

Complex Systems and Mechanistic Explanations

1. MECHANISTIC EXPLANATION

Our aim is to develop a cognitive model of the dynamics of scientific theorizing that is grounded in actual scientific practice. Our focus is on one kind of explanation, one involved in understanding the behavior of complex systems in biology and psychology. Examples of the complex systems we have in mind are the physiological system in yeast that is responsible for alcoholic fermentation, and the psychological system responsible for memory of spatial locations. As we shall discuss in this section, these explanations, which we refer to as *mechanistic explanations*, propose to account for the behavior of a system in terms of the functions performed by its parts and the interactions between these parts. The heuristics of decomposition and localization are central to our analysis of the development of mechanistic explanations. We shall discuss these heuristics in some detail in this chapter, especially in sections 2 and 3 of this chapter. Parts II and III will be concerned with illustrating their role, and also their development under other influences. These heuristics, or family of heuristics, can be thought of as imposing assumptions about the organization of the systems being explained. We shall examine these assumptions and the kinds of organization of actual systems for which these assumptions are likely to succeed and those on which they will likely fail.

By calling the explanations *mechanistic*, we are highlighting the fact that they treat the systems as producing a certain behavior in a manner analogous to that of machines developed through human technology. A machine is a composite of interrelated parts, each performing its own functions, that are combined in such a way that each contributes to producing a behavior of the system. A mechanistic explanation identifies these parts and their organization, showing how the behavior of the machine is a consequence of the parts and their organization. What counts as mechanistic, though, changes with social context. Scientists will appeal analogically to the principles they know to be operative in artificial contrivances as well as in natural systems that are already adequately understood. The state of technology and natural science at any given time thus plays a significant role in determining the plausibility and limits of mechanistic explanations (cf. Gregory 1981). From the universe of the *Timaeus*, through the Archimedian analogues of Galileo and the clockwork universe

of Newton, to the recent focus on servo-mechanisms and computers, the available analogues were important factors in determining which mechanistic models scientists advanced.

The nature and plausibility of mechanistic models is also influenced by characteristics of human thinking, especially by the proclivity of humans to trace operations in a linear or step-by-step fashion. This is especially evident when we consider the forms of organization possible. Many machines are simple, consisting of only a handful of parts that interact minimally or in a linear way. In these machines we can trace and describe the events occurring straightforwardly, relating first what is done by one component, then how this affects the next. Such machines induce little cognitive strain. Some machines, however, are much more complex: one component may affect and be affected by several others, with a cascading effect; or there may be significant feedback from "later" to "earlier" stages. In the latter case, what is functionally dependent becomes unclear. *Interaction* among components becomes critical. Mechanisms of this latter kind are *complex systems*. In the extreme they are *integrated systems*. In such cases, attempting to understand the operation of the entire machine by following the activities in each component in a brute force manner is liable to be futile.

A major part of developing a mechanistic explanation is simply to determine what the components of a system are and what they do. In broad outline, there are two strategies available to analyze and isolate component functions. The first is to isolate components physically within the system and then determine what each does (the goal is to use the knowledge of components to reconstruct how the system as a whole operates). The second strategy is to conjecture how the behavior of the system might be performed by a set of component operations, and then to identify components within the system responsible for the several subtasks. The former is the *analytic* method of Etienne Condillac, which played a role in the development and promotion of both Lavoisier's chemistry and Cuvier's functional anatomy. The latter is the explanatory program of functionalism in contemporary philosophy of mind (see Cummins 1983; Dennett 1978; Lycan 1981a, b), which for contrast we will refer to as a *synthetic* strategy (cf. Posner and McLeod 1982). An analytic strategy constructs from the bottom up; a synthetic strategy projects from the top down.

The analytic strategy is confronted by the fact that smoothly operating systems conceal their component operations. As we will see repeatedly in ensuing chapters, the breakdown of normal functioning often provides better insight into the mechanisms than does normal functioning. In the absence of natural error, failure can be induced. Simon develops the point clearly:

A bridge, under its usual conditions of service, behaves simply as a relatively smooth level surface on which vehicles can move. Only when it has been overloaded do we learn the physical properties of the materials from which it is built. (1981, p. 17)

Although overtaxing a system induces malfunction, which can be an important clue to the functional properties of component parts, this is a relatively crude method—one as likely to lead to catastrophic breakdown and chaos as to insight. For that reason, researchers prefer to differentiate components and their functions by altering specific activities *within* a system. Such *inhibitory* or *deficit studies* allow us to determine a physical component's contributions to the system by inhibiting its operations and then observing resulting deficits in overall system behavior. The best-known examples of such studies are ablation studies in nineteenth-century physiology (see Harrington 1987), but chemical poisoning studies in biochemistry, and the use of x-rays in the genetic research of Beadle and Tatum, follow the same logic.

Inhibitory studies are problematic for a number of reasons. Perhaps the most serious danger is the temptation to infer, from the fact that a specific experimental manipulation interrupted or inhibited a particular activity of the system, that the part of the system damaged is the component responsible for that activity. As R. L. Gregory pointed out over thirty years ago,

Although the effects of a particular type of ablation may be specific and repeatable, it does not follow that the causal connection is simple, or even that the region affected would, if we knew more, be regarded as functionally important for the output—such as memory or speech—which is observed to be upset. It could be the case that some important part of the mechanism subserving the behavior is upset by the damage although it is at most indirectly related, and it is just this which makes the discovery of a fault in a complex machine so difficult. (1961, p. 323)

Simplistic uses of functional deficit studies might lead us, for example, to conclude that a resistor in a radio is a hum-suppressor because the radio hums when the resistor is removed (Gregory 1968, p. 99).

Localization based on deficit studies is often erroneous. What is required is some means of figuring out, from the observed deficits, what the component in question positively contributes to the system when it is functioning normally. These are the functions that a good mechanistic explanation will localize in the parts. We will see that an important guide to finding these functions is the simultaneous use of a variety of inhibitory techniques. This can sometimes allow the investigator to determine component functions, but such an account will be complete and compelling

only when there is an explanation of how these components interact in effecting the normal operation of the system.

An alternative analytical technique follows an opposite approach, stimulating a physical component and observing the behavioral effects on the whole system. If extra stimulation produces an identifiable surplus, we can sometimes infer that under normal conditions the component was responsible for that which is now generated in excess. We shall generally refer to such investigations as *excitatory studies*. The best-known examples are stimulation studies in neuroscience in which electrical stimulation is applied directly to the cortex. Biochemical studies in which potential metabolic intermediaries are injected exhibit a similar logic. Once again, however, there is a temptation to infer that, because excitation to a physical system enhances or induces a particular effect, the stimulated part is necessarily the seat of the function in question. With both inhibitory and excitatory studies, the natural operation of the system is modified; the experimental techniques may be the source of experimental artifacts and may not be diagnostic of normal operations within the system. They are, however, the only techniques available in some cases. Even lacking a clear conception of the organization of the system, analytical strategies allow the experimentalist to probe the system and its organization.

A synthetic strategy requires some prior hypothesis about the organization and operation of the system. From an initial hypothesis about the underlying mechanisms, one formulates a model of how the system functions. The empirical task involves testing performance projected on the model against the actual behavior of the system. One discipline using model studies is AI, in which researchers propose and implement a set of operations in order to perform an activity that would require intelligence if performed by a human. Such model studies are also used in other domains. In a later section we will discuss biochemical research in which comprehensive proposals were advanced as to the intermediate chemical reactions in an overall physiological process. This theoretical work was sometimes supported by actual development of artificial systems intended to show that a mechanism such as the one being proposed could in fact carry out the needed process.

As plausible as such a model might be, it might also turn out to be radically misconceived. One such case is Justus Liebig's (1842) general account of nutrition. Liebig proposed a model of nutrition which plotted a complete set of metabolic transformations, based only on information about the chemical composition of food and waste products. He was guided by the assumption that because food substances are more complex than waste products, all that an animal does is break down these substances to release the energy stored in them. Using his knowledge of basic

chemical reactions, Liebig proposed a bold and brilliant model of animal metabolism. It was also wrong. Though consistent with the data Liebig used, it foundered on physiological data introduced by Bernard, who later showed that animals not only break down complex foodstuff, but also synthesize substances necessary for life (see Bechtel 1982).¹

The synthetic strategy is traditionally regarded as speculative. Perhaps it is. Speculative stages are nonetheless important to all science, if only in identifying possible mechanisms in contexts in which no workable mechanisms were envisioned. The subsequent testing of these models can and does induce the development and elaboration of more adequate models. However, speculation without empirical constraints is as likely to produce spurious explanations as correct ones.² In some cases, one might posit component operations that are, in fact, as complex as the overall operation. We will then have no net explanatory gain. In other cases, one might propose only one among many possible mechanisms and have no warrant for thinking it is the one actually utilized. This is hardly better than having no explanation at all.

The analytic and synthetic strategies are complementary. Inhibitory and excitatory studies can provide empirical data appropriate for evaluating synthetic models. Moreover, synthetic models can provide a theoretical framework in which to interpret information obtained from the empirical studies employing analytical strategies. Even together, though, the strategies hardly provide a fail-safe methodology. The dangers of spurious explanation and premature localization still confront the scientist.

Several factors can make the process of developing mechanistic explanations using techniques such as these extremely challenging. To begin with, as the level and significance of interactions increases, the complexity of the explanatory problems increases as well. The task of constructing an explanation for a given domain might be viewed as one of finding a sufficient number of variables, the constraints on the values of those variables, and the dynamic laws that are functions of those variables. These laws make it possible to use the model to predict future states of affairs from descriptions of an earlier time. The number of variables, once again, defines the dimensionality of the state space or problem space. A complete state description will pick out a single point in a multidimensional space corresponding to a specific value for each variable in the relevant part of the universe at the time. The dynamic laws are functions from states to subsequent states within this state space.

Let us focus just on the question of how complex the state space will be that we need to consider in order to represent the state of the domain. An increase in the number and significance of interactions posited in the constraints and laws requires an increase in the dimensionality of the state space needed to represent the domain. An example may help. Richard

Lewontin (1974) explains that in the case of genetics we can think of the problem as one of predicting the genetic composition of a population from its composition at an earlier time. As we consider greater numbers of loci and alleles, the number of variables needing to be considered increases and, hence, so does the dimensionality of the state space. What is pertinent to our purposes is how interaction affects the number of variables, and consequently the number of dimensions in the state space. If we assume that genes at different loci segregate independently, then the predictive problem reduces to one of finding solutions to independent problems for each locus, and accordingly the proper unit of analysis will be allelic frequencies. In that case, the dimensionality of the space needed to represent the state of the genome is a linear function of the number of alleles and the number of loci. With n loci and a alleles at each locus, the dimensionality of the problem space approximates the product of the two. If, on the other hand, we assume that there is significant linkage between genes at different loci, then, Lewontin urges, the proper unit of analysis will be gametic frequencies. As the number of loci increases, the number of possible gametes increases exponentially: in the case of two alleles at each locus, with two loci, there will be four gametic classes; with three loci, there will be eight gametic classes; and with ten, there will be over a thousand. The expected dimensionality for the state space will be one less than the number of allelic or gametic classes. This means that the anticipated dimensionality of the state space needed to represent the state of the genome will also increase linearly with allelic classes as the unit, and exponentially when using gametic frequencies. The results are detailed in Table 2.1.

Explanatory demands are further aggravated by the relative stability of the systems involved. This is part of what helps to conceal the contributions of the individual components when we examine a normally operating system. As we noted previously, if the goal is to reveal what the parts contribute to the operation of the whole system, we generally must find techniques to perturb the system. When investigating the behavior of *self-organizing* systems, the theorist must contend with the fact that these systems will maintain or determine an equilibrium even in the face of considerable perturbations, both internal and external. Consequently, establishing what the parts contribute will be difficult. In the limiting case, self-organizing systems will be homeostatic—that is, they will maintain a predetermined equilibrium state despite perturbations. The genetic system, in which endonucleases repair damage to nuclear DNA, is one example of such a homeostatic system. In more complicated cases, the systems are better thought of as *self-regulating*—that is, they will modulate activity at varying levels depending on other influences. For example, co-enzymes in cell metabolism adapt the breakdown of foodstuff to the work performed by the cell.

n	$a = 2$		$a = 5$		$a = 10$	
1	1	1	4	4	9	10^{+1}
2	2	3	8	24	18	10^2
3	3	7	12	124	27	10^3
4	4	15	16	624	36	10^4
5	5	31	20	3,124	45	10^5
6	6	63	24	15,624	56	10^6

Table 2.1. Increase in the Dimensionality of a Problem in Genetics as a Function of Linkage. With n loci and a alleles at each locus, the dimensionality of the problem depends critically on linkage. Assuming independent assortment (indicated in plain text), the proper unit of analysis will be allelic frequencies. The dimensionality of the state space increases as a linear function of n . With linkage effects (indicated in bold text), the proper unit of analysis is gametic frequencies. The dimensionality then increases exponentially. (Based on Lewontin 1974.)

When a system's behavior is relatively constant despite variations external to it, we can safely ignore the environment and focus only on the internal mechanisms to specify the parts, their interactions, and their contributions to the behavior of the system. But when the system adapts to the environment, as homeostatic and self-regulating systems do, we cannot simplify the task in this way. The sensitivity to environmental changes means that the parts operate differently under altered conditions and so further conceal from view how they behave when the system is operating normally. In general, interaction among the various components makes it difficult to isolate independent contributions from their coordinated output.

2. DECOMPOSITION AND LOCALIZATION

We now turn to the heuristic strategies of decomposition and localization. As we will see in the chapters to follow, these strategies have been used by a wide variety of researchers in a wide variety of disciplines, from nineteenth-century brain science and early twentieth-century investigations into chromosome structure to more recent work on language and cognition.

Decomposition allows the subdivision of the explanatory task so that the task becomes manageable and the system intelligible. Decomposition assumes that one activity of a whole system is the product of a set of subordinate functions performed in the system. It assumes that there are but a small number of such functions that together result in the behavior we are studying, and that they are minimally interactive. We start with the assumption that interaction can be handled additively or perhaps linearly.

Whether these assumptions are realistic or not is an open question; indeed, at the outset we often simply do not know. The extent to which the assumption of decomposability is realistic can be decided only a posteriori, by seeing how closely we can approximate system behavior by assuming it. We may be led to erroneous explanations, but it may be the only way to begin the task of explaining and understanding complex systems. The failure of decomposition is often more enlightening than is its success: it leads to the discovery of additional important influences on behavior.

Localization is the identification of the different activities proposed in a task decomposition with the behavior or capacities of specific components. In some cases we may be able to identify (through fairly direct means) the physical parts of the system in which we can localize different component functions. In other cases we may have to rely on various functional tools for determining that there are such parts, without being able to identify them; for example, we may be able selectively to inhibit their operation and observe the consequences on behavior. We need not assume that a single part in this sense is a spatially contiguous unit; in fact, we know that in many cases it is not. A functional unit may be distributed spatially within the system. Localization does entail a realistic commitment to the functions isolated in the task decomposition and the use of appropriate techniques to show that *something* is performing each of these functions.

In one extreme form, decomposition and localization assume that a single component within the system is responsible for some range of phenomena exhibited by the system. For example, it is assumed that the posterior cerebral lobes are responsible for vision, that the cell nucleus is responsible for genetic control, and that there is a specialized enzyme responsible for catalyzing a given chemical reaction within the cell. This, the simplest assumption, often guides the first explanatory models, even if it seldom survives. We refer to it as *simple* or *direct localization*, being simple both in focusing on single components and in imposing the fewest constraints. The simplest case, however, is often far too simple (cf. Ch. 4). The behavior to be explained may be at best the product of several independent components rather than of one. No one component can be assigned sole responsibility. It also may be necessary to assume there is some interaction and some differentiation of function. Lacking simple localization, the alternative is to localize a set of component functions and assume linear interaction will explain the behavior of the system. We refer to this as *complex* or *indirect localization*, having not only a complex organization, but complex constraints on the problem.

Pursuing decomposition and localization is to impose an assumption about the nature of the system whose activities one is trying to explain: it is assuming that it is *decomposable*. A decomposable system is modular in

character, with each component operating primarily according to its own intrinsically determined principles. Thus, each component is dependent at most upon inputs from other components, influences other components only by its outputs, and has a specific, intrinsic function. The notion of decomposability stems from Simon and constitutes the descriptive counterpart of localization. Localization presupposes that we are confronted with a modular organization such that the components of the system can be subjected to separate study and investigation; it requires that the components have discrete intrinsic functions intelligible in isolation, even if such functions do not independently replicate those of the system as a whole.

An extreme form of a decomposable system is an *aggregative system*,³ which is a species of *simply decomposable* system. System behavior in such cases is a linear or aggregative function of component behavior (cf. Levins 1970, p. 76). Whatever organization is present is not a significant determinant of the relevant systemic properties. In an important paper, "Forms of Aggregativity," Wimsatt lays down four conditions of aggregativity (1986a, pp. 260–68):

1. Intersubstitutability of parts;
2. Qualitative similarity with a change in the number of parts;
3. Stability under reaggregation of parts; and
4. Minimal interactions among parts.

We emphasize the last condition, primarily because the focus of our investigation is the discovery of organizational properties that fix the interaction of the parts and determine their significance for system behavior. However, we do not see a useful way to elaborate on this condition independently of the first three, and we think that—with a suitable decomposition—when this last condition is satisfied the other three generally will be satisfied as well.⁴ In offering the final condition, Wimsatt gives us this statement, with some reservations: "There are no cooperative or inhibitory interactions among parts of the system" (*ibid.*, p. 269). As an example, lateral inhibition in the retina leads to a lower activation among neurons than would be predicted if we attended only to their stimulation level; this is because an elevated activation level for one neuron decreases the activation level of adjacent neurons.

Few interesting dynamic systems are strictly aggregative. For example, fluid flow or the movement of a herd *approximates* aggregative motion, though even these *only* approximate aggregative systems and some of the most interesting work concerns why this is so. When the relevant systemic properties are at least partially determined by the organization of the system, we no longer have aggregativity. Or, more realistically, to the extent that organization determines systemic behavior, the system is non-

aggregative. In the simplest departures from aggregativity, we may still maintain intersubstitutability; however, when this also fails, we have what we call *composite systems*. There are two species of composite systems, which differ in terms of the role played by systemic organization: In *component systems*, the behavior of the parts is intrinsically determined. In these cases it is feasible to determine component properties in isolation from other components, despite the fact that they interact. The organization of the system is critical for the functioning of the system as a whole, but provides only secondary constraints on the functioning of constituents. In *integrated systems*, systemic organization is significantly involved in determining constituent functions. There may be, for example, mutual correction among subsystems, or feedback relations that are integral to constituent functioning. Thus, as we will see, although work on cell metabolism by, for example, Neuberger and Thunberg in the early decades of this century treated metabolic processes as linear and sequential, the discovery of coenzymes and their function has made it clear that linear models are oversimplified in the extreme (see Bechtel 1986a, and Chapter 7). Systemic organization then provides primary constraints on constituent functioning, and constituent functioning is no longer intrinsically determined. Some of the most interesting and bewildering problems arise with such systems. Richard Levins comments:

This is a [type of] system in which the component subsystems have evolved together, and are not even obviously separable; in which case it may be conceptually difficult to decide what are the really relevant component subsystems. Thus, for example, we might consider that a simpler multicellular organism is composed of cells, and yet the cells may be more profitably regarded, under other circumstances, as simply spatial subdivisions, partly isolated, of an organism. (1970, p. 77)

Dependence of components on each other is frequently mutual and may wholly blur any distinction among them. Thus, mitochondria were once independent organisms—though they are now clearly but parts of a cell, integrated into cell metabolism (Margulis 1970), and we cannot now understand how they function if we neglect their incorporation in, and integration into, the complex activities of the cell.

Composite and integrated component systems correspond to two types of organization in Simon's scheme (component systems are Simon's *nearly decomposable* systems). To the extent that components perform independent functions and send their outputs to other parts, we have *strict* or *simple decomposability*. *Near decomposability* imposes less stringent limits, as Simon explains:

- (1) In a nearly decomposable system, the short-run behavior of each of the component subsystems is approximately independent of the short-run behavior of

the other components; (2) in the long run the behavior of any one of the components depends in only an aggregate way on the behavior of the other components. (1969, p. 210)

For near decomposability, individual components must be controlled by intrinsic factors, in the manner of composite systems. As components become less governed by intrinsic factors, we enter the domain of integrated composite systems, which are *minimally decomposable*.

A system will be nearly decomposable to the extent that the causal interactions *within* subsystems are more important in determining component properties than are the causal interactions *between* subsystems (cf. Simon 1969, p. 209). Wimsatt (1972, p. 72) suggests that we characterize such systems in terms of a parameter that is a “measure of the relative magnitudes of intra- and inter-systemic interactions for these subsystems.” As we noted above, the heuristics of decomposition and localization assume a degree of decomposability. At a minimum they assume that the system is at least nearly decomposable. A parameter of the sort Wimsatt offers, then, would be an estimate of the likely error in predictions based on models developed using decomposition and localization (for further discussion, see Richardson 1982).

Whether decomposition and localization will succeed or fail in a given case, these heuristics are important because they provide us with a tractable strategy for attacking the explanatory problems complex systems present. Recent research in the psychology of judgment indicates that humans have great difficulty comprehending cases with more than a few interacting variables. Humans *cannot* use information involving large numbers of components or complex interactions of components, and even when the problem tasks are computationally tractable, human beings *do not* approach them in this way. Complex systems are computationally as well as psychologically unmanageable for humans.

3. HIERARCHY AND ORGANIZATION

We need now to explore possible reasons for thinking that the sorts of systems encountered in biology or psychology are likely to be decomposable or nearly decomposable—and hence amenable to mechanistic explanations developed through decomposition and localization—or only minimally decomposable or not decomposable at all—and hence not amenable to explanation using these heuristics. To explore this issue it is useful first to recognize that in talking about components and whole systems we are construing nature as incorporating a hierarchy of levels. When parts interact with each other, we can view this as a *horizontal* process; that is, there is interaction between units at one level. When we focus on how components combine into larger units, which in turn may interact with other

larger units, we are addressing a *vertical* question about the relations between levels. If the degree of interaction among components when they come to form wholes does not obliterate the components as autonomous entities, the result is a *decomposable part-whole hierarchy*.

A hierarchical organization also facilitates tractability by providing for theoretical economy. In explaining the behavior of the system we can gain independent characterizations of each component, ignoring both the contributions of other components at the same level as well as the influences operative at higher or lower levels. For example, programming in higher-level languages allows us to bypass the means by which the commands are executed. If the system is decomposable, there will be relatively little information lost in such a representation: "In studying the interaction of two large molecules, generally we do not need to consider in detail the interactions of nuclei of the atoms belonging to the one molecule with the nuclei of the atoms belonging to the other" (Simon 1969, p. 218).

Moreover, there is considerable evidence that information is held in human memory within organized structures and that the very intelligibility of a domain may depend on its being represented as a decomposable hierarchy. In work by Simon and his collaborators, it has been shown that experts in various disciplines differ from novices not just in the amount they know, but in the way their knowledge is organized. Expert chess players, for example, are readily able to reconstruct board positions of games that they have observed for only a few seconds, whereas novices are able to locate only a few pieces. This difference disappears, however, when the board positions do not make strategic sense. Chase and Simon (1973a, b) contend that the differences are due to the fact that experts recognize and remember patterns among pieces and treat these patterns as units (see also Gilmartin and Simon 1973). Analogously, in tasks evaluating the efficiency of recall and recovery, memory is facilitated by texts with specific forms of organization and inhibited by others. Free recall will even impose organization when none is present in the text.

Simon holds that hierarchical organization is also a general phenomenon in nature. He argues that hierarchies arise because the forces governing interactions between objects typically do not form an equally distributed continuum. The strongest forces govern interactions at the lowest level and give rise to reasonably stable units at a middle level. In many cases these lower-level forces may not determine which among a variety of complexes at the higher level will be realized. They constrain but do not determine the results. Weaker forces then come to play in determining the relationships between these middle-level entities.

As an illustration, Simon considers chemical forces. The forces responsible for atomic structure are stronger than those that determine the composition of the molecules made of them. Similarly, the forces determining

the structure of macromolecules are weaker than those that determine their composition.

Thus, protons and neutrons of the atomic nucleus interact strongly through the pion fields, which dispose of energies of some 140 million electron volts each. The covalent bonds that hold molecules together, on the other hand, involve energies only on the order of 5 electron volts. And the bonds that account for the tertiary structure of large macromolecules, hence for their biological activity, involve energies another order of magnitude smaller—around one-half of an electron volt. It is precisely this sharp gradation in bond strengths at successive levels that causes the system to appear hierarchic and to behave so. (Simon 1973a, p. 9)

An equilibrium between the forces at the lowest level defines a set of stable systems. The forces at the higher level then determine the relationship between these units and their combination into other units. In biochemical processes it has been clear since the work of Linus Pauling and Max Delbrück (1940) that quantum mechanical forces are insufficient to explain reactions among complex molecules in the cell; much weaker forces are required for the intermolecular interactions. Hydrogen bonds are especially important, both in antibody formation and in the classic work on the double helix by Watson and Crick (1953). The resulting structures again form into stable units, with their interrelations defined by still weaker forces.

The argument so far for nature comprising decomposable hierarchies assumes that the strengths of the forces for binding components into structures are not continuously distributed. This may or may not be true, but Simon offers other, more general, arguments. For example, he appeals to evolutionary considerations, arguing that complex systems are more likely to evolve if they are hierarchical and decomposable. He assumes that the lower-level forces insure that components can arise independently, existing as stable units in their own right. All that would then be necessary is the formation of stable combinations that would meet the demands at the higher level. Given complex macromolecules, living cells would be combinations formed from them, dependent only on the operation of higher-level forces. Selection could then serve to fine-tune the system without large-scale disruption. To quote once again from Simon,

The loose horizontal coupling of the components of hierarchic systems has great importance for evolutionary processes just as the loose vertical coupling does. The loose vertical coupling permits the stable subassemblies to be treated as simple givens, whose dynamic behavior is irrelevant to assembling the larger structures, only their equilibrium properties affecting system behavior at the higher levels.

The loose horizontal coupling permits each subassembly to operate dynamically in independence of the detail of others; only the inputs it requires and the outputs it produces are relevant for the larger aspects of system behavior. (1973a, p. 16)

To illustrate the principle, Simon (1969) tells the following tale of two watchmakers: Each makes fine watches with 1000 parts of diverse sorts. One of the two, Tempus, uses an hierarchical design with component parts: each watch has ten components, and each component has ten components, etc. The other, Hora, uses a horizontal design. All the parts must be in place before the watch will stay together. Each of the watchmakers is interrupted periodically to take orders for additional watches. Hora's work suffers dramatically, because every interruption results in a loss of all the work on the current watch. Tempus' work suffers too, though not so dramatically, because all that is lost is the work on the current subassembly. The moral is a general one: complex structures are more efficiently constructed if they are composed of stable subassemblies. Simon shows that additional levels of intermediate structure will further increase stability; consequently, the time required to assemble a system with any given number of units is inversely proportional to the number of intervening levels. He then goes on to apply the same principle to the evolution of biological units: decomposability increases the evolutionary rate (*ibid.*, pp. 200ff.). As a result, complex systems arise more readily when they consist of stable subsystems.

As we noted in Chapter 1, the assumption that a system is nearly decomposable and hierarchical is not just motivated by theoretical arguments; human cognitive strategies make such an assumption natural. This does not mean, however, that it is realistic. There are in fact reasons to suspect that many natural systems are not decomposable even though they are hierarchical. Simon concentrates on the division of systems into component parts. In minimizing interconnections between these parts, and treating them as autonomous, Simon sidesteps discussion of what binds parts together, making them *parts* of a complex system. If nothing imposes systemic structure, we have an aggregative system. Systemic behavior is an additive function of component behavior, at least if we ignore threshold effects. We have a hierarchy in name only. With composite systems, interaction between units is critical; indeed, it is constitutive of the higher-level units. Interaction is what makes composites useful for explanatory purposes; however, interaction also compromises the autonomy of components.

The mode of organization is important. Grobstein (1973) distinguishes between *facultative* and *obligate* organization, in a transparent analogy to social symbioses (cf. Richardson 1982). A facultative organization allows

members to disperse and recombine. Individuals can function as part of a more complex system, but are also capable of independent activity; for example, baboons will often forage in groups, but they are also able to forage independently. Facultative organization is thus nearly decomposable. An obligate organization, by contrast, is one in which interdependence has significantly compromised the capacity for independent activity, and the system is thus only minimally decomposable. As Grobstein says, the properties of higher-level structures "are in some sense immanent in the properties of the components, [though] . . . such properties tend to be lost if components of a set are dispersed or if a set is dissociated from its context or superset" (1973, p. 45). Some flowering plants are wholly dependent on birds or insects as vectors for pollination; and some of these birds and insects, in turn, are specialized, feeding on only one type, or a few types, of flowering plants. The latter form of organization is particularly important, as Grobstein recognizes:

Their components are very different in properties when in isolation or in the collective, and the collective, once formed, is not reversible. This is the case with most higher organisms. . . . A complex multicellular organism represents an extreme case in which very special conditions are required to maintain individual cells or individual organs outside of the collective relationship. (Ibid., p. 34)

Not only are the relationships among constituents of the system important for explaining the system's operation, but the constituents themselves have no independent, isolable function.

Simon's theoretical arguments supporting the ubiquity of nearly decomposable systems with facultative organization rest on evolutionary considerations. Such considerations, though, will not support the conclusion that complex systems are generally decomposable. Levins (1970) and Wimsatt (1972) point out that divergence and coadaptation will decrease the decomposability of a system with time: once aggregated, components can diverge and specialize in functions while maintaining stability. Evolvability does not insure stability; thus, while decomposable hierarchies may be more likely to evolve, the considerations advanced do not necessarily favor maintaining decomposability once they are formed. It will remain an open question whether complex natural systems are necessarily decomposable hierarchies.

4. CONCLUSION: FAILURE OF LOCALIZATION

Whether natural systems are hierarchically structured will influence how successful decomposition and localization will be as heuristic strategies. In this chapter we have described decomposition and localization as heu-

ristics for developing mechanistic explanations of complex systems and have examined the assumptions these heuristics make about the nature of the system we need to explain. While there are considerations favoring the occurrence of decomposable hierarchies, there are also considerations pointing toward only minimally decomposable, integrated systems. Thus, there are clearly risks in assuming complex natural systems are hierarchical and decomposable. There are always some risks that stem not from the assumption of decomposable hierarchies, but from specific errors in developing the decomposition and localization. In a case of false localization, a complex system may manifest a component organization, but not the specific component analysis attributed to it. In this case the way the system operates is misrepresented. This may be because there is an alternative component organization at the same level, or because we have adopted the wrong level of analysis altogether. We will consider cases subsequently in which the initial analysis is misguided in these ways.

More radical errors are also possible. The separation of systems into isolated components, with the attendant minimization of interactive importance, may blind us to critical factors governing system behavior; in particular, it may blind us to the importance of systemic interaction. We may not have a decomposable system at all, or we may have one that is only minimally decomposable. Simon acknowledges the risk saying, "If there are important systems in the world that are complex without being hierarchic, they may to a considerable extent escape our observation and understanding" (1969, p. 219). The risk, if realized, should be felt in failures of explanation. If the failures are more limited, we are only limited in our explanations. We will also examine failures of a more radical sort in which, though research began with the assumption of near decomposability, high degrees of organization were subsequently recognized and the explanatory approach had to be adapted to accommodate this organization. Though decomposability may be a natural and fruitful starting point, it may be no more than that.