,p\}|(U,N,V))$ bits is the same as the number of states of the system. If we consider a mapping of system states onto strings, the strings enumerate or label the system states. If there are $I(\{x,p\}|(U,N,V))$ bits in each string, then there is a one-to-one mapping of system states onto the strings, and a string uniquely identifies a system state. We say that a string represents a system microstate.

We thus identify the entropy of a physical system as the amount of information necessary to identify a single microstate from a specified macroscopic ensemble. For an ergodic macroscopic system, this definition is a robust one. It does not matter if we consider a typical or an average amount of information. What happens if the system is nonergodic? There are two kinds of nonergodic systems we will discuss: a magnet with a well-defined magnetization below its ordering phase transition (see Section 1.6), and a glass where there are many frozen coordinates describing the local arrangements of atoms (see Section 1.4). Many of these coordinates do not change during the time of a typical experiment. Should we include the information necessary to specify the frozen variables as part of the entropy? We would like to separate the discussion of the frozen variables from the fast ones that are in equilibrium. We use the entropy S to refer to the fast ensemble—the enumeration of the kinetically accessible states of the system. The same function of the frozen variables we will call C.

For the magnet, the amount of information contained in frozen variables is small. For the Ising model of a magnet (Section 1.6), below the magnetization transition only a single binary variable is necessary to specify if the system magnetization is UP or DOWN. We treat the magnet by giving the information about the magnetization explicitly as part of the ensemble description. The amount of information is insignificant compared to the information in the microstate of a system, and therefore is generally ignored.

In contrast, for a glass, the amount of information that is included in the frozen variables is large. How does this information relate to the thermodynamic treatment of the system? The conventional thermodynamic theory of phase transitions does not consider the existence of frozen information. It is designed for systems like the magnet, where this information is insignificant, and thus it does not apply to the glass transition. A different theory is necessary which includes the change from an ergodic to a nonergodic system, or a change from information in fast variables to information in frozen variables. Is there any relationship between the frozen information and the entropy? If they are related at all, there are two intuitive possibilities. One is that we must specify the frozen variables as part of the ensemble, and the amount of information necessary to describe the fast variables is just as large as if there were no frozen variables. The other is that the frozen variables balance against the fast variables so that when there is more frozen information there is less information in the fast variables. In order to determine which is correct, we will need to consider an experiment that measures both. As long as an experiment is being performed in which the frozen variables never change, then the amount of information in the frozen variables is fixed. Thermodynamic experiments only depend on entropy differences. We will need to consider an experiment that changes the frozen variables—for example, heating up a glass until it becomes a liquid or cooling it from a liquid to a glass. In such an experiment the frozen information must be accounted for. The difficulty with a glass is that we do not have an independent way to determine the amount of frozen information. Fortunately, there is another system where we do.

There is an intermediate example between a magnet and a glass that is of considerable interest. The structure of ice has a glasslike frozen disorder of its hydrogen atoms below approximately 100°K. The simplest way to think about this disorder is that it arises from a choice of orientations of the water molecule around the position of the oxygen atom. This means that there is a macroscopic amount of information necessary to specify the static structure of ice. The amount of information associated with this disorder can be calculated directly using a model for the structure of ice that takes into account the correlations between molecular orientations that are needed to form a self-consistent hydrogen structure within the oxygen lattice. A first estimate is based on an average of $3/2$ orientations per molecule or $C = Nk \ln(3/2) = 0.806$ cal/moleK. A review of better calculations is given in a book by Fletcher. The best is $C = 0.8145 \pm 0.0002$ cal/mole°K. The other calculation we need is the amount of entropy in steam. This can be obtained using a slight modification of the ideal gas calculation, that takes into account the rotational and internal vibrational motion of the water molecule.

The key experiment is to measure the change in the entropy of the system as a function of temperature as it is heated from ice all the way to steam. We find the entropy using the standard thermodynamic relationship (Section 1.3)

$$q = TdS \tag{8.3.3}$$

where q is the heat added to the system. At close to a temperature of zero degrees Kelvin ($T = 0$K) the entropy is zero because all motion stops, and there is only one possible state of the system. Thus we would expect

$$S(T) = \int_0^T q/T \tag{8.3.4}$$

—the total amount of entropy added to the system as it is heated up should be the same as the entropy of the gas. However, experimentally there is a difference of 0.82 ± 0.05 cal/moleK between the two. This is the amount of entropy in the gas that was not added to the system as it was heated. The coincidence of two numbers—the amount of entropy missing and the calculation of the information in the frozen structure of the hydrogen atoms, suggests that the missing entropy was present in the original state of the ice.

$$S(T) = C(T = 0) + \int_0^T \frac{q}{T} \tag{8.3.5}$$

This in turn implies that the information in the frozen degrees of freedom was transferred (but conserved) to the fast degrees of freedom. Eq.(8.3.5) is not consistent with the standard thermodynamic relationship in Eq. (8.3.3). Instead it should be modified to read:

$$q = Td(S + C) \tag{8.3.6}$$

This should be understood as implying that adding heat to a system increases the information either of the fast or frozen variables. Adding heat (e.g., to ice) increases the temperature of the system,so that fewer variables are frozen. In this case C decreases and S increases more than would be given by the conventional relationship of Eq. (8.3.3). When heat is not added to a system, we see that there can be processes that change the number of fast degrees of freedom and the number of static degrees of freedom while leaving their sum the same. We will consider this further in later sections.

Eq. (8.3.6) is important enough to present it again from a different perspective. The discussion will help demonstrate its validity by using a theoretical argument (Fig. 8.3.1). Rather than considering it from the point of view of heating ice till it becomes steam, we consider what happens either to ice or to a glass when we cool it down through the transition where degrees of freedom become frozen. In a theoretical description we start,above the freezing-in transition, with an ensemble of systems. As we cool the system we remove heat,and this is reflected in a decrease in the number of possible states of the system. We think of this as a shrinking of the number of elements of the ensemble. However, as we go through the freezing-in transition, the

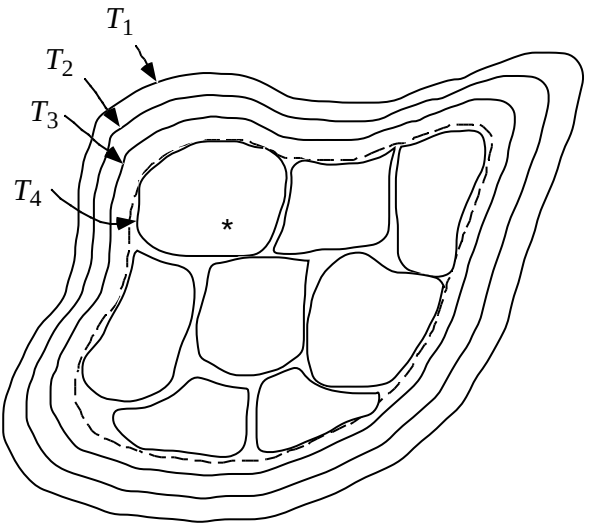

Figure 8.3.1 Schematic illustration of the effect on motion in phase space of cooling through a glass transition. Above the glass transition (T_1,T_2 and T_3) the system is ergodic — it explores the entire phase space. Cooling the system causes the phase space to shrink smoothly. The entropy, the logarithm of the volume of phase space, decreases. Below the glass transition, T_4, the system is no longer ergodic and the phase space breaks up into pieces. A particular system explores only one of the pieces. The total amount of information necessary to specify a particular microstate (e.g. indicated by the *) is the sum of $C/k\ \ln(2)$, the information necessary to specify which piece, and $S/k\ \ln(2)$, the information necessary to specify the particular state within the piece. ∎

ensemble breaks up into disjoint pieces that can not make transitions to each other. Any particular material must be in one of the disjoint pieces. Thus for a particular material we must track only part of the original ensemble. For an incremental decrease in temperature due to an incremental removal of heat, the information needed to identify (describe) a particular microstate is the sum of the information necessary to describe which of the disjoint parts of the ensemble the system is in, plus the information needed to specify which of the microstates the system is in once its ensemble fragment has been specified. This is the meaning of Eq. (8.3.6). The information to specify the ensemble fragment was transferred from the entropy S to the ensemble information C. The reduction of the entropy, S, is not reflected in the amount of heat that is removed.

We are now in a position to give a first definition of complexity. In order to describe a system and its behavior over time, we must describe the ensemble it is in. This information is given by $C/k\ \ln(2)$. If we insist on describing the microstate of the system, we must add the information contained in the fast degrees of freedom $S/k\ \ln(2)$. The question is whether we should insist on describing the microstate. Typically, the

whole point of describing an ensemble is that we don't need to specify the particular microstate. We will return to address this question in greater detail later. However, for now it is reasonable to consider describing the system to be specifying just the ensemble. This implies that the information in the frozen variables $C/k\ln(2)$ is the complexity. For a thermodynamic system in the microcanonical ensemble, the complexity would be given by the (small) number of bits in the specification of the three variables (U,N,V) and the number of bits necessary to specify the type of element (atom,molecule) that is present. The actual amount of information seems not to be precisely defined. For example, we have not identified the number of bits to be used in specifying (U,N,V). As we have seen in the discussion of algorithmic complexity, this is to be expected, since the conventions of how the information is specified are crucial when there is only a small amount.

We have learned from this discussion that for a nonergodic system, the complexity (the frozen ensemble information) is bounded by the sum over the number of fast and static degrees of freedom ($C + S > C$). For material systems, we know in principle how to measure this. As in the case of ice, we heat up the system to the vapor phase where the entropy can be calculated,then subtract the entropy added during the heating process. This gives us the value of $C + S$ at the temperature from which the heating began. If we know that $C >> S$, then the result is the complexity itself. In order for this technique to work at all, the complexity must be large enough so that experimental accuracy can enable its measurement. Estimates we will give later imply that complexities of biological organisms are too small to be measured in this way.

The concept of frozen degrees of freedom immediately raises the question of the time scale in which the experiment is performed. Degrees of freedom that are frozen on one time scale are not on sufficiently longer ones. If our time scale of observation would be arbitrarily long, we would always describe systems in equilibrium. The entropy would then be large and the complexity would be negligible.On the other hand, if our time scale of observation was extremely short so that microscopic motions were detected, then our complexity would be large and the entropy would be negligible. This motivates the introduction of the complexity profile in Section 8.3.3.

Question 8.3.1 Calculate the information necessary to specify the microstate of a mole of an ideal gas at $T = 0°C$ and $P = 1$atm. Use the mass of a helium or neon atom for the mass of the ideal gas particle. This requires a careful investigation of units.A table of fundamental physical constants is given on the following page.

Solution 8.3.1 The entropy of an ideal gas is found in Section 1.3 to be:

$$S = kN[\ln(V/N\lambda(T)^3) + 5/2] \tag{8.3.7}$$

$$\lambda(T) = (h^2/2\pi mkT)^{1/2} \tag{8.3.8}$$

The information content of a microstate is given by Eq. (8.3.2).

Each of the quantities must be evaluated numerically from appropriate tables. A mole of particles is

$$N_0 = 6.0221367 \times 10^{23} \text{/mole} \tag{8.3.9}$$

At the temperature

$$T_0 = 0\ ^\circ\text{C} = 273.13\ ^\circ\text{K} \tag{8.3.10}$$

$$kT_0 = 0.0235384\ \text{eV} \tag{8.3.11}$$

and pressure

$$P_0 = 1\text{atm} = 1.01325 \times 10^5\ \text{Pascal} = 1.01325 \times 10^5\ \text{Newton/m}^2 \tag{8.3.12}$$

the volume (of a mole of particles) of an ideal gas is:

$$V = N_0 kT/P_0 = 22.41410 \times 10^{-3}\ \text{m}^3\text{/mole} \tag{8.3.13}$$

the volume per particle is:

$$V/N = 37219.5\ \text{Å}^3 \tag{8.3.14}$$

At the same temperature we have:

$$\lambda(T) = (2\pi mkT/h^2)^{-1/2} = m[AMU]^{-1/2} \times 1.05633\text{Å} \tag{8.3.15}$$

This gives the total information for a mole of helium gas at these conditions of

$$I = N_0\,(18.5533 + 3/2\,\ln(m[AMU])) = 1.24 \times 10^{25} \tag{8.3.16}$$

Note that the amount of information per particle is only of order 10 bits. ∎

---

hc = 12398.4 eV Å
k = 1.380658x10^{-23} Joule/°K
$R = kN_0$ = 8.3144 Joule/°K/mole
c = 2.99792458 10^8 Meter/second
h = 6.6260755 10^{-34} Joule second
e = 1.60217733 10^{-19} Coulomb
ProtonMass = 1.6726231x10^{-27} kilogram
1 AMU = 1.6605402x10^{-27} kilogram = 9.31494x10^9 eV
M[Helium] = 4.0026 AMU
M[Neon] = 20.179 AMU
M[Helium] c^2 = 3.7284x10^9
M[Neon] c^2 =1.87966x10^{10}

Table 8.3.1 Fundamental constants ∎

8.3.2 *Algorithmic complexity of physical systems*

The complexity of a system is designed to measure the amount of information necessary to describe it, or its behavior. In this section we address the key word "necessary." This word suggests that we are after the minimum amount of information. The minimum amount of information depends on our capabilities of inference from a smaller amount of information. As discussed in Section 8.2.2, logical inference and computation lead to the definition of algorithmic complexity. However, for an ensemble that can be described simply, the algorithmic complexity is no different than the Shannon information.

Since we have established a connection between the complexity of physical systems and representations in terms of character strings, we can apply these results directly to physical systems. A physical system in equilibrium is represented by an ensemble. At any particular time, it is in a single microstate. The specification of this microstate can be compressed by encoding in certain rare cases. However, on average the compression cannot lead to an amount of information significantly different from the entropy (divided by $k \ln(2)$) of the system. This conclusion follows because the microcanonical (or canonical) ensemble can be concisely described. For a nonergodic system like a glass, the microstate description has been separated into two parts. It is no longer true that the ensemble of dynamically accessible states of a particular system is concisely describable. The information in the frozen degrees of freedom is precisely the information necessary to specify the ensemble of dynamically accessible states. The total information, $(C + S)/k \ln(2)$, represents the selection of a microstate from a simple ensemble (microcanonical or canonical). Since the total information cannot be compressed, neither can either of the two parts of the information—the frozen degrees of freedom that we have identified with the complexity, or the additional information necessary to specify a particular microstate. Thus the algorithmic complexity is the same as the information for either part.

We can now, finally, explain the experimental observation that an adiabatic process does not change the entropy of a system (Section 1.3). The algorithmic description of an adiabatic process requires only a few pieces of information, e.g., the size of a force applied over a specified distance. If a new microstate of the system can be described by the original microstate plus the process of adiabatic change, then the amount of information in the microstate has not been changed, and the adiabatic process does not change the microstate algorithmic complexity—the entropy of the system. Like other aspects of statistical mechanics (Section 1.3), this should not be understood as a proof but rather as an explanation of the relationship of the thermodynamic observation to the microscopic properties. Using this explanation, we can identify the nature of an adiabatic process as one that is described microscopically by a small amount of information.

This becomes clearer when we compare adiabatic and irreversible processes. Our argument that an adiabatic process does not change the entropy is based on considering the information necessary to describe an adiabatic process—slowly moving a piston to expand the space available to a gas. An irreversible process could achieve a similar expansion, but would not be thermodynamically the same. Take, for example,

the removal of a partition that separates the gas from a second, initially empty, chamber. The irreversible process of expansion of the gas results in a final state which has a higher entropy (see Question 1.3.4). The removal of a partition in itself does not appear to require a lot of information to describe. One moment after the partition is removed, the entropy of the system is the same as before. To understand how the entropy increases, we must consider the nature of irreversible dynamics.

A key ingredient in our understanding of physical systems is that the time evolution of an isolated system can be obtained from the simple laws of mechanics (classical or quantum). This means that the dynamics of an isolated system conserves the amount of information as well as the energy. Such dynamics are called conservative. If we consider an ensemble of systems starting in a particular region of phase space, the phase space position evolves in time, but the volume of the phase space that is occupied—the entropy—does not change. This conservation of phase space can be understood from our discussion of algorithmic complexity: since the deterministic dynamics of a system can be computed, the algorithmic complexity of the system is conserved. Where does the additional entropy come from for the final equilibrium state after the expansion?

There are two parts to the process of proceeding to a true equilibrium state. In the first part the distinction between the nonequilibrium and equilibrium state is obscured. At first there is macroscopically observable information—the particles are in one half of the chamber. This information is converted to microscopic correlations between atomic positions and momenta. The conversion occurs when the gas expands to fill the chamber, and various currents that follow this expansion become smaller and smaller in extent. The microscopic correlations cannot be observed on a macroscopic scale, and for standard observations the system is indistinguishable from an equilibrium state. The transfer of information from macroscopic to microscopic scale is related to issues of chaos in the dynamics of physical systems, which will be discussed later.

The second part to the process is an actual increase in the entropy of the system. The additional entropy must come from outside the system. In macroscopic physical processes, we are not generally concerned with isolating the system from information transfer, only with isolating the system from energy transfer. Thus we can surmise that the expansion of the gas is followed by an information transfer that enables the entropy to increase to its equilibrium value without changing the energy of the system. Many of the issues related to describing this nonequilibrium process will not be addressed here. We will, however, begin to address the topic of the scale of observation at which correlations appear using the complexity profile in the following section.

8.3.3 *Complexity profile*

General approach In this section we discuss the relationship of microscopic and macroscopic complexity. Our objective is to develop a consistent language for discussing complexity as a function of length scale. In the following section we will discuss the complexity as a function of time scale, which generalizes the discussion of frozen and fast degrees of freedom in Section 8.3.1.

When we describe a system, we are not generally interested in a microscopic description of the positions and velocities of all of the particles. For a thermodynamic system there are only a few macroscopic parameters that we use to describe the system. This is indeed the reason we use entropy as a summary of the many hidden parameters of the system that we are not interested in. The microscopic parameters change too fast and over too small distances to matter for our macroscopic measurements/experience. The same is true more generally about systems that are not in equilibrium: a macroscopic description does not require specifying the position of each atom. This implies that we must develop an understanding of complexity that is not tied to the microscopic description, but is relevant to observations at a particular length and time scale.

This point lies at the root of a conceptual problem in thinking about the complexity of systems. A gas in equilibrium has a large entropy which is its microscopic complexity. This is counter to our understanding of complex systems. Systems in equilibrium are intuitively simpler than nonequilibrium systems such as a human being. In Section 8.3.1 we started to address this problem by identifying the complexity of a nonergodic system as the information necessary to specify the frozen degrees of freedom. We now discuss a more systematic approach to dealing with macroscopic observations.

In order to consider the macroscopic complexity, we have to define what we mean by macroscopic in a formal sense. The concept of macroscopic must be understood in relation to a particular observer. While we often consider experimental results to be independent of the observer, there are various ways in which the observer is essential to the observation. In this context, in which we are concerned with the meaning of macroscopic, considering the observer is essential.

How do we characterize the difference between a microscopic and a macroscopic observer? The most crucial difference is that a microscopic observer is able to distinguish between all inherently distinguishable states of the system, while a macroscopic observer cannot. For a macroscopic observer, many microscopically distinct states appear the same. This is related to our understanding of complexity, because the macroscopic observer need only specify which of the macroscopically distinct states the system is in. The microscopic observer must specify which of the microscopically distinct states the system is in. Thus the macroscopic complexity must always be smaller than the microscopic complexity of a system. Instead of considering a unique macroscopic observer, we will consider a sequence of observers with a progressively poorer ability to distinguish microstates. Using these observers, we will define the complexity profile.

Ideal gas These ideas can be directly applied to the ideal gas. We generally think about a macroscopic observer as having an inability to distinguish fine-scale distances. Thus we expect that the usual uncertainty in particle position Δx will increase for a macroscopic observer. However, we learn from quantum mechanics that a unique microstate of the system is defined using an uncertainty in both position and momentum, $\Delta x \Delta p = h$. Thus for the macroscopic observer to confuse distinct microstates, the product $\Delta x \Delta p$ must be larger than its minimum value—an observation of the system provides measurements of the position and momentum of each particle, whose uncertainty has a product greater than h. We can label our observers by this uncertainty, which we call $\tilde{h}$.

If we retrace our steps to the calculation of the entropy of an ideal gas (Question 1.3.2), we can recognize that essentially the same calculation applies to the complexity with the uncertainty $\tilde{h}$. An observer with the uncertainty $\tilde{h}$ will determine the complexity of the ideal gas according to Eq.(8.3.7) and Eq.(8.3.8), with h replaced by $\tilde{h}$. Thus we define the complexity profile for the ideal gas in equilibrium as:

$$C(\tilde{h}) = S - 3kN \ln(\tilde{h}/h) \qquad \tilde{h} > h \tag{8.3.17}$$

This equation describes a complexity that decreases as the ability of the observer to distinguish states decreases. This is as we expected. Despite the weak logarithmic dependence on $\tilde{h}$, $C(\tilde{h})$ decreases rapidly because the coefficient of the logarithm is so large. By the time $\tilde{h}$ is about 100 times h the complexity profile has become negative for the ideal gases described in Question 8.3.1.

What does a negative complexity mean? It actually means that we have not done the calculation quite right. The counting of states we did for the ideal gas assumed that the particles were well separated from each other. If they begin to overlap then we must count the possible states differently. This overlap is significant precisely when Eq.(8.3.17) becomes negative. If the particles really overlapped then quantum statistics becomes important; the gas is said to be degenerate and satisfies either Fermi-Dirac or Bose-Einstein statistics. In our case the overlap arises only because the observer cannot distinguish different particle positions. In this case, the counting of states is appropriate to a classical ideal gas, as we now explain.

To calculate the complexity as a function of $\tilde{h}$ for an equilibrium state whose entropy is S, we start by calculating the number of microstates that the observer cannot distinguish. The logarithm of this number of microstates, which we call $S(\tilde{h})/k \ln(2)$, is the amount of information necessary to specify a microstate, if the macrostate is known. Thus we have that:

$$C(\tilde{h}) = S - S(\tilde{h}) \tag{8.3.18}$$

To count the number of microstates that the observer cannot distinguish, we note that the possible microstates of a particular particle are grouped together by the observer into bins (regions or cells of position and momentum) of size $(\Delta x \Delta p)^d = \tilde{h}^d$, where $d = 3$ is the dimensionality of space. The observer determines only that a particle is within a certain region. In the classical ideal gas each particle moves independently, so more than one particle may occupy the same microstate. However, this is unlikely. As $\tilde{h}$ increases it becomes increasingly likely that there is more than one particle in a region. If the number of particles in a certain region is n_i, then the number of distinct microstates of the bin that the observer does not distinguish is:

$$\frac{g^{n_i}}{n_i!} \tag{8.3.19}$$

where $g = (\tilde{h}/h)^d$ is the number of microstates within a region. This is the product of the number of states each particle may be in, corrected for particle indistinguishability. The number of microstates of the whole system that appear to the observer to be the same is the product of such terms for each region:

$$\prod_i \frac{g^{n_i}}{n_i!} \tag{8.3.20}$$

From this we can determine the complexity of the state determined by the observer as:

$$C(\tilde{h}) = S - S(\tilde{h}) = S - k\ln\left(\prod_i \frac{g^{n_i}}{n_i!}\right) \tag{8.3.21}$$

If we consider this expression when $g = 1$—a microscopic observer—then n_i is almost always either zero or one and each term in the product is one (a more exact treatment requires treating the statistics of a degenerate gas). Then $C(\tilde{h})$ is S, which means that the microstate complexity is just the entropy. For $g > 1$ but not too large, n_i will still be either zero or one, and we recover Eq. (8.3.17). On the other hand, using this expression it is possible to show that for a large value of g, when the values of n_i are significantly larger than one, the complexity goes to zero.

We can understand this by recognizing that as g increases, the number of particles in each bin increases and becomes closer to the average number of particles in a bin according to the macroscopic probability distribution. This is the equilibrium macrostate. By our conventions we are measuring the amount of information necessary for the observer to specify its observation in relation to the equilibrium state. Therefore, when the average number of particles in a bin becomes close enough to this distribution, there is no information that must be given. To write this explicitly, when n_i is much larger than one we apply Sterling's approximation to the factorial in Eq. (8.3.21) to obtain:

$$C(\tilde{h}) = S - k\sum_i n_i\left(\ln(g/n_i) + 1\right) = S + k\sum_i gP_i\ln(P_i) - kN \tag{8.3.22}$$

where $P_i = n_i/g$ is the probability a particle is in a particular state according to the observer. It is shown in Question 8.3.2 that $C(\tilde{h})$ is zero when P_i is the equilibrium probability for finding a particle in region i (note that i stands for both position and momentum (x,p)).

There are additional smaller terms in Sterling's approximation to the factorial that we have neglected. These terms are generally ignored in calculations of the entropy because they are not proportional to the number of particles. They are, however, relevant to calculations of the complexity:

$$C(\tilde{h}) = S - k\sum_i n_i\left(\ln(g/n_i) + 1\right) + k\sum_i \ln\left(\sqrt{2\pi n_i}\right) \tag{8.3.23}$$

The additional terms are related to fluctuations in the density. This will become apparent when we analyze nonuniform systems below.

We will discuss additional examples of the complexity profile below. First we simplify the complexity profile for observers that measure only the positions and not the momenta of particles.

Question 8.3.2 Show that Eq.(8.3.22) is zero when P_i is the equilibrium probability of locating a particle in a particular state identified by momentum p and position x. For simplicity assume that all g states in the cell have essentially the same position and momentum.

Solution 8.3.2 We calculate an expression for $P_i \rightarrow P(x,p)$ using Boltzmann probability for a single particle (since all are independent):

$$P(x,p) = NZ^{-1}e^{-p^2/2mkT} \tag{8.3.24}$$

where Z is the one particle partition function given by:

$$Z = \sum_{x,p} e^{-p^2/2mkT} = \int \frac{d^3x d^3p}{h^3} e^{-p^2/2mkT} = \frac{V}{\lambda^3} \tag{8.3.25}$$

We evaluate the expression:

$$-k\sum_i gP(x,p)\ln(P(x,p)) + kN \tag{8.3.26}$$

which, by Eq.(8.3.22), we want to show is the same as the entropy. Since all g states in cell i have essentially the same position and momentum, this is equal to:

$$-k\sum_{x,p} P(x,p)\ln(P(x,p)) + kN = k\sum_{x,p}\left(\ln(V/N\lambda^3) + p^2/2mkT\right)\left(N\lambda^3/V\right)e^{-p^2/2mkT} \tag{8.3.27}$$

which is most readily evaluated by recognizing it as:

$$kN + kNZ^{-1}\left(\ln(V/N\lambda^3) - \frac{1}{\beta}\frac{d}{d\beta}\right)Z = kN\left(\ln(V/N\lambda^3) + 5/2\right) \tag{8.3.28}$$

which is S as given in Eq. (8.3.7). ∎

Position without momentum The use of the scale parameter $\Delta x \Delta p$ in the above discussion should trouble us, because we do not generally consider the momentum uncertainty on the macroscopic scale. The resolution of this problem arises because we have assumed that the system has a known energy or temperature. If we know the temperature then we know the thermal velocity or momentum:

$$\Delta p \approx \sqrt{mkT} \tag{8.3.29}$$

It does not make sense to have a momentum uncertainty of a particle that is much greater than this. Using $\Delta x \Delta p = h$ this means there is also a natural uncertainty in position which is the thermal wavelength λ given by Eq. (8.3.8). This is the maximal quantum position uncertainty, unless the observer can distinguish the thermal motion of individual particles. We can now think about a sequence of observers who do not distinguish the momentum of particles (they have a larger uncertainty than the thermal momentum) but have increasing uncertainty in position given by $L = \Delta x$, or $g = (L/\lambda)^d$. For such observers the equilibrium momentum probability distribution

is to be assumed. In this case the number of particles in a cell n_i contributes a term to the entropy that is equal to the entropy of a gas with this many particles in the volume L^d. This gives a total entropy of:

$$S(L) = k \sum_i n_i \left[\ln(L^d / n_i \lambda^3) + 5/2 \right] \tag{8.3.30}$$

and the complexity is:

$$C(L) = S - k \sum_i n_i \left(\ln(g / n_i) + 5/2 \right) \tag{8.3.31}$$

which differs in form from Eq. (8.3.22) only in the constant.

While we generally do not think about measuring momentum, we do measure velocity. This follows from the content of the previous paragraph. We consider observers that measure particle positions at different times and from this they may infer the velocity and indirectly the momentum. Since the observer measures n_i, the determination of velocity depends on the observer's ability to distinguish moving spatial density variations. Thus we consider the measurement of $n(x,t)$, where x has macroscopic meaning as a granular coordinate that has discrete values separated by L. We emphasize,however, that this description of a space- and time-dependent density assumes that the local momentum distribution of the system is consistent with an equilibrium ensemble. The more fundamental description is given by the distribution of particle positions and momenta, $n_i = n(x,p)$. Thus, for example, we can also describe a rotating disk that has no macroscopic changes in density over time, but the rotation is still macroscopic. We can also describe fluid flow in an incompressible fluid. In this section we continue to restrict ourselves to the description of observations at a particular time. The time dependence of observations will be considered in Section 8.3.5.

Thus far we have considered systems that are in generic states selected from the equilibrium ensemble. Equilibrium systems are uniform on all but very microscopic scales, unless we are exactly at a phase transition. Thus, most of the complexity disappears on a scale that is far smaller than typical macroscopic observations. This is not necessarily true about nonequilibrium systems. Systems that are in states that are far from equilibrium can have nonuniform densities of particles.A macroscopic observer will see these macroscopic variations. We will consider a couple of different examples of nonequilibrium states to illustrate some properties of the complexity profile. Before we do this we need to consider the effect of algorithmic compression on the complexity profile.

Algorithmic complexity and error To discuss macroscopic complexity more completely, we turn to algorithmic complexity as a function of scale. The complexity of a system,particularly a nonequilibrium system,should be defined in terms of the algorithmic complexity of its description. This means that patterns that are present in the positions (or momenta) of its particles can be used to simplify the description.

Using this discussion we can reformulate our understanding of the complexity profile. We defined the profile using observers with progressively poorer ability to distinguish microstates. The fraction of the ensemble occupied by these states defined

the complexity. Using an algorithmic perspective we say, equivalently, that the observer cannot distinguish the true state from a state that has a smaller algorithmic complexity. An observer with a value of $g = 2$ cannot distinguish which of two states each particle occupies in the real microstate. Let us label the single particle states using an index that enumerates them. We can then imagine a checkerboard (in six dimensions of position and momentum) where odd indexed states are black and even ones are white. The observer cannot tell if a particle is in a black or a white state. Thus, no matter what the real state is,there is a simpler state where only odd (or only even) indexed states of the particles are occupied, which cannot be distinguished from the real system by the observer. The algorithmic complexity of this state with particles in odd indexed states is essentially the complexity that we determined above, $C(g = 2)$—it is the information necessary to specify this state out of all the states that have particles only in odd indexed states. Thus,in every case, we can specify the complexity of the system for the observer as the complexity of the simplest state that is consistent with the observations—by Occam's razor, this is the state that the observer will use to describe the system.

We note that this is also equivalent to defining the complexity profile as the length of the description as the error allowed in the description increases. The total error as a function of g for the ideal gas is

$$\frac{1}{2}\log\left(\prod \Delta x_i \Delta p_i / h\right) = \frac{1}{2} N \log(g) \tag{8.3.32}$$

where N is the number of particles in the system. The factor of 1/2 arises because the average error is half of the maximum error that could occur. This approach is helpful since it suggests how to generalize the complexity profile for systems that have different types of particles. We can define the complexity profile as a function of the number of errors that are made. This is better than using a particular length scale, which implies a different error for particles of different mass as indicated by Eq.(8.3.8). For conceptual simplicity, we will continue to write the complexity profile as a function of g or of length scale.

Nonequilibrium states Our next objective is to consider nonequilibrium states. When we have a nonequilibrium state,the microstate of the system is simpler than an equilibrium state to begin with. As we mentioned at the end of Section 8.3.2,there are nonequilibrium states that cannot be distinguished from equilibrium states on a macroscopic scale. These nonequilibrium states have microscopic correlations. Thus, the microscopic complexity is lower than the equilibrium entropy, while the macroscopic complexity is the same as in equilibrium:

$$\begin{aligned} C(g) < C_0(g) = S_0 \qquad & g = 1 \\ C(g) = C_0(g) \qquad & g >> 1 \end{aligned} \tag{8.3.33}$$

where we use the subscript 0 to indicate quantities of the equilibrium state. We illustrate this by an example. Using the indexing of single particle states we just introduced, we take a microstate where all particles are in odd indexed states. The mi-

crostate complexity is the same as that of an equilibrium state at $g = 2$, which is less than the entropy of the equilibrium system:

$$C(g = 1) = C_0(g = 2) < C_0(g = 1)$$

However, the complexity of this system for scales of observation $g \geq 2$ is the same as that of an equilibrium system—macroscopic observers do not distinguish them.

This scenario, where the complexity of a nonequilibrium state starts smaller but then quickly becomes equal to the equilibrium state complexity, does not always hold. It is true that the microscopic complexity must be less than or equal to the entropy of an equilibrium system, and that all systems have the same complexity when L is the size of the system. However, what we will show is that the complexity of a nonequilibrium system can be higher than that of the equilibrium system at large scales that are smaller than the size of the system. This is apparent in the case, for example, of a nonuniform density at large scales.

To illustrate what happens for such a nonequilibrium state, we consider a system that has nonuniformity that is characteristic of a particular length scale L_0, which is significantly larger than the microscopic scale λ but smaller than the size of the system. This means that n_i is smooth on finer scales,and there is no particular relationship between what is going on in one region of length scale L_0 and another. The values of n_i will be taken from a Gaussian distribution around the equilibrium value n_0 with a standard deviation of σ. We assume that σ is larger than the natural density fluctuations, which have a standard deviation of $\sigma_0 = \sqrt{n_0}$. For convenience we also assume that σ is much smaller than n_0.

We can calculate both the complexity $C(L)$, and the apparent entropy $S(L)$ for this system. We start by calculating them at the scale L_0. $C(L_0)$ is the amount of information necessary to specify the density values. This is the product of the number of cells V/L^d times the information in a number selected from a Gaussian distribution of width σ. From Question 8.3.3 this is:

$$C(L_0) = k \frac{V}{L_0^d}\left(\frac{1}{2}(1 + \ln(2\pi)) + \ln\sigma\right) \tag{8.3.34}$$

The number of microstates consistent with this macrostate at L_0 is given by the sum of ideal gas entropies in each region:

$$S(L_0) = -k \sum_i n_i \ln(n_i / g) + (5/2)kN \tag{8.3.35}$$

Since σ is less than n_0, this can be evaluated by expanding to second order in $\delta n_i = n_i - n_0$:

$$S(L_0) = S_0 - k \sum_i \frac{(\delta n_i)^2}{2n_0} = S_0 - \frac{kV\sigma^2}{2L_0^d n_0} \tag{8.3.36}$$

where S_0 is the entropy of the equilibrium system, and we used $\langle \delta n_i^2 \rangle = \sigma^2$. We note that when $\sigma = \sigma_0$ the logarithmic terms in the complexity reduce to the extra terms

found in Eq. (8.3.23). Thus, these terms are the information needed to describe the equilibrium fluctuations in the density.

We can understand the behavior of the complexity profile of this system. By construction, the minimum amount of information needed to specify the microstate is $C(\lambda) = S(L_0) + C(L_0)$. This is the sum over the entropy of equilibrium gases with densities n_i in volumes L_0^d, plus $C(L_0)$. Since $S(L_0)$ is linear in the number of particles, while $C(L_0)$ is logarithmic in σ and therefore logarithmic in the number of particles, we conclude that $C(L_0)$ is much smaller than $S(L_0)$. For $L > \lambda$ the complexity profile $C(L)$ decreases like that of an equilibrium ideal gas. The term $S(L_0)$ is eliminated at a microscopic length scale larger than λ but much smaller than L_0. However, $C(L_0)$ remains. Due to this term the complexity crosses that of an equilibrium gas to become larger. For length scales up to L_0 the complexity is essentially constant and equal to Eq.(8.3.34). Above L_0 it decreases to zero as L continues to increase by virtue of the effect of combining the different n_i into fewer regions. Combining the regions results in a Gaussian distribution with a standard deviation that decreases as the square root of the number of terms $\sigma \to \sigma(L_0/L)^{d/2}$. Thus, the complexity and entropy profiles for $L > L_0$ are:

$$\begin{aligned} C(L) &= k\frac{V}{L^d}\left(\frac{1}{2}(1+\ln(2\pi)) + \ln\sigma\left(L_0/L\right)^{d/2}\right) \\ S(L) &= S_0 - \frac{kV\sigma^2}{2(LL_0)^{d/2}n_0} \end{aligned} \tag{8.3.37}$$

This expression continues to be valid until there is only one region left,and the complexity goes to zero. The precise way the complexity goes to zero is not described by Eq. (8.3.37), since the Gaussian distribution does not apply in this limit.

There are several comments that we can make that are relevant to understanding complexity profiles in general. First we see that in order for the macroscopic complexity to be higher than that in equilibrium, the entropy at the same scale must be reduced $S(L) < S_0$. This is necessary because the sum $S(L) + C(L)$—the total information necessary to specify a microstate—cannot be greater than S_0. However, we also note that the reduction in $S(L)$ is much larger than the increase in $C(L)$. The ratio between the two is given by:

$$\frac{\delta S(L)}{\delta C(L)} = -\frac{\sigma^2}{2n_0}\frac{L^{d/2}}{L_0^{d/2}}\frac{1}{\ln(\sigma/\sigma_0)} \tag{8.3.38}$$

For $\sigma > \sigma_0 = \sqrt{n_0}$ this is greater than one. We can understand this result in two ways. First, a complex macroscopic system must be far from equilibrium, and therefore must have a much smaller entropy than an equilibrium system. Second, a macroscopic observer makes many errors in determining the microstate,and therefore if the microstate is similar to an equilibrium state,the observer cannot distinguish the two and the macroscopic properties must also be similar to an equilibrium state. For every bit of information that distinguishes the macrostate, there must be many bits of difference in the microstate.

In calculating the complexity of the system at a particular scale, we assumed that the observer was in error in obtaining the position and momentum of each particle. However, we assumed that the number of particles within each bin was determined exactly. Thus the complexity we calculated is the information necessary to specify the number of particles precise to the single particle. This is why even the equilibrium density fluctuations were described. An alternative, more reasonable, approach assumes that particle counting is also subject to error. For simplicity we can assume that the error is a fraction of the number of particles counted. For macroscopic systems this fraction is much larger than the equilibrium fluctuations, which therefore need not be described. This approach also modifies the form of the complexity profile of the nonuniform gas in Eq.(8.3.37). The error in measurement increases as $n_0(L)\quad L^d$ with the scale of observation. Letting $\gamma m_0(L)$ be the error in a measurement of particle number, we write:

$$C(L)=k\frac{V}{L^d}\left(\frac{1}{2}(1+\ln(2\pi))+\ln\frac{\sigma L_0^{d/2}}{\gamma n_0(L)L^{d/2}}\right)\approx k\frac{V}{L^d}\ln\frac{\sigma L_0^{3d/2}}{\gamma n_0(L_0)L^{3d/2}} \qquad (8.3.39)$$

The consequence of this modification is that the complexity decreases somewhat more rapidly as the scale of observation increases. The expression for the entropy in Eq. (8.3.37) is unchanged.

Question 8.3.3 What is the information in a number (character) selected from a Gaussian distribution of standard deviation σ?

Solution 8.3.3 Starting from a Gaussian distribution (Eq. 1.2.39),

$$P(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/2\sigma^2} \qquad (8.3.40)$$

we calculate the information (Eq. 8.2.2):

$$\begin{aligned} I &= -\int dxP(x)\log(P(x))=\int dxP(x)\left(\log(\sqrt{2\pi}\sigma)+\ln(2)x^2/2\sigma^2\right) \\ &=\log(\sqrt{2\pi}\sigma)+\ln(2)/2 \end{aligned} \qquad (8.3.41)$$

where the second term in the integral can be evaluated using $<x^2>=\sigma^2$.

We note that this result is to be interpreted as the information in a discrete distribution of integral values of x, like a random walk, that in the limit of large σ gives a Gaussian distribution. The units that are used to measure σ define the precision to which the values of x are to be described. It thus makes sense that the information to specify an integer of typical magnitude σ is essentially $\log(\sigma)$. ∎

8.3.4 *Time dependence—chaos and the complexity profile*

General approach In describing a system, we are interested in macroscopic observations over time, $n(x,t)$. As with the uncertainty in position, a macroscopic observer is not able to distinguish the time of observation within less than a certain time in-

terval $T = \Delta t$. To define what this means, we say that the system is represented by an ensemble with probability $P_{L,T}(n(x;t))$, or more generally $P_{L,T}(n(x,p;t))$. The different microstates that occur during the time interval T are all part of this ensemble. This may appear different than the definition we used for the spatial uncertainty. However, the definitions can be restated in a way that makes them appear equivalent. In this restatement we recognize that the observer performs measurements that are, in effect, averages over various possible microscopic measurements. The average measurements over space and time represent the system (or system ensemble) that is to be described by the observer. This representation will be discussed further in Section 8.3.6. The use of an ensemble is convenient because the observer may only measure one quantity, but we can consider various quantities that can be measured using the same degree of precision. The ensemble represents all possible measurements with this degree of precision. For example, the observer can measure correlations between particle positions that are fixed over time. If we averaged the density $n(x,t)$ over time, these correlations could disappear because of the movement of the whole system. However, if we average over the ensemble, they do not. We define the complexity profile $C(L,T)$ as the amount of information necessary to specify the ensemble $P_{L,T}(n(x,t))$. A description at a finer scale contains all of the information necessary to describe the coarser scale. Thus, $C(L,T)$ is a monotonic decreasing function of its arguments. A direct analysis is discussed in Question 8.3.4. We start, however, by considering the effect on $C(L,T)$ of prediction and the lack of predictability in chaotic dynamics.

Predictability and chaos As discussed earlier, a key ingredient in our understanding of physical systems is that the time evolution of an isolated system (or a system whose interactions with its environment are specified) can be obtained from the simple laws of mechanics starting from a complete microscopic description of the position and momenta of the particles. Thus, if we use a small enough L and T, so that each particle can be distinguished, we only need to specify $P_{L,T}(n(x,t))$ over a short period of time (or the simultaneous values of position and momentum) in order to predict the behavior over all subsequent times. The laws of mechanics are also reversible. We describe the past as well as the future from the description of a system at a particular time. This must mean that information is not lost over time. Systems that do not lose information over time are called conservative systems.

However, when we increase the spatial scale of observation, L, then the information loss—the complexity reduction—also limits the predictability of a system. We are not guaranteed that by knowing $P_{L,T}(n(x,t))$ at a scale L we can predict the system behavior. This is true even if we are only concerned about predicting the behavior at the scale L. We may need additional smaller-scale information to describe the time evolution of the system. This is precisely the origin of the study of chaotic systems discussed in Section 1.1. Chaotic systems take information from smaller scales and bring it to larger scales. Chaotic systems may be contrasted with dissipative systems that take information from larger scales to smaller scales. If we perturb (disturb) a dissipative system, the effect disappears over time. Looking at such a system at a particular time, we cannot tell if it was perturbed at some time far enough in the past. Since the information on a microscopic scale must be conserved, we know that the

information that is lost on the macroscopic scale must be preserved on the microscopic scale. In this sense we can say that information has been transferred from the macroscopic to the microscopic scale. For such systems, we cannot describe the past from present information on a particular length scale.

The degree of predictability is manifest when we consider that the complexity of a system $C(L,T)$ at a particular L and T depends also on the duration of the description—the limits of $t \quad [t_1,t_2]$. Like the spatial extent of the system, this temporal extent is part of the system definition. We typically keep these limits constant as we vary T to obtain the complexity profile. However, we can also characterize the dependence of the complexity on the time limits t_1,t_2 by determining the rate at which information is either gained or lost for a chaotic or stable system. For complex systems, the flow of information between length scales is bidirectional—even if the total amount of information at a particular scale is preserved, the information may change over time by transfer to or from shorter length scales. Unlike most theories of currents, information currents remain relevant even though they may be equal and opposite. All of the information that affects behavior at a particular length scale, at any time over the duration of the description, should be included in the complexity.

It is helpful to develop a conceptual image of the flow of information in a system. We begin by considering a conservative, nonchaotic and nondissipative system seen by an observer who is able to distinguish $2^{C(L)/k\ln(2)} = e^{C(L)/k}$ states. $C(L)/k\ln(2)$ is the amount of information necessary to describe the system during a single time interval of length T. For a conservative system the amount of information necessary to describe the state at a particular time does not change over time. The dynamics of the system causes the state of the system to change over time among these states. The sequence of states could be described one by one. This would require

$$N_T C(L)/k\ln(2) \tag{8.3.42}$$

bits, where $N_T = (t_2 - t_1)/T$ is the number of time intervals. However, we can also describe the state at a particular time (e.g., the initial conditions) and the dynamics. The amount of information to do this is:

$$(C(L) + C_t(L,T))/k\ln(2) \tag{8.3.43}$$

$C_t(L,T)/k\ln(2)$ is the information needed to describe the dynamics. For a nonchaotic and nondissipative system we can show that this information is quite small. We know from the previous section that the macrostate of the system of complexity $C(L)$ is consistent with a microstate which has the same complexity. The microstate has a dynamics that is simple, since it follows the dynamics of standard physical law. The dynamics of the simple microstate also describes the dynamics of the macrostate, which must therefore also be simple. Therefore Eq.(8.3.43) is smaller than Eq.(8.3.42) and the complexity is $C(L,T) = C(L) + C_t(L,T) \approx C(L)$. This holds for a system following conservative, nonchaotic and nondissipative dynamics.

For a system that is chaotic or dissipative, the picture must be modified to accommodate the flow of information between scales. From the previous paragraph we conclude that all of the interesting (complex) dynamics of a system is provided by in-

formation that comes from finer scales. The observer does not see this information before it appears in the state of the system—i.e.,in the dynamics. If we allow ourselves to see the finer-scale information we can track the flow of information that the observer does not see. In a conventional chaotic system,the flow of information can be characterized by its Lyaponov exponents. For a system that is described by a single real valued parameter, $x(t)$, the Lyaponov exponent is defined as an average over:

$$h = \ln((x'(t) - x(t))/(x'(t-1) - x(t-1))) \tag{8.3.44}$$

where unprimed and primed coordinates indicate two different trajectories. We can readily see how this affects the information needed by an observer to describe the dynamics. Consider an observer at a particular scale, L. The observer sees the system in state $x(t-1)$ at time $t-1$, but he determines $x(t-1)$ only within a bin of width L. Using the dynamics of the system that is assumed to be known, the observer can determine the state of the system at the next time. This extrapolation is not precise, so the observer needs additional information to specify the next location. The amount of information needed is the logarithm of the number of bins that one bin expands into during one time step. This is precisely $h/\ln(2)$ bits of information. Thus, the complexity of the dynamics for the observer is given by:

$$C(L,T) = C(L) + C_t(L,T) + N_T k h \tag{8.3.45}$$

where we have used the same notation as in Eq. (8.3.43).

A physical system that has many dimensions,like the microscopic ideal gas, will have one Lyaponov exponent for each of $6N$ dimensions of position and momentum. If the dynamics is conservative then the sum over all the Lyaponov exponents is zero,

$$\sum_i h_i = \log(\prod_i \Delta x_i(t)\Delta p_i(t) / \prod_i \Delta x_i(t-T)\Delta p_i(t-T)) = 0 \tag{8.3.46}$$

where $\Delta x_i(t) = x_i'(t) - x_i(t)$ and $\Delta p_i(t) = p_i'(t) - p_i(t)$. This follows directly from conservation of volumes of phase space in conservative dynamics. However, while the sum over all exponents is zero, some of the exponents may be positive and some negative. These correspond to chaotic and dissipative modes of the dynamics. We can imagine the flow of information as consisting of two streams, one going to higher scales and one to lower scales. The complexity of the system is given by:

$$C(L,T) = C(L) + C_t(L,T) + N_T k \sum_{i\, h_i>0} h_i \tag{8.3.47}$$

As indicated, the sum is only over positive values.

Two cautionary remarks about the application of Lyaponov exponents to complex physical systems are necessary. Unlike many standard models of chaos,a complex system does not have the same number of degrees of freedom at every scale. The number of independent bits of information describing the system above a particular scale is given by the complexity profile, $C(L)$. Thus,the flow of information between scales should be thought of as due to a number of closed loops that extend from a particular lowest scale up to a particular highest scale. As the scale increases,the complexity

decreases. Thus, so does the maximum number of Lyaponov exponents. This means that the sum over Lyaponov exponents is itself a function of scale. More generally, we must also be concerned that $C(L)$ can be time dependent,as it is in many irreversible processes.

The second remark is that over time the cycling of information between scales may bring the same information back more than once. Eq. (8.3.47) does not distinguish this,and therefore may include multiple counting of the same information. We should understand this expression as an upper bound on the complexity.

Time scale dependence Once we have chaotic behavior, we can consider various descriptions of the time dependence of the behavior seen by a particular observer. All of the models we considered in Chapter 1 are applicable. The state of the system may be selected at random from a particular distribution (ensemble) of states at successive time intervals. This is a special case of the more general Markov chain model that is described by a set of transition probabilities. Long-range correlations that are not easily described by a Markov chain may also be important in the dynamics.

In order to discuss the complexity profile as a function of T, we consider a Markov chain model. From the analysis in Question 8.3.4 we learn that the loss of complexity with time scale occurs as a result of cycles in the dynamics. These cycles need not be deterministic; they may be stochastic—cycles that do not repeat indefinitely but rather can occur one or more times through the probabilistic selection of successive states. Thus,a high complexity for large T arises when there is a large space of states with low chance of repetition in the dynamics. The highest complexity would arise from a deterministic dynamics with cycles that are longer than T. This might seem to contradict our previous conclusion, where the deterministic dynamics was found to be simple. However, a complex deterministic dynamics can arise if the successive states are specified by information from a smaller scale.

Question 8.3.4 Consider the information in a Markov chain of N_T states at intervals T_0 given by the transition matrix $P(s'|s)$. Assume the complexity of specifying the transition matrix—the complexity of the dynamics —$C_t = C(P(s'|s))$,is itself small.(See Question 8.3.5 for the case of a complex deterministic dynamics.)

a. Show that the more deterministic the chain is, the less information it contains.

b. Show that for an observer at a longer time scale consisting of two time steps ($T = 2T_0$) the information is reduced. Hint: Use convexity of information as described in Question 1.8.8, $f(\ x\ ) > \ f(x)$, for the function $f(x) = -x\log(x)$.

c. Show that the complexity does not decrease for a system that does not allow 2-cycles.

Solution 8.3.4 When the complexity of the dynamics is small, then the complexity of the Markov chain is given by:

$$C = C(s) + C_t + N_T k \ln(2) I(s'|s) \tag{8.3.48}$$

where the terms correspond to the information in the initial state of the system,the information in the dynamics and the incremental information per update needed to specify the next state. The relationship between this and Eq.(8.3.47) should be apparent. This expression does not hold if C_t is large, because ifit is larger than $N_T C(s)$,then the chain is more concisely described by specifying each of the states of the system (see Question 8.3.5).

The proof of (*a*) follows from realizing that the more deterministic the system is,the smaller is $I(s'|s)$. This may be used to define how deterministic the dynamics is.

To analyze the complexity of the Markov chain for an observer at time scale $2T_0$, we need to combine successive system states into an unordered pair—the ensemble of states seen by the observer. We use the notation $\{s', s\}$ for a pair of states. Thus, we are considering a new Markov chain of transitions between unordered pairs. To analyze this we need two probabilities:the probability of a pair and the transition probability from one pair to the next. The latter is the new transition matrix. The probability of a particular pair is:

$$P(\{s_1,s_2\}) = \begin{cases} P(s_1|s_2)P(s_2)+P(s_2|s_1)P(s_1) & s_2 \neq s_1 \\ P(s_1|s_1)P(s_1) & s_2 = s_1 \end{cases} \tag{8.3.49}$$

where $P(s)$ is the probability of a particular state of the system and the two terms in the upper line correspond to the probability of starting from s_1 to make the pair, and starting from s_2 to make the pair. The transition matrix for pairs is given by

$$\begin{aligned} P(\{s_1',s_2'\}|\{s_1,s_2\}) = \big[&\big(P(s_1'|s_2')P(s_2'|s_1)+P(s_2'|s_1')P(s_1'|s_1)\big)P(s_1|s_2)P(s_2) \\ &+\big(P(s_1'|s_2')P(s_2'|s_2)+P(s_2'|s_1')P(s_1'|s_2)\big)P(s_2|s_1)P(s_1)\big] / P(\{s_1,s_2\}) \end{aligned} \tag{8.3.50}$$

which is valid only for $s_1 \neq s_2$ and for $s_1' \neq s_2'$. Other cases are treated like Eq.(8.3.49). Eq.(8.3.50) includes all four possible ways of generating the sequence of the two pairs. The normalization is needed because the transition matrix is the probability of $\{s_1' \neq s_2'\}$ occurring, assuming the pair $\{s_1, s_2\}$ has already occurred.

To show (*b*) we must prove that the process of combining the states into pairs reduces the information necessary to describe the chain. This is apparent since the observer loses the information about the state order within each pair. To show it from the equations, we note from Eq.(8.3.49) that the probability of a particular pair is larger than or equal to the probability of each of the two possible unordered pairs. Since the probabilities are larger, the information is smaller. Thus the information contained in the first pair is smaller for $T = 2$ than for $T = 1$. We must show the same result for each successive pair. The transition probability can be seen to be an average over two terms in the round parenthesis. By convexity, the information in the average is less than the average information of each term.Each of the terms is a sum

over the probabilities of two possible orderings, and is therefore larger than or equal to the probability of either ordering. Thus,the information needed to specify any pair in the chain is smaller than the corresponding information in the chain of states.

Finally, to prove (*c*) we note that the less the order of states is lost when we combine states into pairs, the more complexity is retained. If transitions in the dynamics can only occur in one direction,then we can infer the order and information is not lost. Thus, for $T = 2$ the complexity is retained if the dynamics is not reversible—there are no 2-cycles. From the equations we see that if only one of $P(s_1|s_2)$ and $P(s_2|s_1)$ can be nonzero, and similarly for $P(s'_1|s'_2)$ and $P(s'_2|s'_1)$, then only one term survives in Eq. (8.3.49) and Eq. (8.3.50) and no averaging is performed. For arbitrary T the complexity is the same as at $T = 1$ if the dynamics does not allow loops of size less than or equal to T. ∎

Question 8.3.5 Calculate the maximum information that might in principle be necessary to specify completely a deterministic dynamics of a system whose complexity at any time is $C(L)$. Contrast this with the maximum complexity of describing N_T steps of this system.

Solution 8.3.5 The number of possible states of the system is $2^{C(L)/k\ln(2)}$. Each of these must be assigned a successor by the dynamics. The maximum possible information to specify the dynamics arises if there is no algorithm that can specify the successor, so that each successor must be identified out of all possible states. This would require $2^{C(L)/k\ln(2)}C(L)/k\ln(2)$ bits.

The maximum complexity of N_T steps is just $N_T C(L)$, as long as this is smaller than the previous result. Which is generally a reasonable assumption. ∎

A simple example of chaotic behavior that is relevant to complex systems is that of a mobile system—an animal or human being—where the motion is internally directed.A description of the system behavior, even at a length scale larger than the system itself, must describe this motion. However, the motion is determined by information contained on a smaller length scale just prior to its occurrence. This satisfies the formal requirements for chaotic behavior regardless of the specifics of the motion involved. Stated differently, the large-scale motion would be changed by modifications of the internal state of the system. This is consistent with the sensitivity of chaotic motion to smaller scale changes.

Another example of information transfer between different scales is related to adaptability, which requires that information about the external environment be represented in the organism.This generally involves the transfer of information between a larger scale and a smaller scale.Specifically, between observed phenomena and their representation in the synapses of the nervous system.

When we describe a system at a particular moment of time,the complexity of the system at its own scale or larger is zero—or a constant if we include the description of the equilibrium system. However, when we consider the description of a system over

time, then the complexity is larger due to the system motion. Increasing the scale of observation continues to result in a progressive decrease in complexity. At a scale that is larger than the system itself, it is the motion of the system as measured by its location at successive time intervals that is to be described. As the scale becomes larger, smaller scale motions are not observed,and a simpler description of motion is possible. The observer only notes changes in position that are larger than the scale of observation.

A natural question that can be asked in this context is whether the motion of the system is due to external influences or due to the system itself. For example,a particle moving in a fluid may be displaced by the motion of the fluid. This should be considered different from a mobile bacteria. Similarly, a basketball in a game moves through its trajectory not because of its own volition, but rather because of the volition of the players. How do we distinguish this from a system that moves due to its own actions? More generally, we must ask how we must deal with the environmental influences for a system that is not isolated. This question will be dealt with in Section 8.3.6 on behavioral complexity. Before we address this question,in the next section we discuss several aspects of the complexity profile, including the relationship of the complexity of the whole to the complexity of its parts.

8.3.5 *Properties of complexity profiles of systems and components*

General properties We can readily understand some of the properties that we would expect to find in complexity profiles of systems that are difficult to calculate directly. Fig. 8.3.2 illustrates the complexity profile for a few systems. The paragraphs that follow describe some of their features.

For any system, the complexity at the smallest values of L, T is the microscopic complexity—the amount of information necessary to describe a particular microstate. For an equilibrium state this is the same as the thermodynamic entropy, which is the entropy of a system observed on an arbitrarily long time scale.This is not true in general because short-range correlations decrease the microstate complexity, but do not affect the apparent macroscopic entropy. We have thus also defined the entropy profile $S(L,T)$ as the amount of information necessary to determine an arbitrary microstate consistent with the observed macrostate. From our discussion of nonergodic systems in Section 8.3.1 we might also conclude that at any scale L, T the sum of the complexity $C(L,T)$ and the entropy $S(L,T)$ of the system (the fast degrees of freedom) should add up to the microscopic complexity or macroscopic entropy

$$C(0,0) \approx S(\infty,\infty) \approx C(L,T) + S(L,T) \tag{8.3.51}$$

However, this is valid only under special circumstances—when the macroscopic state is selected at random from the ensemble of macrostates,and the microstate is selected at random from the possible microstates.A glass may satisfy this requirement; however, other complex systems need not.

For a typical system in equilibrium, as L,T is increased the system rapidly becomes homogeneous in space and time. Specifically, the density of the system is

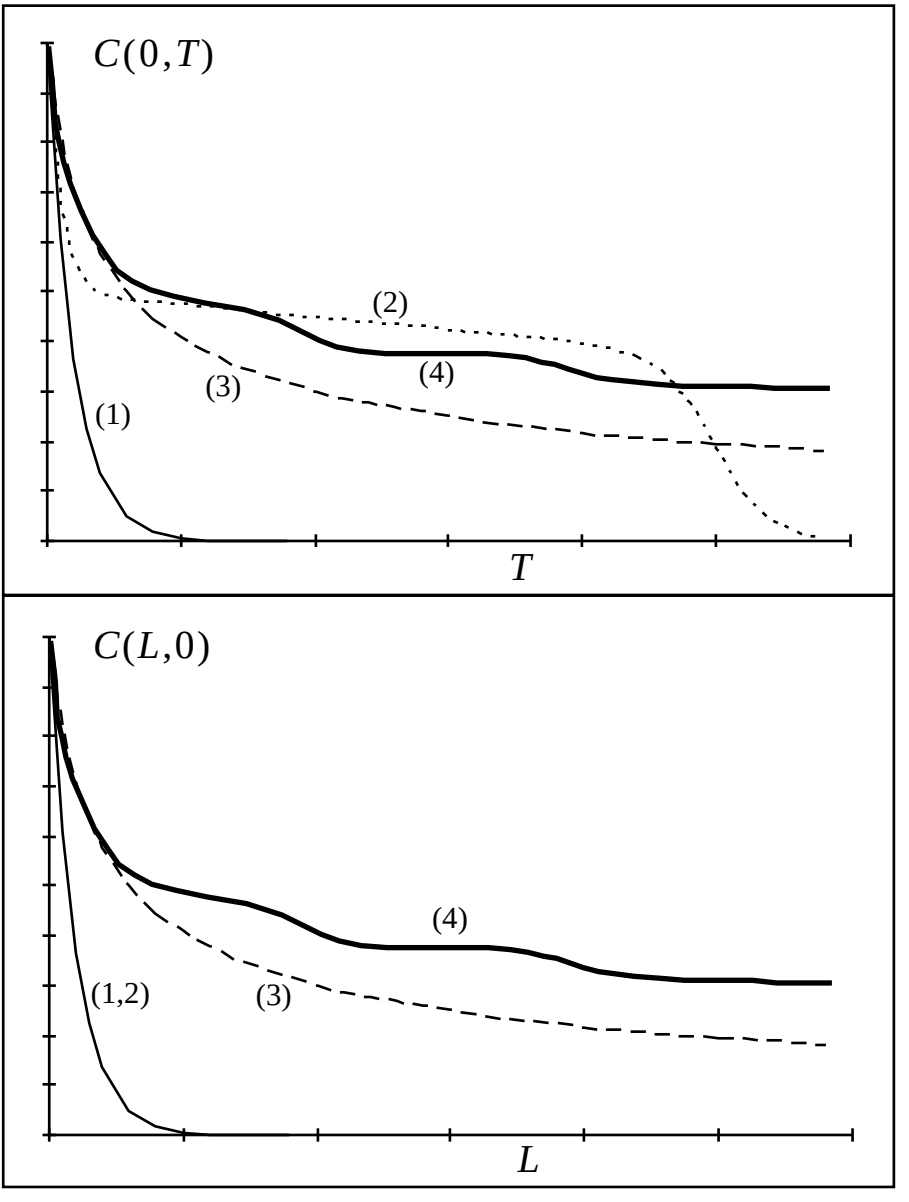
C(0,T)
(1)
(2)
(3)
(4)
T
C(L,0)
(1,2)
(3)
(4)
L

Figure 8.3.2 Schematic plots of the complexity profile $C(L,T)$ of four different systems. $C(L,T)$ is the amount of information necessary to describe the system ensemble as a function of the length scale, L, and time scale, T, of observation. Top panel shows the time scale dependence, bottom panel shows the length scale dependence. (1) An equilibrium system has a complexity profile that is sharply peaked at $T = 0$ and $L = 0$. Once the length or time scale is beyond the correlation length or correlation time respectively, the complexity is just the macroscopic complexity associated with thermodynamic quantities (U,N,V), which vanishes on any reasonable scale. (2) For a glass the complexity profile as a function of time scale $C(0,T)$ decays rapidly at first due to averaging over atomic vibrations; it then reaches a plateau that represents the frozen degrees of freedom. At much longer time scales the complexity profile decays to its thermodynamic limit. Unlike $C(0,T)$, $C(L,0)$ of a glass decays like a thermodynamic system because it is homogeneous in space. (3) A magnet at a second-order phase transition has a complexity profile that follows power-law behavior in both length and time scale. Stochastic fractals capture this kind of behavior. (4) A complex biological organism has a complexity profile that should follow similar behavior to that of a fractal. However it has plateau-like regions that correspond to crossing the scale of internal components, such as molecules and cells. ∎

uniform in space and time,aside from unobservable small fluctuations, once the scale of observation is larger than either the correlation length or the correlation time of the system. Indeed, this might be taken to be the definition of the correlation length and time—the scale at which the microscopic information becomes irrelevant to the properties of the system. Beyond the correlation length,the average behavior characteristic of the macroscopic scale is all that remains,and the complexity profile is constant at all length and time scales less than the size of the system.

We can contrast the complexity profile of a thermodynamic system with what we expect from various complex systems. For a glass, the complexity profile is quite different in time and in space. A typical glass is uniform if L is larger than a microscopic correlation length. Thus, the complexity profile of the glass is similar to an equilibrium system as a function of L. However, it is different as a function of T. The frozen degrees of freedom that make it a nonergodic system at typical time scales of observation guarantee this. At typical values of T the temporal ensemble of the system includes the states that are reached by vibrational modes of the system, but not the atomic rearrangements characteristic of fluid motion. Thus, the atomic vibrations cannot be observed except at microscopic values of T. However, a significant part of the microscopic description remains necessary at longer time scales. Correspondingly, a plateau in the complexity profile extends up to characteristic time scales of human observation. At a temperature-dependent and much longer time scale, the complexity profile declines to its thermodynamic limit. This time scale, the relaxation time, is accessible near the glass transition temperature. For lower temperatures it is not. Because the glass is uniform in space, the plateau should be relatively flat and end abruptly. This is because spatial uniformity indicates that the relaxation time is essentially a local property with a narrow distribution. A more extended spatial coupling would give rise to a grading of the plateau and a broadening of the time scale at which the plateau disappears.

More generally, for a complex system we expect that many parameters will be required to describe its properties at all length and time scales, at least up to some fraction of the spatial and temporal scale of the system itself. Starting from the microscopic complexity, the complexity profile should not be expected to fall smoothly. In biological organisms, we can expect that as we increase the scale of observation, there will be particular length scales at which details will be lost. Plateaus in the profile are related to the existence of well-defined levels of description. For example, an identifiable level of cellular behavior would correspond to a plateau, because over a range of length scales larger than the cell, a full accounting of cellular properties, but not of the internal behavior of the cell, must be given. There are many cells that have a characteristic size and are immobile. However, because different cell populations have different sizes and some cells are mobile, the sharpness of the transition should be smoothed. We can at least qualitatively identify several different plateaus. At the shortest time scale the atomic vibrations will be averaged out to end the first plateau. Larger atomic motions or molecular behavior will be averaged out on a second, larger scale. The internal cellular behavior will then be averaged out. Finally, the internal behavior of tissues and organs will be averaged out on a still longer length and time scale. It is the degrees of freedom that remain relevant on the longest length scale that are key to the complexity of the system. These degrees of freedom manifest the concept of emergent collective behavior. Ultimately, they must be traceable back to the microscopic degrees of freedom. Describing the connection between the microscopic parameters and macroscopically relevant parameters has occupied our attention in much of this book.

Mathematical models that best capture the complexity profile of a complex system are fractals (see Section 1.10). Mathematical fractals with no granularity (no smallest length scale) have infinite complexity. However, if we define a smallest length scale, corresponding to the atomic length scale of a physical system, and we define a longest length scale that is the size of the system, then we can plot the spatial complexity profile of a fractal-like system. There are two quite distinct kinds of mathematical fractals, deterministic and stochastic fractals. The deterministic fractals are specified by an algorithm with only a few parameters, and thus their algorithmic complexity is small. Examples are the Kantor set or the Sierpinski gasket. The algorithm describes how to create finer and finer scale detail. The only difficulty in specifying the fractal is specifying the number of levels to which the algorithm should be iterated. This information (the number of iterations) requires a parameter whose length grows logarithmically with the ratio of the size of the system to the smallest length scale. Thus, a deterministic fractal has a complexity profile that decreases logarithmically with observation length scale L, but is very small on all length scales.

Stochastic fractals are qualitatively different. In such fractals, there are random choices made at every scale of the structure. Stochastic fractals can be based upon the Kantor set or Sierpinski gasket, by including random choices in the algorithm. They may also be systems representing the spatial structure of various stochastic processes. Such a system requires information to describe its structure on every length scale. A stochastic fractal is a member of an ensemble, and its algorithmic as well as ensemble complexity will scale as a power law of the scale of observation L. As L increases, the

amount of information is reduced, but there is no length scale smaller than the size of the system at which it is completely lost. Time series that have fractal behavior—that have power-law correlations—would also display a power-law dependence of their complexity profile as a function of T. The simplest physical model that demonstrates such fractal properties in space and time is an Ising model at its second-order transition point. At this transition there are fluctuations on all spatial and temporal scales that have power-law behavior in both. Observers with larger values of L can see the behavior of the correlations only on the longer length scales. A renormalization treatment, discussed in Section 1.10, can give the value of the complexity profile. These examples illustrate how microscopic information may become irrelevant on larger length scales, while leaving collective information that remains relevant at the longer scales.

The complexity profile enables us to consider again the definition of a complex system. As we stated, it seems intuitive that a complex system is complex on many scales. This strengthens the identification of the fractal model of space and time as a central model for the understanding of complex systems. We have also gained an understanding of the difference between deterministic and stochastic fractal systems. We see that the glass is complex in its temporal behavior, but not in its spatial behavior, and therefore is only a partial example of a complex system. If we want to identify a unique complexity of a system, there is a natural space and time scale at which to define it. For the spatial scale, L_s, we consider a significant fraction of the system—one-tenth of its size. For the temporal scale, T_s, we consider the relaxation (autocorrelation) time of the behavior on this same length scale. This is essentially the maximal complexity for this length scale, which would be the same as setting $T = 0$. However, we could also take a natural time scale of $T_s = L_s / v_s$ where v_s is a characteristic velocity of the system. This form makes the increase in time scale for larger length scales (systems) apparent. Leaving out the time scale,since it is dependent on the space scale, we can write the complexity of a system s as

$$C_s = C_s(L_s) = C_s(L_s, L_s / v_s) \approx C_s(L_s, 0) \tag{8.3.52}$$

In Section 1.10 we discussed generally the scaling of quantities as a function of the precision to which we describe the system.One of the central questions in the field of complex systems is understanding how complexity scales. This scaling is concretized by the complexity profile.One of the objectives is to understand the ultimate limits to complexity. Given a particular length or time scale, we ask what is the maximum possible complexity at that scale.One could say that this complexity is limited by the thermodynamic entropy; however, there are further limitations. These limitations are established by the nature of physical law that establishes the dynamics and interactions of the components. Thus it is unlikely that atoms can be attached to each other in such a way that the behavior of each atom is relevant to the spatiotemporal behavior of an organism at the length and time scale relevant to a human being. The details of behavior must be lost as we observe on longer length and time scales; this results in a loss of complexity. The complexity scaling of complex organisms should follow a line like that given in Fig. 8.3.2. The highest complexity of an organism results

from the retention of the greatest significance of details. This is in contrast to thermodynamic systems, where all of the degrees of freedom average out on a very short length and time scale. At this time we do not know what limits can be placed on the rate of decrease of complexity with scale.

Components and systems As we discussed in Chapter 2, a complex system is formed out of a hierarchy of interdependent subsystems. Thus, relevant to various questions about the complexity profile is an understanding of the complexity that may arise when we bring together complex systems to form a larger complex system. In general it is not clear that bringing together many complex systems must give rise to a collective complex system. This was discussed in Chapter 6, where one example was a flock of animals. Here we can provide additional meaning to this statement using the complexity profile. We will discuss the relationship of the complexity of components to the complexity of the system they are part of. To be definite, we can consider a flock of sheep. The example is chosen to expand our view toward more general application of these ideas. The general statements we make apply to any system formed out of subsystems.

Let us assume that we know the complexity of a sheep, $C_{sheep}(L_{sheep})$, the amount of information necessary to describe the relevant behaviors of eating, walking, reproducing, flocking, etc.,at a length scale of about one-tenth the size of the sheep. For our current purposes this might be a lot of information contained in a large number of books, or a little information contained in a single paragraph of text.Later, in Section 8.4, we will obtain an estimate of the complexity as, of order, one book or 10^7 bits.

We now consider a flock of N sheep and construct a description of this flock. We begin by taking information that describes each of the sheep. Combining these descriptions, we have a description of the flock. This information is,however, highly redundant. Much of the information that describes one sheep can also be used to describe other sheep. Of course there are differences in size and in behavior. However, having described one sheep in detail we can describe the differences, or we can describe general characteristics of sheep and then specialize them for each of the individual sheep. Using this strategy, a description of the flock will be shorter than the sum of the lengths of the descriptions of each of the sheep. Still, this is not what we really want. The description of the flock behavior has to be on its own length scale L_{flock}, which is much larger than L_{sheep}. So we shift our observation of behavior to this longer length scale and find that most of the details of the individual sheep behavior have become irrelevant to the description of the flock. We describe the flock behavior in terms of sheep density, grazing activity, migration, reproductive rates, etc. Thus we write that:

$$C_{flock} = C_{flock}(L_{flock}) << C_{flock}(L_{sheep}) << NC_{sheep}(L_{sheep}) = NC_{sheep} \qquad (8.3.53)$$

where N is the number of sheep in the flock. Among other conclusions, we see that the complexity of a flock may actually be smaller than the complexity of one sheep.

More generally, the relationship between the complexity of the collective complex system and the complexity of component systems is crucially dependent on the existence of coherence and correlations in the behavior of the components that can arise either from common origins for the behavior or from interactions between the

components. We first describe this qualitatively by considering the two inequalities in Eq. (8.3.53). The second inequality arises because different sheep have the same behavior. In this case their behavior is coherent. The first inequality arises because we change the scale of observation and so lose the behavior of an individual sheep. There is a trade-off between these two inequalities. If the behaviors of the sheep are independent,then their behavior cannot be observed on the longer scale.Specifically, the movement of one sheep to the right is canceled by another sheep that starts at its right and moves to the left. Thus, only correlated motions of many sheep can be observed on a longer scale.On the other hand,if their behaviors are correlated,then the complexity of describing all of them is much smaller than the sum of the separate complexities. Thus, having a large collective complexity requires a balance between dependence and independence of the behavior of the components.

We can discuss this more quantitatively by considering the example of the nonuniform ideal gas. The loss of information for uncorrelated quantities due to combining them together is described by Eq.(8.3.37). To construct a model where the quantities are correlated, we consider placing the same densities in a region of scale $L_1 > L_0$. This is the same model as the previous one, but now on a length scale of L_1. The new value of σ is $\sigma_1 = \sigma(L_1/L_0)^d$. This increase of the standard deviation causes an increase in the value of the complexity for all scales greater than L_1. However, for $L < L_1$ the complexity is just the complexity at L_1, since there is no structure below this scale. A comparative plot is given in Fig. 8.3.3.

We can come closer to considering the behavior of a collection of animals by considering a model for their motion. We start with a scale L_0 just larger than the animal, so that we do not describe its internal structure—we describe only its location at successive intervals of time. The characteristic time over which a sheep moves a distance L_0 is T_0. We will use a model for sheep motion that can illustrate the effect of coherence of many sheep, as well as the effect of coherent motion of an individual sheep over time. To do this we assume that an individual sheep moves in a straight line for a distance qL_0 in a time qT_0 before choosing a new direction to move in at random. For simplicity we can assume that the direction chosen is one of the four compass directions, though this is not necessary for the analysis. We will use this model to calculate the complexity profile of an individual sheep. Our treatment only describes the leading behavior of the complexity profile and not various corrections.

For $L = L_0$ and $T = T_0$, the complexity of describing the motion is exactly 2 bits for every q steps to determine which of the four possible directions the sheep will move next. Because the movement is in a straight line, and the changes in direction are at well-defined intervals, we can reconstruct the motion from the measurements of any observer with $L < qL_0$ and $T < qT_0$. Thus the complexity is:

$$C(L,T) = 2N_T/q \qquad L < qL_0,\ T < qT_0 \tag{8.3.54}$$

Once the scale of observation is greater than qL_0, the observer does not see every change in direction. The sheep is moving in a random walk where each step has a length qL_0 and takes a time qT_0, but the observer does not see each step. The distance traveled is proportional to the square root of the time,and so the sheep moves a dis-

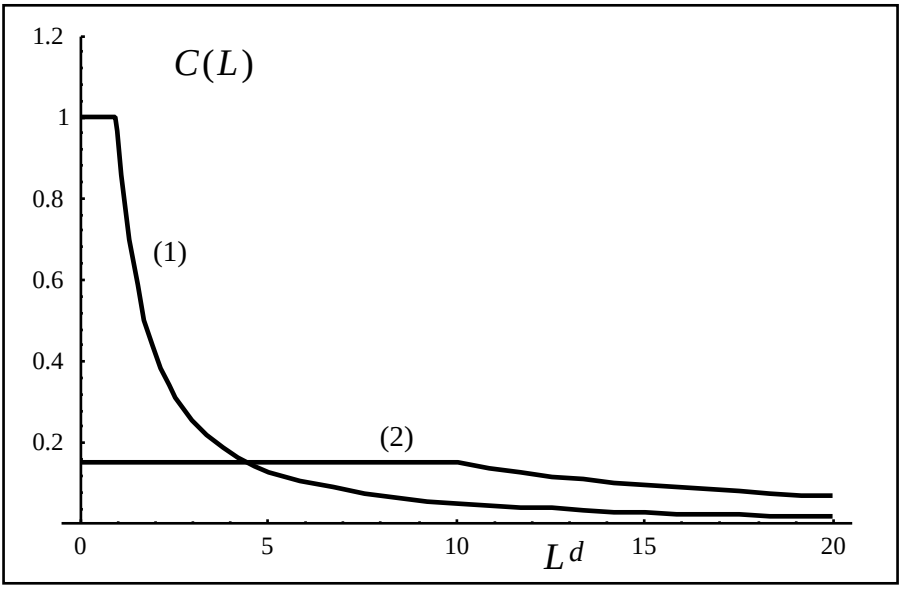

Figure 8.3.3 Plot of the complexity of a nonuniform gas (Eq. (8.3.37)), for two cases. The first (1) has a correlation in its nonuniformity at a scale L_0 and the second (2) at a scale $L_1 > L_0$. The magnitude of the local deviations in the density are the same in the two cases. The second case has a lower complexity at smaller scales but a higher complexity at the larger scales. Because the complexity decreases rapidly with scale, to show the effects on a linear scale L_1 was taken to be only $\sqrt[3]{10}L_0$, and the horizontal axis is in units of L^3 measured in units of L_0^3. Eq (8.3.39) would give similar results but the complexity would decay still more rapidly. ∎

tance L once in every $(\sigma_0/L)^2$ steps, where $\sigma_0 = qL_0$ is the standard deviation of the random walk in each dimension. Every time the sheep travels a distance L we need 2 bits to describe its motion, and thus we have a complexity:

$$C(L,T) = 2\frac{N_T}{q}\frac{\sigma_0^2}{L^2} = 2N_T\frac{qL_0^2}{L^2} \qquad L > qL_0,\ T < qT_0 \tag{8.3.55}$$

We note that at $L = qL_0$ Eq. (8.3.54) and Eq. (8.3.55) are equal.

To obtain the complexity profile for long times scales $T > qT_0$, but short length scales $L < qL_0$, we use a simplified "blob" picture to combine the successive positions of the sheep into an ensemble of positions. For T only a few times qT_0 we can expect that the ensemble would enable us to reconstruct the motion—the complexity is the same as Eq.(8.3.54). However, eventually the ensemble of positions will overlap and form a blob. At this point the movement of the sheep will be described by the movement of the blob, which itself undergoes a random walk. The standard deviation of this random walk is proportional to the square root of the number of steps:

$\sigma = \sigma_0\sqrt{T/qT_0}$. Since this is larger than L, the amount of information is essentially that of selecting a value from a Gaussian distribution of this standard deviation:

$$C(L,T) = 2\frac{N_T}{q}\min(1, \frac{qT_0}{T}(1+\log(\frac{\sigma}{L})) = 2\frac{N_T}{q}\min(1, \frac{qT_0}{T}(1+\log(\frac{L_0}{L}\sqrt{\frac{qT}{T_0}})) \qquad L<\sigma,\ T>qT_0 \qquad (8.3.56)$$

There are a few points to be made about this expression. First, we use the minimum of two values to select the crossover point between the behavior in Eq. (8.3.54) and the blob behavior. As we mentioned above, the blob behavior only occurs for T significantly greater than qT_0. The simplest way to identify the crossover point is when the new estimate of the complexity becomes lower than our previous value. The second point is that we have chosen to adjust the constant term added to the logarithm so that when $L = \sigma$ the complexity matches that given by Eq.(8.3.55), which describes the behavior when L becomes large. Thus the limit on Eq.(8.3.55) should be generalized to $L > \sigma$. This minor adjustment enables the complexity to be continuous despite our rough approximations, and does not change any of the conclusions.

We can see from our results (Fig. 8.3.4) how varying q affects the complexity. Increasing q decreases the complexity at the scale of a sheep, $C(L,T) \quad 1/q$ in Eq. (8.3.54). However, it increases the complexity at longer scales $C(L,T) \quad q$ in Eq.(8.3.55). This is a straightforward consequence of increasing the coherence of the motion over time. We also see that the complexity at long times decays inversely proportional to the time but is relatively insensitive to q. The value of q primarily affects the crossover point to the long time behavior.

We now use two different assumptions to calculate the complexity of the flock. If the movement of all of the sheep is coherent, then the complexity of the flock for length scales greater than the size of the flock is the same as the complexity of a sheep for the same length scales. This is apparent because describing the movement of a single sheep is the same as describing the entire flock. We now see the significance of increasing q. Increasing q increases the flock complexity until qL_0 reaches L_1, where L_1 is the size of the flock. Thus we can increase the complexity of the whole at the cost of reducing the complexity of the components.

If the movement of sheep are independent of each other, then the flock displacements—the displacements of its center of mass—are of characteristic size $\sigma/\sqrt{N}$ (see Eq.5.2.21). We might be concerned that the flock will disperse. However, as in our discussions of polymers in Section 5.2, interactions that would keep the sheep together need not affect the motion of their center of mass. We could also introduce into our model a circular reflecting boundary (a moving pen) around the flock, with its center at the center of mass. Since the motion of the sheep with this boundary does not require additional information over that without it,the complexity is the same. In either case, the complexity of flock motion ($L > L_1$) is obtained as:

$$C(L,T) = 2N_T\frac{qL_0^2}{NL^2} \qquad L>\sigma \qquad (8.3.57)$$

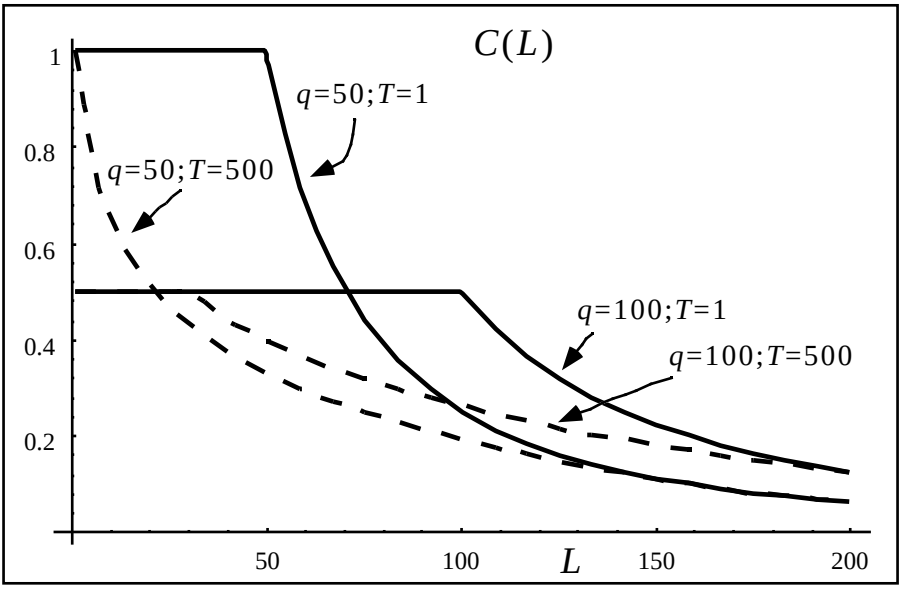

Figure 8.3.4 The complexity profile is plotted for a model of the movement of sheep as part of a flock. Increasing the distance a sheep moves in a straight line (coherence of motion in time), q, decreases the complexity at small length scales and increases the complexity at large length scales. Solid lines and dashed lines show the complexity profile as a function of length scale for a time scale $T = 1$ and $T = 500$ respectively. ∎

This is valid for all L if σ is less than L_1. If we choose T to be very large, Eq. (8.3.56) applies, with σ replaced by $\sigma/\sqrt{N}$. We see that when the motion of sheep are independent,the flock complexity is much lower than before—it decreases inversely with the number of sheep when $L > \sigma$. Even in this case, however, increasing q increases the flock complexity. Thus coherence in the behavior of a single sheep in time, or coherence between different sheep, increases the complexity of the flock. However, the maximum complexity of the flock is just that of an individual sheep, and this arises only for coherent behavior when all movements are visible on the scale of the flock. Any movements of an individual sheep that are smaller than the scale of the flock disappear on the scale of the flock. Thus even for coherent motion, in general the flock complexity is smaller than the complexity of a sheep.

This example illustrates the effect of coherent behavior. However, we see that even with coherent motion the complexity of a flock at its scale cannot be larger than the complexity of the sheep at its own scale. This is a problem for us, because our study of complex systems is focused upon systems whose complexity is larger than their components. Without this possibility, there would be no complex systems. To obtain a higher complexity of the whole we must modify this model. We must assume

more generally that the motion of a sheep is describable using a set of patterns of behavior. Coherent motion of sheep still lead to a similar (or lower) complexity. To increase the complexity, the motion of the flock must have more complex patterns of motion. In order to achieve such patterns, the motions of the individual sheep must be neither independent nor coherent—they must be correlated motions that combine patterns of sheep motion into the more complex patterns of flock motion. This is possible only if there are interactions between them, which have not been included here. It should now be clear that the objective of learning how the complexity of a system is related to the complexity of its components is central to our study of complex systems.

Question 8.3.6 Throughout much of this book our working definition of complex systems or complex organisms as articulated in Section 1.3 and developed further in Chapter 2 was that a complex system has a behavior that is dependent on all of its parts. In particular, that it is impossible to take part of a complex organism away without affecting the behavior of the whole and behavior of the part. How is this definition related to the definition of complexity articulated in this section?

Solution 8.3.6 Our quantitative concept of complexity is a measure of the information necessary to describe the system behavior on its own length scale. If the system behavior is complex,then it must require many parameters to describe. These parameters are related to the description of the system on a smaller length scale, where the parts of the system are manifest because we can distinguish the description of one part from another. To do this we limit $P_{L,T}(n(x,t))$ to the domain of the part. The behavior of a system is thus related to the behavior of the parts. The more these are relevant to the system behavior, the greater is the system complexity. The information that describes the system behavior must be relevant on every smaller length scale. Thus, we have a direct relationship between the definition of a complex system in terms of parts and the definition in terms of information. Ultimately, the information necessary to describe the system behavior is determined by the microscopic description of atomic positions and motions. The more complex a system is, the more its behavior depends on smaller scale components. ∎

Question 8.3.7 When we defined interdependence we did not consider the dependence of an animal on air as a relevant example. Explain.

Solution 8.3.7 We can now recognize that the use of information as a characterization of behavior enables us to distinguish various forms of dependency. In particular, we see that the dependence of an animal on air is simple, since the necessary properties of air are simple to describe. Thus, the degree of interdependence of two systems should be measured as the amount of information necessary to replace one in the description of the other. ∎

8.3.6 *Behavioral complexity*

Our ability to describe a system arises from measurements or observations of its behavior. The use of system descriptions to define system complexity does not directly take this into account. The complexity profile brought us closer by acknowledging the observer in the space and time scale of the description. By acknowledging the scale of observation, we obtained a mechanism for distinguishing complex systems from equilibrium systems,and a systematic method for characterizing the complexity of a system. There is another approach to reaching the complexity profile that incorporates the observer and system relationship in a more satisfactory manner. It also enables us to consider directly the interaction of the system with its environment, which was not included previously. To introduce the new approach, we return to the underpinning of descriptive complexity and present the concept of behavioral complexity.

In Shannon's approach to the study of information in communication systems, there were two quantities of fundamental interest. The first was the information content of an individual message, and the second was the average information provided by a particular source. The discussion of algorithmic complexity was based on a consideration of the information provided by a particular message—specifically, how much it could be compressed. This carried over into our discussion of physical systems when we introduced the microscopic complexity of a system as the information contained in a particular microscopic realization of the system. When all messages, or all system states, have the same probability, then the information in the particular message is the same as the average information, and we can write:

$$I(\{x,p\}|(U,N,V)) = -\log P(\{x,p\}) = -\sum_{\{x,p\}} P(\{x,p\})\log(P(\{x,p\})) \quad (8.3.58)$$

The expression on the right, however, has a different purpose. It is a quantity that characterizes the ensemble rather than the individual microstate. It is a characterization of the source rather than of any particular message.

We can pursue this line of reasoning by considering more carefully how we might characterize the source of the information, rather than the messages.One way to characterize the source is to determine the average amount of information in a message. However, if we want to describe the source to someone, the most essential information is to give a description of the kinds of messages that will be received—the ensemble of possible messages. Thus to characterize the source we need a description of the probability of each kind of message. How much information do we need to describe these probabilities? We call this the behavioral complexity of the source.

A few examples in the context of a source of messages will serve to illustrate this concept. Any description of a source must assume a language that is to be used. We assume that the language consists of a list of characters or messages that can be received from the source, along with their probabilities.A delimiter (:) is used to separate the messages from their probability. For convenience, we will write probabilities in decimal notation. A second delimiter (,) is used to separate different members of the list.A source that gives zeros and ones at random with equal probability would be described by {1:0.5,0:0.5}. It is convenient to include the length of a message in our

description of the source. Thus we might describe a source with length $N = 1000$ character messages, each character zero and one with equal probability, as: {1000(1:0.5, 0:0.5)}. The message complexity of this source would be given by N, the length of a message. However, the behavioral complexity is given by (in this language): two decimal digits,two characters (1, 0),the number representing N (requiring $\log(N)$ characters) and several delimiters. We could also specify an ASCII language source by a table of this kind that would consist of 256 elements and the probabilities of their occurrence in some database. We see that the behavioral complexity is quite distinct from the complexity of the messages provided by a source. In particular in the above example it can be larger, if $N = 1$, or it can be much smaller, if N is large.

This definition of the behavioral complexity of a source runs into a minor problem, because the probabilities are real numbers and would generally require arbitrary numbers of digits to describe. To overcome this problem,there must be a convention assumed about the limit of precision that is desired in describing the source. In principle,this precision is related to the number of messages that might be received. This convention could be part of the language, or could be defined by the specification itself. The description of the source can also be compressed using the principles of algorithmic complexity.

As we found above,the behavioral complexity can be much smaller than the information complexity of a particular message—if the source provides many random digits, the complexity of the message is high but the complexity of the source is low because we can characterize it simply as a source of random numbers. However, if the probability of each message must be independently specified, the behavioral complexity of a source is much larger than the information content of a particular message. If a particular message requires N bits of information,then the number of possible messages is 2^N. Listing all of the possible messages requires $N\,2^N$ bits, and specifying each probability with Q bits would give us a total of $(N + Q)2^N$ bits to describe the source. This could be reduced if the messages are placed in an agreed-upon order; then the number of bits is $Q2^N$. This is still exponentially larger than the information in a particular message. Thus, the complexity of an arbitrary source of messages of a particular length is much larger than the complexity of the messages it sends.

We are interested in the behavioral complexity when our objective is to use the messages that we receive to understand the source, rather than to make use of the information itself. Behavioral complexity becomes particularly useful when it is smaller than the complexity of a message, because it enables us to anticipate or predict the behavior of the source.

We now apply these thoughts about the source as the system of interest, rather than the message as the system of interest, to a discussion of the properties of physical systems. To make the connection between source and system, we consider an observer of a physical system who performs a number of measurements. We might imagine the measurements to consist of subjecting the system to light at various frequencies and measuring their scattering and reflection (looking at the system), observations of animals in the wild or in captivity, or physical probes of the system. We consider each measurement to be a message from the system to the observer. We must,

however, take note that any measurement consists of two parts,the conditions or environment in which the observation was performed and the behavior of the system under these conditions. We write any observation as a pair (e,a), where e represents the environment and a represents a measurement of system properties (action) under the circumstances of the environment e. The observer, after performing a number of measurements, writes a description of the observations. This description characterizes the system. It captures the properties of the list of measurements, rather than of one particular measurement. It may or may not explicitly contain the information of each measurement. Alternatively, it may assign probabilities to a particular measurement. We would like to define the behavioral complexity as the amount of information contained in the observer's description. However, we must be careful how we do this because of the presence of the environmental description e.

In order to clarify this point, and to make contact between behavioral complexity and our previous discussion of descriptive complexity, we first consider the physical system of interest to be essentially isolated. Then the environmental description is irrelevant, and an observation consists only of the system measurement a. The list of measurements is the set $\{a\}$. In this case it is relatively easy to see that the behavioral complexity of a physical system is its descriptive complexity—the set of all measurements characterizes completely the state of the system.

If the entire set of measurements is performed at a single instant, and has arbitrary precision, then the behavioral complexity is the microstate complexity of the system. The result of any measurement can be obtained from a description of the microstate, and the set of possible measurements determines the microstate.

For a set of measurements performed over time on an equilibrium system, the behavioral complexity is the ensemble complexity—the number of parameters necessary to specify its ensemble. A particular message is a measurement of the system properties, which in principle might be detailed enough to determine the instantaneous positions and momenta of all of the particles. However, the list of measurements is determined by the ensemble of states the system might have. As in Section 8.3.1, we conclude that the complexity of an equilibrium system is the complexity of describing its ensemble—specifying (U,N,V) and other parameters like magnetization that result from the breaking of ergodicity. For a glass, the ensemble information is the information in the frozen coordinates previously defined as the complexity. More generally, for a set of measurements performed over an interval of time T—or at one instant but with time determination error T—and with spatial position determination errors given by L, we recover the complexity profile.

We now return to consider a system that is not isolated but subject to an environmental influence so that an observation consists of the pair (e,a) (Fig. 8.3.5). The complexity of describing such messages also contains the complexity of the environment e. Does this mean that our system description must indude its environment and that the complexity of the system is dependent on the complexity of the environment? Complex systems or simple systems interact and respond to the environment in which they are found. Since the system response a is dependent on the environment e, there is no doubt that the complexity of a is dependent on the complexity of e. Three

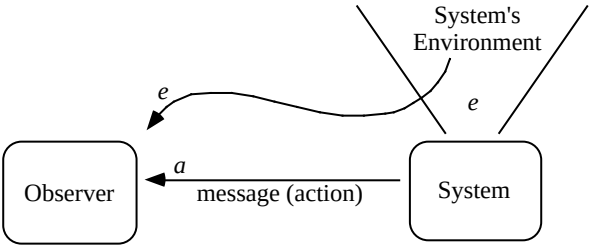

Figure 8.3.5 The observation of system behavior involves measurements both of the system's environment, e, and the system's actions, a, in response to this environment Thus we should characterize a system as a function, $a = f(e)$, where the function f describes its actions in response to its environment. It is generally simpler to describe a model for the system structure, which is also a model of f, rather than a list of all of its environment-action (e,a) pairs. ∎

examples illustrate how the environmental influence is important. The tail of a dog has a particular motion that can be described, and the complexity can be characterized. However, we may want to attribute much of this complexity to the rest of the dog rather than to the tail. Similarly, the motion of a particle suspended in a liquid follows Brownian motion, the description of which might be better attributed to the liquid than to the particle. Clearer yet is the example of the behavior of a basketball during a basketball game. These examples generalize to the consideration of any system, because measuring the properties of a system in an environment may cause us to be measuring the influence of the environment, rather than the system. The observer must describe the system behavior as a response to a particular environment, rather than just the behavior itself. Thus, we do not characterize the system by a list of actions $\{a\}$ but rather by the list of pairs $\{(e,a)\}$ where our concern is to describe f the functional mapping $a = f(e)$ from the environment e to the response a. Once we realize this, we can again affirm that a full microscopic description of the physical system is enough to give all system responses. The point is that the complexity of a system should not include the complexity of the influence upon it, but just the complexity of its response. This response is a property of the system and is determined by a complete microscopic description. Conversely, a full description of behavior subject to all possible environments would require complete microscopic information.

However, within a range of environments and with a desired degree of precision (spatial and temporal scale) it is possible to provide less information and still describe the behavior. We consider the ensemble of messages (measurements) to have possible times of observation over a range of times given by T and errors in position determination L. Describing the ensemble of responses gives us the behavioral complexity profile $C_b(L,T)$.

When the influence of the environment is not important, $C(L,T)$ and $C_b(L,T)$ are the same. When the environment matters,it is also important to characterize the information that is relevant about the environment. This is related to the problem of

prediction, because predicting the system behavior in the future requires information about the environment. As we have defined it,the descriptive complexity is the information necessary to predict the behavior of the system over the time interval $t_2 - t_1$. We can characterize the environmental influence by generalizing Eq. (8.3.47) to include a term that describes the rate of information transfer from the environment to the system:

$$\begin{aligned} C(L,T) &= C_b(L,T) + N_T C_e(L,T) \\ C_b(L,T) &= C_b(L) + C_t(L,T) + N_T k \sum_{i:h_i>0} h_i \end{aligned} \tag{8.3.59}$$

where $C_e(L)/k\ln(2)$ is the information about the environment necessary to predict the state of the system at the next time step, and $C_b(L)$ is the behavioral complexity at one time interval. Because the system itself is finite,the amount of information about the universe that is relevant to the system behavior in any interval of time must also be finite. We note that because the system affects the environment, which then affects the system, Eq.(8.3.59) as written may count information more than once. Thus,this expression as written is an upper bound on the complexity. We noted this point also with respect to the Lyaponov exponents after Eq. (8.3.47).

This use of behavior/response rather than a description to characterize a system is related to the use of response functions in physics, or input/output relationships to describe artificial systems. The response function can (in principle) be completely derived from the microscopic description of a system. It is more directly relevant to the system behavior in response to environmental influences, and thus is essential for direct comparison with experimental results.

Behavioral complexity suggests that we should consider the system behavior as represented by a function $a = f(e)$. The input to the function is a description of the environment; the output is the response or action. There is a difficulty with this approach in that the complexity of functions is generically much larger than that of the system itself. From the discussion in Section 8.2.3 we know that the description of a function would require an amount of information given by $C_f = C_a 2^{C_e}$, where C_e is the environmental complexity, and C_a is the complexity of the action. Because the environmental influence leads to an exponentially large complexity, it is clear that often the most compact description of the system behavior will give its structure rather than its response to all inputs. Then, in principle, the response can be derived from the structure. This also implies that the behavior of physical systems under different environments cannot be independent. We note that these conclusions must also apply to human beings as complex systems that respond to their environment (see Question 8.3.8).

Question 8.3.8 Discuss the following statements with respect to human beings as complex systems: "The most compact description of the system behavior will give its structure rather than its response to all inputs," and "This implies that the behavior of physical systems under different environments cannot be independent."

Solution 8.3.8 The first statement is relevant to the discussion of behaviorism as an approach to psychology (see Section 3.2.8). It says that the idea of describing human behavior by cataloging reactions to environmental stimuli is ultimately an inefficient approach. It is more effective to use such measurements to construct a model for the internal functioning of the individual and use this model to describe the measured responses. The model description is much more concise than the description of all possible responses.

Moreover, from the second statement we know that the model can describe the responses to circumstances that have not been measured. This also means that the use of such models may be effective in predicting the behavior of an individual. Specifically, that reactions of a human being are not independent of past reactions to other circumstances. A model that incorporates the previous behaviors may have some ability to predict the behavior to new circumstances. This is part of what we do when we interact with other individuals—we construct models that represent their behavior and then anticipate how they will react to new circumstances.

The coupling between the reaction of a human being under one circumstance to the reaction under a different circumstance is also relevant to our understanding of human limitations. Optimizing the response through adaptation to a set of environments according to some goal is a process that is limited in its effectiveness due to the coupling between responses to different circumstances. An individual who is effective in some circumstances may have qualities that lead to ineffective behavior under other circumstances. We will discuss this in Chapter 9 in the context of considering the specialization of human beings in society. This point is also applicable more generally to living organisms and their ability to consume resources and avoid predators as discussed in Chapter 6. Increasing complexity enables an organism to be more effective, but the effectiveness under a variety of circumstances is limited by the interdependence of responses. This is relevant to the observation that living organisms generally consume limited types of resources and live in particular ecological niches. ∎

8.3.7 *The observer and recognition*

The explicit existence of an observer in the definition of behavioral complexity enables us to further consider the role of the observer in the definition of complexity. What assumptions have we made about the properties of the observer? One of the assumptions that we have made is that the observer is more complex than the system. What happens if the complexity of the system is greater than the complexity of the observer? The complexity of an observer is the number of bits that may be used to describe the observer. If the observer is described by fewer bits than are needed to describe the system, then the observer will be unable to contain the description of the system that is being observed. In this case, the observer will construct a description of the system that is simpler than the system actually is. There are several possible ways that the observer may simplify the description of the system. One is to reject the

observation of all but a few kinds of messages. The other is to artificially limit the length of messages described. A third is to treat complex variability of the source as random—described by simple probabilities. These simplifications are often done in our modeling of physical systems.

An inherent problem in discussing behavioral complexity using environmental influence is that it is never possible to guarantee that the behavior of a system has been fully characterized. For example, a rock can be described as "just sitting there," if we want to describe the complexity of its motion under different environments. Of course the nature of the environment could be changed so that other behaviors will be realized. We may, for example,discover that the rock is actually a camouflaged animal. This is an inherent problem in behavioral complexity: it is never possible to characterize with certainty the complexity of a system under circumstances that have not been measured. All such conclusions are extrapolations. Performing such extrapolations is an essential part of the use of the description of a system. This is a general problem that applies to quantitative scientific modeling as well as the use of experience in general.

Finally, we describe the relevance of recognition to complexity. The first comment is related to the recognition of sets of numbers introduced briefly in Section 8.2.3. We introduced there the concept of recognition complexity of a set that relies upon a recognizer (a special kind of TM called a predicate that gives a single bit output) that can identify the system under discussion. Specifically, when presented with the system it says, "This is it," and when presented with any other system it says, "This is not it." We define the complexity of a system (or set of systems) as the complexity of the simplest recognizer of the system (or set of systems). There are some interesting features of this definition.First we realize that this definition is well suited to describing classes of systems. A description or model of a class of systems must identify common attributes rather than specific behaviors.A second interesting feature is that the complexity of the recognizer depends on the possible universe of systems that it can be presented with. For example,the complexity of recognizing cows depends on whether we allow ourselves to present the recognizer with all domestic animals, all known biological organisms on earth, all potentially viable biological organisms, or all possible systems. Naturally, this is an important issue in the field of pattern recognition, where the complexity of designing a system to recognize a particular pattern is strongly dependent on the universe of possibilities within which the pattern must be recognized. We will return to this point later when we consider the properties of human language in Section 8.4.1.

A different form of complexity related to recognition may be abstracted from the Turing test of artificial intelligence. This test suggests that we will achieve an artificial representation of intelligence when it becomes impossible to determine whether we are interacting with an artificial or actual human being. We can assume that Turing had in mind only a limited type of interaction between the observer "we" and the systems being observed—either the real or artificial representation of a human being. This test, which relies upon an observer to recognize the system, can serve as the basis for an additional definition of complexity. We determine the minimal possible

complexity of a model (simulated representation) of the system which would be recognized by a particular observer under particular circumstances as the system. The complexity of this model we call the substitution complexity. The sensitivity of this definition to the nature of the observer and the conditions of the observation is manifest. In some ways this definition,however, is implicit in all of our earlier definitions. In all cases, the complexity measures the length of a representation of the system. Ultimately we must determine whether a particular representation of the system is faithful. The "we" in the previous sentence is some observer that must recognize the system behavior in the constructed representation.

We conclude this section by reviewing some of the main concepts that were introduced. We noted the sensitivity of complexity to the spatial and temporal scale relevant to the description or response. The complexity profile formally takes this into account. If necessary, we can define the unique complexity of a system to be its complexity profile evaluated at its own scale.A more complete characterization of the system uses the entire complexity profile. We found that the mathematical models most closely associated with complexity—chaos and fractals—were both relevant. The former described the influence of microscopic information over time. The latter described the gradual rather than rapid loss of information with spatial and temporal scale. We also reconciled the notion of information as a measure of system complexity with the notion of complex systems as composed out of interdependent parts.Our next objective is to concretize this discussion further by estimating the complexity of particular systems.

8.4 Complexity Estimation

There are various difficulties associated with obtaining specific values for the complexity of a particular system. There are both fundamental and practical problems. Fundamental problems such as the difficulty in determining whether a representation is maximally compressed are important. However, before this is an issue we must first obtain a representation.

One approach to obtaining the complexity of a system is to construct a representation. The explicit representation should then be used to make a simulation to show that the system behavior is reproduced. If it is,then we know that the length of the representation is an upper bound on the complexity of the system. We can hope, however, that it will not be necessary to obtain explicit representations in order to estimate complexities. The objective of this section is to discuss various methods for estimating the complexity of systems with which we are familiar. These approaches make use of representations that we cannot simulate, however, they do have recognizable relationships to the system.

Measuring complexity is an experimental problem. The only reason that we are able to discuss the complexity of various systems is that we have already made many measurements of the properties of various systems. We can make use of the existing information to construct estimates of their complexity. A specific estimation method is not necessarily useful for all systems.

Our objective in this section is limited to obtaining "ballpark" estimates of the complexity of systems. This means that our errors will be in the exponent rather than in the number itself. We would be very happy to have an estimate of complexity such as $10^{3\pm1}$ or $10^{7\pm2}$. When appropriate, we keep track of half-decades using factors of three, such as in 3×10^4. These rough estimates will give us a first impression of the degree of complexity of many of the systems we would like to understand. It would tell us how difficult (very roughly) they are to describe. We will discuss three methods—(1) use of intuition and human language descriptions, (2) use of a natural representation tied to the system existence, where the principle example is the genome of living organisms, and (3) use of component counting. Each of these methods has flaws that will limit our confidence in the resulting estimates. However, since we are trying to find rough estimates, we can still take advantage of them. Consistency of different methods will give us some confidence in our estimates of complexity.

While we will discuss the complexity of various systems,our focus will be on determining the complexity of a human being. Our final estimate,$10^{10\pm2}$ bits will be obtained by combining the results of different estimation techniques in the following sections. The implications of obtaining an estimate of human complexity will be discussed in Section 8.4.4. We start,however, by noting that the complexity of a human being can be bounded by the physical entropy of the collection of atoms from which he or she is formed. This is roughly the entropy of a similar weight of water, about 10^{31} bits. This is the value of $S/k \ln 2$. As usual, we have assumed that there is nothing associated with a human being except the material of which he or she is formed, and that this material is described by known physical law. This entropy is an upper bound to the information necessary to specify the complete human being. The meaning of this number is that if we take away the person and we replace all of the atoms according to a specification of 10^{31} bits of information, we have replaced microscopically each atom where it was. According to our understanding of physical law, there can be no discernible difference. We will discuss the implications for artificial intelligence in Section 8.4.4, where we consider whether a computer could simulate the dynamics of atoms in order to simulate the behavior of the human being.

The entropy of a human being is much larger than the complexity estimate we are after, because we are interested in the complexity at a relevant spatial and temporal scale. In general we consider the complexity of a system at the natural scale defined in Section 8.3.5, one-tenth the size of the system itself,and the relaxation time of the behavior on this same length scale. We could also define the complexity by the observer. For example,the maximum visual sensitivity of a human being is about 1/100 of a second and 0.1 mm. For either case, observing only at this spatial and temporal scale decreases dramatically the relevance of the microscopic description. The reduction in information is hard to estimate directly. To estimate the relevant complexity, we must consider other techniques. However, since most of the information in the entropy is needed to describe the position of molecules of water undergoing vibrations, we can guess that the complexity is significantly smaller than the entropy.

8.4.1 *Human intuition—language and complexity*

The first method for estimation of complexity—the use of human intuition and language—is the least controlled/scientific method of obtaining an estimate of the complexity of a system. This approach,in its most basic form,is precisely what was asked in Question 8.2.1. We ask someone what they believe the complexity of the system is. It is assumed that the person we ask is somewhat knowledgeable about the system and also about the problem of describing systems. Even though it appears highly arbitrary, we should not dismiss this approach too readily because human beings are designed to understand complex systems. It could be argued that much of our development is directed toward enabling us to construct predictive models of various parts of the environment in which we live. The complexity of a system is directly related to the amount of study we need in order to master or predict the behavior of a system. It is not accidental that this is the fundamental objective of science—behavior prediction. We are quite used to using the word "complexity"in a qualitative manner and even in a comparative fashion—this is more complex or less complex than something else. What is missing is the quantitative definition. In order for someone to give a quantitative estimate of the complexity of a system,it is necessary to provide a definition of complexity that can be readily understood.

One useful and intuitive definition of complexity is the amount of information necessary to describe the behavior of a system. The information can be quantified in terms of representations people are familiar with—the amount of text/the number of pages /the number of books. This can be sufficient to cause a person to build a rough mental model of the system description, which is much more sophisticated than many explicit representations that might be constructed. There is an inherent limitation in this approach mentioned more generally above—a human being cannot directly estimate the complexity of an organism of similar or greater complexity than a human being. In particular, we cannot use this approach directly to estimate the complexity of human beings. Thus we will focus on simpler animals first. For example, we could ask the question in the following way: How much text is necessary to describe the behavior of a frog? We might emphasize for clarification that we are not interested in comparative frogology, or molecular frogology. We are just interested in a description of the behavior of a frog.

To gain additional confidence in this approach, we may go to the library and find descriptions that are provided in books. Superficially, we find that there are entire books devoted to a particular type of insect (mosquito, ant, butterfly), as there are books devoted to the tiger or the ape. However, there is a qualitative difference between these books. The books on insects are devoted to comparative descriptions, where various types of, e.g., mosquitoes, from around the world, their physiology, and/or their evolutionary history are described. Tens to hundreds of types are compared in a single book. Exceptional behaviors or examples are highlighted. The amount of text devoted to the behavior of a particular type of mosquito could be readily contained in less than a single chapter. On the other hand,a book devoted to tigers may describe only behavior (e.g., not physiology), and one devoted to apes

would describe only a particular individual in a manner that is limited to only part of its behaviors.

Does the difference in texts describing insects and tigers reflect the social priorities of human beings? This appears to be difficult to support. The mosquito is much more relevant to the well-being of human beings than the tiger. Mosquitoes are easier to study in captivity and are more readily available in the wild. There are films that enable us to observe the mosquito behavior at its own scale rather than at our usual larger scale. Despite such films, there is no book-length description of the behavior of a mosquito. This is true despite the importance of knowledge of its behavior to prevention of various diseases. Even if there is some degree of subjectivity to the complexity estimates obtained from the lengths of descriptions found in books, the use of existing books is a reasonable first attempt to obtain complexity estimates from the information that has been compiled by human beings. We can also argue that when there is greater experience with complexity and complexity estimation, our ability to use intuition or existing texts will improve and become important tools in complexity estimation.

Before applying this methodology, however, we should understand more carefully the basic relationship of language to complexity. We have already discussed in Section 1.8 the information in a string of English characters. A first estimate of 4.8 bits per character could be based upon the existence of 26 letters and 1 space. In Question 1.8.12, the best estimate obtained was 3.3 bits per character using a Markov chain model that included correlations between adjacent characters. To obtain an even better estimate, we need to have a model that includes longer-range correlations between characters. The most reliable estimates have been obtained by asking people to guess the next character in an English text. It is assumed that people have a highly sophisticated model for the structure of English and that the individual has no specific knowledge of the text. The guesses were used to establish bounds on the information content. We can summarize these bounds as 0.9 ± 0.3 bits/character. For our present discussion, the difference between high and low bounds (a factor of 2) is not significant. For convenience we will use 1 bit/character for our conversion factor. For larger quantities of text, this corresponds to values given in Table 8.4.1.

Our estimate of information in text has assumed a strictly narrative English text. We should also be concerned about figures that accompany descriptive materials. Does the conventional wisdom of "a picture is worth a thousand words" make sense? We can consider this both from the point of view of direct compression of the picture, and the

Amount of text	Information in text	Text with figures
1 char	1 bit	-
1 page = 3000 char	3×10^3 bit	10^4
1 chapter = 30 pages	10^5 bit	3×10^5
1 book = 10 chapters	10^6 bit	3×10^6

Table 8.4.1 Information estimates for straight English text and illustrated text. ∎

possibility of replacing the figure by descriptive text. A thousand words corresponds to 5×10^3 characters or bits,about two pages of text. Descriptive figures such as graphs or diagrams often consist of a few lines that can be concisely described using a formula and would have a smaller complexity. Photographs are formed of highly correlated graphical information that can be compressed. In a black and white photograph 5×10^3 bits would correspond to a 70×70 grid of completely independent pixels. If we recall that we are not interested in small details, this seems reasonable as an upper bound. Moreover, the text that accompanies a figure generally describes its essential content. Thus when we ask the key question—whether two pages of text would be sufficient to describe a typical figure and replace its function in the text—this seems a somewhat generous but not entirely unreasonable value. A figure typically occupies half of a page that would be otherwise occupied by text. Thus, for a highly illustrated book, on average containing one figure and one-half page of text on each page, our estimate of the information content of the book would increase from 10^6 bits by a factor of 2.5 to roughly 3×10^6 bits. If there is one picture on every two pages, the information content of the book would be doubled rather than tripled. While it is not really essential for our level of precision, it seems reasonable to adopt the convention that estimates using descriptions of behavioral complexity include figures. We will do so by increasing the previous values by a factor of 3 (Table 8.4.1). This will not change any of the conclusions.

There is another aspect of the relationship of language to complexity. A language uses individual words (like "frog") to represent complex phenomena or systems (like the physical system we call a frog). The complexity of the word "frog" is not the same as the complexity of the frog. Why is this possible? According to our discussion of algorithmic complexity, the smallest possible representation of a complex system has a length in bits which is equal to the system complexity. Here we have an example of a system—frog—whose representation "frog" is manifestly smaller than its complexity.

The resolution of this puzzle is through the concept of recognition complexity discussed in Section 8.3.7.A word is a member of an ensemble of words,and the systems that are described by these words are an ensemble of systems. It is only necessary that the ensemble of words be matched to the ensemble of systems described by the words,not the whole ensemble of possible systems. Thus,the complexity of a word is not related to the complexity of the system, but rather to the complexity of specifying the system—the logarithm of the number of systems that are part of the shared experience of the individuals who are communicating. This is the central point of recognition complexity. For a human being with experience and memory of only a limited number of the set of all complex systems, to describe a system one must identify it only in comparison with the systems in memory, not with those possible in principle.

Another way to think about this is to consider a human being as analogous to a special UTM with a set of short representations that the UTM can expand to a specific limited subset of possible long descriptions. For example, having memorized a play by Shakespeare,it is only necessary to invoke the name to retrieve the whole play. This is,indeed,the essence of naming—a name is a short reference to a complex system. All words are names of more complex entities.

In this way, language provides a systematic mechanism for compression of information. This implies that we should not use the length of a word to estimate the complexity of a system that it refers to. Does this also invalidate the use of human language to obtain complexity estimates? On one hand, when we are asked to describe the behavior of a frog, we assume that we must describe it without reference to the name itself. "It behaves like a frog" is not a sufficient description. There is a presumption that a description of behavior is made to someone without specific knowledge. An estimate of the complexity of a frog would be much higher than the complexity of the word "frog." On the other hand, the words that would be used to describe a frog also refer to complex entities or actions. Consistency in different estimates of the amount of text necessary to describe a frog might arise from the use of a common language and experience. We could expand the description further by requiring that a person explain not only the behavior of the frog, but also the meaning of each of the words used to describe the behavior of the frog. At this point, however, it is more constructive to keep in mind the subtle relationship between language and complexity as part of our uncertainty, and take the given estimates at face value. Ultimately, the complexity of a system is defined by the condition that all possible (in principle) behaviors of the same complexity could be described using the same length of text. We accept the possibility that language-based estimates of complexity of biological organisms may be systematically too small because they are common and familiar. We may nevertheless have relative complexities estimated correctly.

Finally, we can argue that when we estimate the complexity of systems that approach the complexity of a human being, the estimation problems becomes less severe. This follows because of our discussion of universality of complexity given in Section 8.2.2. Specifically, that the more complex a system is, the less relevant specific knowledge is, and the more universal are estimates of complexity. Nevertheless, ultimately we will conclude that the inherent compression in use of language for describing familiar complex systems is the greatest contributor to uncertainty in complexity estimates.

There is another approach to the use of human intuition and language in estimating complexity. This is by reference to computer languages. For someone familiar with computer simulation, we can ask for the length of the computer program that can simulate the behavior of the system—more specifically, the length of the program that can simulate a frog. Computer languages are generally not very high in information content, because there are a few commands and variables that are used throughout the program. Thus we might estimate the complexity of a program not by characters, but by program lines at several bits per program line. Consistent with the definition of algorithmic complexity, the estimate of system complexity should also include the complexity of the compiler and of the computer operating system and hardware. Compilers and operating systems are much more complex than many programs by themselves. We can bypass this problem by considering instead the size of the execution module—after application of the compiler.

There are other problems with the use of natural or artificial language descriptions, including:

1. Overestimation due to a lack of knowledge of possible representations. This problem is related to the difficulty of determining the compressibility of information. The assumption of a particular length of text presumes a kind of representation. This choice of representation may not be the most compact. This may be due to the form of the representation—specifically English text. Alternatively, the assumption may be in the conceptual (semantic) framework. An example is the complexity of the motion of the planets in the Ptolemaic (earth-centered) representation compared to the Copernican (sun-centered) representation. Ptolemy would give a larger complexity estimate than Copernicus because the Ptolemaic system requires a much longer description—which is the reason the Copernican system is accepted as "true" today.
2. Underestimation due to lack of knowledge of the full behavior of the system. If an individual is familiar with the behavior of a system only under limited circumstances,the presumption that this limited knowledge is complete will lead to a complexity estimate that is too low. Alternatively, lack of knowledge may also result in too high estimates if the individual extrapolates the missing knowledge from more complex systems.
3. Difficulty with counting. Large numbers are generally difficult for people to imagine or estimate. This is the advantage of identifying numbers with length of text, which is generally a more familiar quantity.

With all of these limitations in mind, what are some of the estimates that we have obtained? Table 8.4.2 was constructed using various books. The lengths of linguistic descriptions of the behavior of biological organisms range from several pages to several books. Insects and fish are at pages,frogs at a chapter, most mammals at approximately a book, and monkeys and apes at several books. These numbers span the range of complexity estimates.

We have concluded that it is not possible to use this approach to obtain an estimate of human complexity. However, this is not quite true. We can apply this method by taking the highest complexity estimate of other systems and using this as a close lower bound to the complexity of the human being. By close lower bound we mean that the actual complexity should not be tremendously greater. According to our

Animal	Text length	Complexity (bits)
Fish	a few pages	$3x10^4$
Grasshopper, Mosquito	a few pages to a chapter	10^5
Ant (one, not colony)	a few pages to a chapter	10^5
Frog	a chapter or two	$3x10^5$
Rabbit	a short book	10^6
Tiger	a book	$3x10^6$
Ape	a few books	10^7

Table 8.4.2 Estimates of the approximate length of text descriptions of animal behavior ∎

experience,the complexity estimates of animals tend to extend up to roughly a single book. Primates may be estimated somewhat higher, with a range of one to tens of books. This suggests that human complexity is somewhat larger than this latter number—approximately 10^8 bits, or about 30 books. We will see how this compares to other estimates in the following sections.

There are several other approaches to estimating human complexity based upon language. The existence of book-length biographies implies a poor estimate of human complexity of 10^6 bits. We can also estimate the complexity of a human being by the typical amount of information that a person can learn.Specifically, it seems to make sense to base an estimate on the length of a college education, which uses approximately 30 textbooks. This is in direct agreement with the previous estimate of 10^8 bits. It might be argued that this estimate is too low because we have not included other parts of the education (elementary and high school and postgraduate education) or other kinds of education/information that are not academic. It might also be argued that this is too high because students do not actually know the entire content of 30 textbooks. One reason this number appears reasonable is that if the complexity of a human being were much greater than this,there would be individuals who would endure tens or hundreds of college educations in different subjects. The estimate of roughly 30 textbooks is also consistent with the general upper limit on the number of books an individual can write in a lifetime. The most prolific author in modern times is Isaac Asimov, with about 500 books. Thus from such text-based self-consistent evidence we might assume that the estimate of 10^8 bits is not wrong by more than one to two orders of magnitude. We now turn to estimation methods that are not based on text.

8.4.2 *Genetic code*

Biological organisms present us with a convenient and explicit representation for their formation by development—the genome. It is generally assumed that most of the information needed to describe the physiology of the organism is contained in genetic information. For simplicity we might think of DNA as a kind of program that is interpreted by decoding machinery during development and operation. In this regard the genome is much like a Turing machine tape (see Section 1.9), even though the mechanism for transcription is quite different from the conventional Turing machine. Some other perspectives are given in Section 7.1.Regardless of how we ultimately view the developmental process and cellular function, it appears natural to associate with the genome the information that is necessary to specify physiological design and function. It is not difficult to determine an upper bound to the amount of information that is contained in a DNA sequence. Taken at face value,this provides us with an estimate of the complexity of an organism. We must then inquire as to the approximations that are being made. We first discuss the approach in somewhat greater detail.

Considering the DNA as an alphabet of four characters provided by the four nucleotides or bases represented by A (adenine) T (tyrosine) C (cytosine) G (guanine), a first estimate of the information contained in a DNA sequence would be $N \log(4) = 2N$. N is the length of the DNA chain. Since DNA is formed of two com-

Organism	Genome length (base pairs)	Complexity (bits)
Bacteria (E. coli)	10^6–10^7	10^7
Fungi	10^7–10^8	10^8
Plants	10^8–10^{11}	$3x10^8$–$3x10^{11}$
Insects	10^8–$7x10^9$	10^9
Fish (bony)	$5x10^8$–$5x10^9$	$3x10^9$
Frog and Toad	10^9–10^{10}	10^{10}
Mammals	$2x10^9$–$3x10^9$	10^{10}
Man	$3x10^9$	10^{10}

Table 8.4.3 Estimates of complexity based upon genome length. Except for plants, where there is a particularly wide range of genome lengths, a single number is given for the information contained in the genome, because the accuracy does not justify more specific numbers. Genome lengths and ranges are representative. ∎

plementary nucleotide chains in a double helix, its length is measured in base pairs. While this estimate neglects many corrections, there are a number of assumptions that we are making about the organism that give a larger uncertainty than some of the corrections that we can apply. Therefore as a rough estimate,this is essentially as good an estimate as we can obtain from this methodology at present.Specific numbers are given in Table 8.4.3. We see that for a human being, the estimate is nearly 10^{10} bits, which is somewhat larger than that obtained from language-based estimates in the previous section. What is more remarkable is that there is no systematic trend of increasing genome length that parallels our expectations of increasing organism complexity based on estimates of the last section. Aside from the increasing trend from bacteria to fungi to animals/plants,there is no apparent trend that would suggest that genome length is correlated with our expectations about complexity.

We now proceed to discuss limitations in this approach. The list of approximations given below is not meant to be exhaustive, but it does suggest some of the difficulties in determining the information content even when there is a clear first numerical value to start from.

a. A significant percentage of DNA is "non-coding." This DNA is not transcribed for protein structures. It may be relevant to the structural properties of DNA. It may also contain other useful information not directly relevant to protein sequence. Nevertheless, it is likely that information in most of the base pairs that are non-coding is not essential for organism behavior. Specifically, they can be replaced by many other possible base pair sequences without effect. Since 30%–50% of human DNA is estimated to be coding, this correction would reduce the estimated complexity by a factor of two to three.

b. Direct forms of compression: as presently understood,DNA is primarily utilized through transcription to a sequence of amino acids. The coding for each amino acid is given by a triple of bases. Since there are many more triples ($4^3 = 64$) than amino acids (twenty) some of the sequences have no amino acid counterpart,and

there are more than one sequence that map onto the same amino acid. This redundancy means that there is less information in the DNA sequence. Taking this into account by assigning a triple of bases to one of twenty characters that represent amino acids would give a new estimate of $(N/3)\log(20) = 1.4N$. To improve the estimate further, we would include the relative probability of the different amino acids, and correlations between them.

c. General compression: more generally, we can ask how compressed the DNA encoding of information is. We can rely upon a basic optimization of function in biology. This might suggest that some degree of compression is performed in order to reduce the complexity of transmission of the information from generation to generation. However, this is not a proof, and one could also argue in favor of redundancy in order to avoid susceptibility to small changes. Moreover there are likely to be inherent limitations on the compressibility of the information due to the possible transcription mechanisms that serve instead of decompression algorithms. For example,ifa molecule that is to be represented has a long chain of the same amino acid, e.g., asp-asp-asp-asp-asp-asp-asp-asp-asp-asp-asp-asp-asp-asp-asp-asp-asp-asp-asp, it would be interesting if this could be represented using a chemical equivalent of (18)asp. This requires a transcription mechanism that repeats segments—a DNA loop. There are organisms that are known to have highly repetitive sequences (e.g., 10^7 repetitions) forming a significant fraction of their genome. Much of this may be non-coding DNA.

Other forms of compression may also be relevant. For example, we can ask if there are protein components/subchains that can be used in more than one protein. This is relevant to the general redundancy of protein design. There is evidence that the genome does uses this property for compression by overlapping the regions that code for several different proteins. A particular region of DNA may have several coding regions that can be combined in different ways to obtain a number of different proteins. Transcription may start from distinct initial points. Presumably, the information that describes the pattern of transcriptions is represented in the noncoding segments that are between the coding segments. Related to the issue of DNA code compression are questions about the complexity of protein primary structure in relation to its own function—specifically, how much information is necessary to describe the function of a protein. This may be much less than the information necessary to specify its primary structure (amino acid sequence). This discussion is approaching issues of the scale at which complexity is measured—at the atomic scale where the specific amino acid is relevant, or at the molecular scale at which the enzymatic function is relevant. We will mention this limitation again in point (*d*).

d. Scale of representation:the genome codes for macromolecular and cellular function of the biological organism. This is much less than the microscopic entropy, since it does not code the atomic vibrations or molecular diffusion. However, since our concern is for the organism's macroscopic complexity, the DNA is likely to be coding a far greater complexity than we are interested in for multicellular

organisms. The assumption is that much of the cellular chemical activity is not relevant to a description of the behavior on the scale of the organism. If the DNA were representing the sum of the molecular or cellular scale complexity of each of the cells independently, then the error in estimating the complexity would be quite large. However, the molecular and cellular behavior is generally repeated throughout the organism in different cells. Thus, the DNA is essentially representing the complexity of a single cellular function with the additional complication of representing the variation in this function. To the extent that the complexity of cellular behavior is smaller than that of the complete organism,it may be assumed that the greatest part of the DNA code represents the macroscale behavior. On the other hand, if the organism behavior is comparatively simple,the greater part of the DNA representation would be devoted to describing the cellular behavior.

e. Completeness of representation: we have assumed that DNA is the only source of cellular information. However, during cell division not only the DNA is transferred but also other cellular structures,and it is not clear how much information is necessary to specify their function. It is clear, however, that DNA does not contain all the information. Otherwise it would be possible to transfer DNA from one cell into any other cell and the organism would function through control by the DNA. This is not the case. However, it may very well be that the description of all other parts of the cell, including the transcription mechanisms, only involves a small fraction of the information content compared to the DNA (for example,10^4–10^6 bits compared to 10^7–10^{11} bits in DNA).Similar to our point (*d*), the information in cellular structures is more likely to be irrelevant for organisms whose complexity is high. We could note also that there are two sources of DNA in the eukaryotic cell, nuclear DNA and mitochondrial DNA. The information in the nuclear DNA dominates over the mitochondrial DNA, and we also expect it to dominate over other sources of cellular information. It is possible, however, that the other sources of information approach some fraction (e.g., 10%) of the information in the nuclear DNA, causing a small correction to our estimates.

f. We have implicitly assumed that the development process of a biological organism is deterministic and uniquely determined by the genome. Randomness in the process of development gives rise to additional information in the final structure that is not contained in the genome. Thus, even organisms that have the same DNA are not exactly the same. In humans, identical twins have been studied in order to determine the difference between environmental and genetic influence. Here we are not considering the macroscale environmental influence, but rather the microscale influence. This influence begins with the randomness of molecular vibrations during the developmental process. The additional information gained in this way would have to play a relatively minor functional role if there is significance to the genetic control over physiology. Nevertheless,a complete estimate of the complexity of a system must include this information. Without considering different scales of structure or behavior, on the macroscale we should

not expect the microscopic randomness to affect the complexity by more than a factor of 2,and more likely the effect is not more than 10% in a typical biological organism.

g. We have also neglected the macroscale environmental influences on behavior. These are usually described by adaptation and learning. For most biological organisms,the environmental influences on behavior are believed to be small compared to genetic influences. Instinctive behaviors dominate. This is not as true about many mammals and even less true about human beings. Therefore,the genetic estimate becomes less reliable as an upper bound for human beings than it is for lower animals. This point will be discussed in greater detail below.

We can see that the assumptions discussed in (*a*), (*b*), (*c*) and (*d*) would lead to the DNA length being an overly large estimate of the complexity. Assumptions discussed in (*e*), (*f*) and (*g*) imply it is an underestimate.

One of the conceptual difficulties that we are presented with in considering genome length as a complexity estimate is that plants have a much higher DNA length than animals. This is in conflict with the conventional wisdom that animals have a greater complexity of behavior than plants.We might adopt one of two approaches to understanding this result: first, that plants are actually more complex than animals, and second, that the DNA representation in plants does not make use of, or cannot make use of, compression algorithms that are present in animal cells.

If plants are systematically more complex than animals, there must be a general quality of plants that has higher descriptive and behavioral complexity. A candidate for such a property is that plants are generally able to regenerate after injury. This inherently requires more information than the reliance upon a specific time history for development. In essence,there must be some form of actual blueprint for the organism encoded in the genome that takes into account many possible circumstances. From a programming point of view, this is a multiply reentrant program. To enable this feature may very well be more complex, or it may require a more redundant (longer) representation of the same information. It is presumed that the structure of animals has such a high intrinsic complexity that representation of a fully regenerative organism would be impossible. This idea might be checked by considering the genome length of animals that have greater ability to regenerate. If they are substantially longer than similar animals without the ability to regenerate, the explanation would be supported. Indeed, the salamander, which is the only vertebrate with the ability to regenerate limbs, has a genome of 10^{11} base pairs. This is much larger than that of other vertebrates, and comparable to that of the largest plant genomes.

A more general reason for the high plant genome complexity that is consistent with regeneration would be that plants have systematically developed a high complexity on smaller (molecular and cellular) rather than larger (organismal) scales. One reason for this would be that plant immobility requires the development of complex molecular and cellular mechanisms to inhibit or survive partial consumption by other organisms. By our discussion of the complexity profile in Section 8.3, a high complexity on small scales would not allow a high complexity on larger scales. This

explanation would also be consistent with our understanding of the relative simplicity of plants on the larger scale.

The second possibility is that there exists a systematic additional redundancy of the genome in plants. This might be the result of particular proteins with chains of repetitive amino acids.A protein formed out of a long chain of the same amino acid might be functionally of importance in plants,and not in animals. This is a potential explanation for the relative lengths of plant genome and animal genome.

One of the most striking features of the genome lengths found for various organisms is their relative uniformity. Widely different types of organisms have similar genome lengths, while similar organisms may have quite different genome lengths. One explanation for this that might be suggested is that genome lengths have increased systematically with evolutionary time. It is hard, however, to see why this would be the case in all but the simplest models of evolution. It makes more sense to infer that there are constraints on the genome lengths that have led it to gravitate toward a value in the range 10^9–10^{10}. Increases in organism complexity then result from fewer redundancies and better compression, rather than longer genomes. In principle, this could account for the pattern of complexities we have obtained.

Regardless of the ultimate reason for various genome lengths, in each case the complexity estimate from genome length provides an upper bound to the genetic component of organism complexity (c.f. points (*e*), (*f*) and (*g*) above). Thus,the human genome length provides us with an estimate of human complexity.

8.4.3 *Component counting*

The objective of complexity estimation is to determine the behavioral complexity of a system as a whole. However, one of the important clues to the complexity of the system is its composition from elements and their interactions. By counting the number of elements, we can develop an understanding of the complexity of the system. However, as with other estimation methods,it must be understood that there are inherent problems in this approach. We will find that this method gives us a much higher estimate than the other methods. In using this method we are faced with the dilemma that lies at the heart of the ability to understand the nature of complex systems—how does complex behavior arise out of the component behavior and their interactions? The essential question that we face is: Assuming that we have a system formed of N interacting elements that have a complexity C_0 (or a known distribution of complexities),how can the complexity C of the whole system be determined? The maximal possible value would be NC_0. However, as we discussed in Section 8.3,this is reduced both by correlations between elements and by the change of scale from that of the elements to that of the system. We will discuss these problems in the context of estimating human complexity.

If we are to consider the behavioral complexity of a human being by counting components, we must identify the relevant components to count. If we count the number of atoms, we would be describing the microscopic complexity. On the other hand, we cannot count the number of parts on the scale of the organism (one) because the problem in determining the complexity remains in evaluating C_0. Thus

the objective is to select components at an intermediate scale. Of the natural intermediate scales to consider, there are molecules, cells and organs. We will tackle the problem by considering cells and discuss difficulties that arise in this context. The first difficulty is that the complexity of behavior does not arise equally from all cells. It is generally understood that muscle cells and bone cells are largely uniform in structure. They may therefore collectively be described in terms of a few parameters, and their contribution to organism behavior can be summarized simply. In contrast,as we discussed in Chapter 2,the behavior of the system on the scale of the organism is generally attributed to the nervous system. Thus,aside from an inconsequential number of additional parameters, we will consider only the cells of the nervous system. If we were considering the behavior on a smaller length scale, then it would be natural to also consider the immune system.

In order to make more progress, we must discuss a specific model for the nervous system and then determine its limitations. We can do this by considering the behavior of a model system we studied in detail in Chapter 2—the attractor neural network model.Each of the neurons is a binary variable. Its behavior is specified by whether it is ON or OFF. The behavior of the network is,however, described by the values of the synapses. The total complexity of the synapses could be quite high if we allowed the synapses to have many digits of precision in their values, but this does not contribute to the complexity of the network behavior. Given our investigation of the storage of patterns in the network, we can argue that the maximal number of independent parameters that may be specified for the operation of the network consists of the neural firing patterns that are stored. This corresponds to $\alpha_c N^2$ bits of information, where N is the number of neurons,and $\alpha_c \approx 0.14$ is a number that arose from our analysis of network overload.

There are several problems with applying this formula to biological nervous systems. The first is that the biological network is not fully connected. We could apply a similar formula to the network assuming only the number of synapses N_s that are present, on average, for a neuron. This gives a value $\alpha_c N_s N$. This means that the storage capacity of the network is smaller, and should scale with the number of synapses. For the human brain where N_s has been estimated at 10^4 and $N \approx 10^{11}$, this would give a value of $0.1 \times 10^4 \times 10^{11} = 10^{14}$ bits. The problem with this estimate is that in order to specify the behavior of the network, we need to specify not only the imprinted patterns but also which synapses are present and which are absent.Listing the synapses that are present would require a set of number pairs that would specify which neurons each neuron is attached to. This list would require roughly $NN_s\log(N) = 3 \times 10^{16}$ bits, which is larger than the number of bits of information in the storage itself. This estimate may be reduced by a small amount, if, as we expect, the synapses of a neuron largely connect to neurons that are nearby. We will use 10^{16} as the basis for our complexity estimate.

The second major problem with this model is that real neurons are far from binary variables. Indeed, a neuron is a complex system. Each neuron responds to particular neurotransmitters, and the synapse between two specific neurons is different from other synapses. How many parameters would be needed to describe the behavior of an individual neuron,and how relevant are these parameters to the complexity

of the whole system? Naively, we might think that taking into account the complexity of individual neurons gives a much higher complexity than that considered above. However, this is not the case. We assume that the parameters necessary to describe an individual neuron correspond to a complexity C_0, and it is necessary to specify the parameters of all of the neurons. Then the complexity of the whole system would include C_0N bits for the neurons themselves. This would be greater than 10^{16} bits only if the complexity of the individual neurons were larger than 10^5. A reasonable estimate of the complexity of a neuron is roughly 10^3–10^4 bits. This would give a value of $C_0N = 10^{13}$–10^{14} bits, which is not a significant amount by comparison with 10^{16} bits. By these estimates, the complexity of the internal structure of a neuron is not greater than the complexity of its interconnections.

Similarly, we should consider the complexity of a synapse, which multiplies the number of synapses. Synapses are significantly simpler than the neurons. We may estimate their complexity as no more than 10 bits. This would be sufficient to specify the synaptic strength and the type of chemicals involved in transmission. Multiplying this by the total number of synapses (10^{15}) gives 10^{16} bits. This is the same as the information necessary to specify the list of synapses that are present.

Combining our estimates for the information necessary to specify the structure of neurons, the structure of synapses and the list of synapses present, we obtain an estimate for complexity of 10^{16} bits. This estimate is significantly larger than the estimate found from the other two approaches. As we mentioned before, there are two fundamental difficulties with this approach that make the estimate too high—correlations among parameters and the scale of description.

Many of the parameters enumerated above are likely to be the same, giving rise to the possibility of compression of the description. Both the description of an individual neuron and the description of the synapses between them can be drastically simplified if all of them follow a pattern. For example, the visual system involves processing of a visual field where the different neurons at different locations perform essentially the same operation on the visual information. Even if there are smooth variations in the parameters that describe both the neuron behavior and the synapses between them, we can describe the processing of the visual field in terms of a small number of parameters. Indeed, one would guess (an intuition-based estimate) that processing of the visual field is quite complicated (more than 10^2 bits) but would not exceed 10^3–10^5 bits altogether. Since a substantial fraction of the number of neurons in the brain is devoted to initial visual processing, the use of this reduced description of the visual processing would reduce the estimate of the complexity of the whole system.

Nevertheless, the initial visual processing does not involve more than 10% of the number of neurons. Even if we eliminate all of their parameters, the estimate of system complexity would not change. However, the idea behind this construction is that whenever there are many neurons whose behavior can be grouped together into particular functions, then the complexity of the description is reduced. Thus if we can describe neurons as belonging to a particular class of neurons (category or stereotype), then the complexity is reduced. It is known that neurons can be categorized; however, it is not clear how many parameters remain once this categorization has been done.

When we think about grouping the neurons together, we might also realize that this discussion is relevant to the consideration of the influence of environment and genetics on behavior. If the number of parameters necessary to describe the network greatly exceeds the number of parameters in the genetic code, which is only 10^{10} bits, then many of these parameters must be specified by the environment. We will discuss this again in the next section.

On a more philosophical note, we comment that parameters that describe the nervous system also include the malleable short-term memory. While this may be a small part of the total information, our estimate of behavioral complexity should raise questions such as, How specific do we have to be? Should the content of short-term memory be included? The argument in favor would be that we need to represent the human being in entirety. The argument against would be that what happened in the past five minutes or even the past day is not relevant and we can reset this part of the memory. Eventually we may ask whether the objective is to represent the specific information known by an individual or just his or her "character."

We have not yet directly addressed the role of substructure (Chapter 2) in the complexity of the nervous system. In comparison with a fully connected network, a network with substructure is more complex because it is necessary to specify the substructure, or more specifically which neurons (or which information) are proximate to which. However, in a system that is subdivided by virtue of having fewer synapses between subdivisions, once we have counted the information that is present in the selection of synapses,as we have done above,the substructure of the system has already been included.

The second problem of estimating complexity based on component counting is that we do not know how to reduce the complexity estimate based upon an increase of the length scale of observation. The estimate we have obtained for the complexity of the nervous system is relevant to a description of its behavior on the scale of a neuron (it does, however, focus on cellular behavior most relevant to the behavior of the organism). In order to overcome this problem, we need a method to assess the dependence of the organism behavior on the cellular behavior. A natural approach might be to evaluate the robustness of the system behavior to changes in the components. Human beings are believed to lose approximately 10^6 neurons every day (even without alcohol) corresponding to the loss of a significant fraction of the neurons over the course of a lifetime. This suggests that individual neurons are not crucial to determining human behavior. It implies that there may be a couple of orders of magnitude between the estimate of neuron complexity and human complexity. However, since the daily loss of neurons corresponds only to a loss of 1 in 10^5 neurons, we could also argue that it would be hard for us to notice the impact of this loss. In any event, our estimate based upon component counting, 10^{16}, is eight orders of magnitude larger than the estimates obtained from text and six orders of magnitude larger than the genome-based estimate. To account for this difference we would have to argue that 99.999% of neuron parameters are irrelevant to human behavior. This is too great a discrepancy to dismiss based upon such an argument.

Finally, we can demonstrate that 10^{16} is too large an estimate of complexity by considering the counting of time rather than the counting of components. We consider a minimal time interval of describing a human being to be of order 1 second, and we allow for each second 10^3 bits of information. There are of order 10^9 seconds in a lifetime. Thus we conclude that only, at most, 10^{12} bits of information are necessary to describe the actions of a human. This estimate assumes that each second is independently described from all other seconds, and no patterns of behavior exist. This would seem to be a very generous estimate. We can contrast this number with an estimate of the total amount of information that might be imprinted upon the synapses. This can be estimated as the total number of neuronal states over the course of a lifetime. For a neuron reaction time of order 10^{-2} seconds, 10^{11} neurons, and 10^9 seconds in a lifetime, we have 10^{22} bits of information. Thus we see that the total amount of information that passes through the nervous system is much larger than the information that is represented there, which is larger than the information that is manifest in terms of behavior. This suggests either that the collective behavior of neurons requires redundant information in the synapses, as discussed in Section 8.3.6, or that the actions of an individual do not fully represent the possible actions that the individual would take under all circumstances. The latter possibility returns us to the discussion of Eq.(8.3.47) and Eq.(8.3.59), where we commented that the expression is an upper bound, because information may cycle between scales or between system and environment. Under these circumstances, the potential complexity of a system under the most diverse set of circumstances is not necessarily the observed complexity. Both of our approaches to component counting (spatial and temporal) may overestimate the complexity due to this problem.

8.4.4 *Complexity of human beings, artificial intelligence, and the soul*

We begin this section by summarizing the estimates of human complexity from the previous sections, and then turn to some more philosophical considerations of its significance. We have found that the microscopic complexity of a human being is in the vicinity of 10^{30} bits. This is much larger than our estimates of the macroscopic complexity—language-based 10^8 bits, genome-based 10^{10} bits and component (neuron)-counting 10^{16} bits. As discussed at the end of the last section, we replace the spatial component-counting estimate with the time-counting upper bound of 10^{12} bits. We will discuss the discrepancies between these numbers and conclude with an estimate of $10^{10\pm2}$ bits.

We can summarize our understanding of the different estimates. The language-based estimate is likely to be somewhat low because of the inherent compression achieved by language. One way to say this is that a college education, consisting of 30 textbooks, is based upon childhood learning (nonlinguistic and linguistic) that provides meaning to the words, and therefore contains comparable or greater information. The genome-based complexity is likely to be a too-large estimate of the influence of genome on behavior, because genome information is compressible and because much of it must be relevant to molecular and cellular function. The component-

counting estimate suggests that the information obtained from experience is much larger than the information due to the genome—specifically, that genetic information cannot specify the parameters of the neural network. This is consistent with our discussion in Section 3.2.11 that suggested that synapses store learned information while the genome determines the overall structure of the network. We must still conclude that most of the network information is not relevant to behavior at the larger scale. It is redundant, and /or does not manifest itself in human behavior because of the limited types of external circumstances that are encountered. Because of this last point, the complexity for describing the response to arbitrary circumstances may be higher than the estimate that we will give, but should still be significantly less than 10^{16} bits.

Our estimate of the complexity of a human being is $10^{10\pm2}$ bits. The error bars essentially bracket the values we obtained. The main final caveat is that the difficulty in assessing the possibility of information compression may lead to a systematic bias to high complexities. For the following discussion,the actual value is less important than the existence of an estimate.

Consideration of the complexity of a human being is intimately related to fundamental issues in artificial intelligence. The complexity of a human being specifies the amount of information necessary to describe and, given an environment, predict the behavior of a human being. There is no presumption that the prediction would be feasible using present technology. However, in principle,there is an implication of its possibility. Our objective here is to briefly discuss both philosophical and practical implications of this observation.

The notion of reproducing human behavior in a computer (or by other artificial means) has traditionally been a major domain of confrontation between science and religion,and science and popular thought. Some of these conflicts arise because of the supposition by some religious philosophers of a nonmaterial soul that is presumed to animate human beings. Such nonmaterial entities are rejected in the context of science because they are, by definition,not measurable. It may be helpful to discuss some of the alternate approaches to the traditional conflict that bypass the controversy in favor of slightly modified definitions.Specifically, we will consider the possibility of a scientific definition of the concept of a soul. We will see that such a concept is not necessarily in conflict with notions of artificial intelligence. Instead it is closely related to the assumptions of this field.

One way to define the concept of soul is as the information that describes completely a human being. We have just estimated the amount of this information. To understand how this is related to the religious concept of soul, we must realize that the concept of soul serves a purpose. When an individual dies,the existence of a soul represents the independence of the human being from the material of which he or she is formed. If the material of which the human being is made were essential to its function, then there would be no independent functional description. Also, there would be no mechanism by which we could reproduce human behavior without making use of precisely the atoms of which he or she was formed. In this way the description of a soul suggests an abstraction of function from matter which is consistent with abstractions that are familiar in science and modern thought, but might not be consis-

tent with more primitive notions of matter. A primitive concept of matter might insist that the matter of which we are formed is essential to our functioning. The simplest possible abstraction would be to state (as is claimed by physics) that the specific atoms of which the human being are formed are not necessary to his or her function. Instead, these atoms may be replaced by other indistinguishable atoms and the same behavior will be found. Artificial intelligence takes this a large step further by stating that there are other possible media in which the same behavior can be realized. A human being is not directly tied to the material of which he is made. Instead there is a functional description that can be implemented in various media, of which one possible medium is the biological body that the human being was implemented in, when we met him or her.

Viewed in this light, the statement of the existence of a soul appears to be the same as the claim of artificial intelligence—that a human being can be reproduced in a different form by embodying the function rather than the mechanism of the human being. There is, however, a crucial distinction between the religious view and some of the practical approaches of artificial intelligence. This difference is related to the notion of a universal artificial intelligence, which is conceptually similar to the model of universal Turing machines. According to this view there is a generic model for intelligence that can be implemented in a computer. In contrast, the religious view is typically focused on the individual identity of an individual human being as manifest in a unique soul. We have discussed in Chapter 3 that our models of human beings are to be understood as nonuniversal and would indeed be better realized by the concept of representing individual human beings rather than a generic artificial intelligence. There are common features to the information processing of different individuals. However, we anticipate that the features characteristic of human behavior are predominantly specific to each individual rather than common. Thus the objective of creating artificial human beings might be better described as that of manifesting the soul of an individual human.

We can illustrate this change in perspective by considering the Turing test for recognizing artificial intelligence. The Turing test suggests that in a conversation with a computer we may not be able to distinguish it from a human being. A key problem with this prescription is that there is no specification of which human being is to be modeled. Human beings have varied complexity, and interactions are of varied levels of intimacy. It would be quite easy to reproduce the conversation of a mute individual, or even an obsessed individual. Which human being did Turing have in mind? We can go beyond this objection by recognizing that in order to fool us into thinking that the computer is a human being, except for a very casual conversation, the computer would have to represent a single human being with a name, a family history, a profession, opinions and a personality, not an abstract notion of intelligence. Finally, we may also ask whether the represented human being is someone we already know, or someone we do not know, prior to the test.

While we bypassed the fundamental controversy between science and religion regarding the presence of an immaterial soul, we suspect that the real conflict between the approaches resides in a different place. This conflict is in the question of the

intrinsic value of a human being and his place in the universe. Both the religious and popular view would like to place an importance on a human being that transcends the value of the matter of which he is formed. Philosophically, the scientific perspective has often been viewed as lowering human worth. This is true whether it is physical scientists that view the material of which man is formed as "just" composed of the same atoms as rocks and water, or whether it is biological scientists that consider the biochemical and cellular structures as the same as,and derived evolutionarily from,animal processes.

The study of complexity presents us with an opportunity in this regard.A quantitative definition of complexity can provide a direct measure of the difference between the behavior of a rock,an animal and a human being. We should recognize that this capability can be a double-edged sword. On the one hand it provides us with a scientific method for distinguishing man from matter, and man from animal, by recognizing that the particular arrangement of atoms in a human being, or the particular implementation of biology, achieves a functionality that is highly complex. At the same time, by placing a number on this complexity it presents us with the finiteness of the human being. For those who would like to view themselves as infinite,a finite complexity may be humbling and difficult to accept. Others who already recognize the inherent limitations of individual human beings,including themselves, may find it comforting to know that this limitation is fundamental.

As is often the case,the value of a number attains meaning though comparison. Specifically, we may consider the complexity of a human being and see it as either high or low. We must have some reference point with respect to which we measure human complexity. One reference point was clear in the preceding discussion—that of animals. We found that our (linguistic) estimates of human complexity placed human beings quantitatively above those of animals, as we might expect. This result is quite reasonable but does not suggest any clear dividing line between animals and man. There is, however, an independent value to which these complexities can be compared. For consistency, we use language-based complexity estimates throughout.

The idea of biological evolution and the biological continuity of man from animal is based upon the concept of the survival demands of the environment on man. Let us consider for the moment the complexity of the demands of the environment. We can estimate this complexity using relevant literature. Books that discuss survival in the wild are typically quite short, 3×10^5 bits. Such a book might describe more than just basic survival—plants to eat and animal hunting—but also various skills of a primitive life such as stone knives, tanning, basket making, and primitive home or boat construction. Alternatively, a book might discuss survival under extreme circumstances rather than survival under more typical circumstances. Even so, the amount of text is not longer than a rather brief book. While there are many individuals who have devoted themselves to living in the wild,there are no encyclopedias of relevant information. This suggests that in comparison with the complexity of a human being, the complexity of survival demands is small. Indeed, this complexity appears to be right at the estimated dividing line between animal (10^6 bits) and man (10^8 bits). It is significant that an ape may have a complexity of ten times the com-

plexity of the environmental demands upon it, but a human being has a complexity of a hundred times this demand. Another way to arrive at this conclusion is to consider primitive man, or primitive tribes that exist today. We might ask about the complexity of their existence and specifically whether the demands of the survival are the same as the complexity of their lives. From books that reflect studies of such peoples we see that the description of their survival techniques is much shorter than the description of their social and cultural activities. A single aspect of their culture might occupy a book, while the survival methods do not occupy even a single one.

We might compare the behavior of primitive man with the behavior of animal predators. In contrast to grazing animals, predators satisfy their survival needs in terms of food using only a small part of the day. One might ask why they did not develop complex cultural activities. One might think, for example, of sleeping lions. While they do have a social life, it does not compare in complexity to that of human beings. The explanation that our discussion provides is that while time would allow cultural activities, complexity does not. Thus, the complexity of such predators is essentially devoted to problems of survival. That of human beings is not.

This conclusion is quite intriguing. Several interesting remarks follow. In this context we can suggest that analyses of animal behavior should not necessarily be assumed to apply to human behavior. In particular, any animal behavior might be justified on the basis of a survival demand. While this approach has also often been applied to human beings—the survival advantages associated with culture, art and science have often been suggested—our analysis suggests that this is not justified, at least not in a direct fashion. Human behavior cannot be driven by survival demands if the survival demands are simpler than the human behavior. Of course, this does not rule out that general aspects or patterns of behavior, or even some specific behaviors, are driven by survival demands.

One of the distinctions between man and animals is the relative dominance of instinctive behavior in animals, as compared to learned behavior in man. It is often suggested that human dependence on learned rather than instinctive behavior is simply a different strategy for survival. However, if the complexity of the demands of survival are smaller than that of a human being, this does not hold. We can argue instead that if the complexity of survival demands are limited, then there is no reason for additional instinctive behaviors. Thus, our results suggest that instinctive behavior is actually a better strategy for overcoming survival demands—because it is prevalent in organisms whose behavior arises in response to survival demands. However, once such demands are met, there is little reason to produce more complex instinctive behaviors, and for this reason human behavior is not instinctively driven.

We now turn to some more practical aspects of the implications of our complexity estimates for the problem of artificial intelligence—or the re-creation of an individual in an artificial form. We may start from the microscopic complexity (roughly the entropy) which corresponds to the information necessary to replace every atom in the human being with another atom of the same kind, or alternatively, to represent the atoms in a computer. We might imagine that the computer could simulate the dynamics of the atoms in order to simulate the behavior of the human being. The

practicality of such an implementation is highly questionable. The problem is not just that the number of bits of storage as well as the speed requirements are beyond modern technology. It must be assumed that any computer representation of this dynamics must ultimately be composed of atoms. If the simulation is not composed out of the atoms themselves, but some controllable representation of the atoms, then the complexity of the machine must be significantly greater than that of a human being. Moreover, unless the system is constructed to respond to its environment in a manner similar to the response of a human being, then the computer must also simulate the environment. Such a task is likely to be formally as well as practically impossible.

One central question then becomes whether it is possible to compress the representation of a human being into a simpler one that can be stored. Our estimate of behavioral complexity, $10^{10\pm2}$ bits, suggests that this might be possible. Since a CD-ROM contains 5×10^9 bits, we are discussing $2 \times 10^{\pm2}$ CD-ROMs. At the lower end of this range, 0.02 CD-ROMs is clearly not a problem. Even at the upper end, two hundred CD-ROMs is well within the domain of feasibility. Indeed, even if we chose to represent the information we estimated to be necessary to describe the neural network of a single individual, 10^{16} bits or 2 million CD-ROMs, this would be a technologically feasible project. We have made no claims about our ability to obtain the necessary information for one individual. However, once this information is obtained, it should be possible to store it. A computer that can simulate the behavior of this individual represents a more significant problem.

Before we discuss the problem of simulating a human being, we might ask what the additional microscopic complexity present in a human body is good for. Specifically, if only 10^{10} bits are relevant to human behavior, what are most of the 10^{31} bits doing? One way to think about this question is to ask why nature didn't build a similar machine with of order 10^{10} atoms, which would be significantly smaller. We might also ask whether we would know if such an organism existed. On our own scale, we might ask why nature doesn't build an organism with a complexity of order 10^{30}. We have already suggested that there may be inherent limitations to the complexity that can be formed. However, there may also be another use of some of the additional large number of microscopic pieces of information.

One possible use of the additional information can be inferred from our arguments about the difference between TM with and without a random tape. The discussion in Section 1.9.7 suggests that it may be necessary to have a source of randomness to allow human qualities such as creativity. This fits nicely with our discussion of chaos in complex system behavior. The implication is that the microscopic information becomes gradually relevant to the macroscopic behavior as a chaotic process. We can assume that most microscopic information in a human being describes the position and orientation of water molecules. In this picture, random motion of molecules affects cellular behavior, specifically the firing of neurons, that ultimately affects human behavior. This does not mean that all of the microscopic information is relevant. Only a small number of bits can be relevant at any time. However, we recognize that in order to obtain a certain number of random bits, there must be a much larger reservoir of randomness. This is one approach to understanding a possible use of the microscopic information content of a human being. Another

approach would ascribe the additional information to the necessary support structures for the complex behavior, but would not attribute to it an essential role as information.

We have demonstrated time and again that it is possible to build a stronger or faster machine than a human being. This has led some people to believe that we can also build a systematically more capable machine—in the form of a robot. We have already argued that the present notion of computers may not be sufficient if it becomes necessary to include chaotic behavior. We can go beyond this argument by considering the problem we have introduced of the fundamental limits to complexity for a collection of molecules. It may turn out that our quest for the design of a complex machine will be limited by the same fundamental laws that limit the design of human beings.One of the natural improvements for the design of deterministic machines is to consider lower temperatures that enable lower error rates and higher speeds, and possibly the use of superconductors. However, the choice of a higher temperature may be required to enable a higher microscopic complexity, which also limits the macroscopic complexity. The mammalian body temperature may be selected to balance two competing effects. At high temperatures there is a high microscopic complexity. However, breaking the ergodic theorem requires low temperatures so that energy barriers can be effective in stopping movement in phase space.A way to argue this point more generally is that the sensitivity of human ears and eyes is not limited by the biological design, but by fundamental limits of quantum mechanics. It may also be that the behavioral complexity of a human being at its own length and time scale is limited by fundamental law. As with the existence of artificial sensors in other parts of the visual spectrum, we already know that machines with other capabilities can be built. However, this argument suggests that it may not be possible to build a systematically more complex artificial organism.

The previous discussion is not a proof that we cannot build a robot that is more capable than a human being. However, any claims that it is possible should be tempered by the respect that we have gained from studying the effectiveness of biological design. In this regard, it is interesting that some of the modern approaches to artificial intelligence consider the use of nanotechnology, which at least in part will make use of biological molecules and methods.

Finally we can say that the concept of an infinite human being may not be entirely lost. Even the lowly TM whose internal (table) complexity is rather small can,in arbitrarily long time and with an infinite storage, reproduce arbitrarily complex behavior. In this regard we should not consider just the complexity of a human being but also the complexity of a human being in the context of his tools. For example, we can consider the complexity of a human being with paper and pen,the complexity of a human being with a computer, or the complexity of a human being with access to a library. Since human beings make use of external storage that is limited only by the available matter, over time a human being, through collaboration with other human beings/generations extending through time,can reproduce complex behavior limited only by the matter that is available. This brings us back to questions of the behavior of collections of human beings, which we will address in Chapter 9.