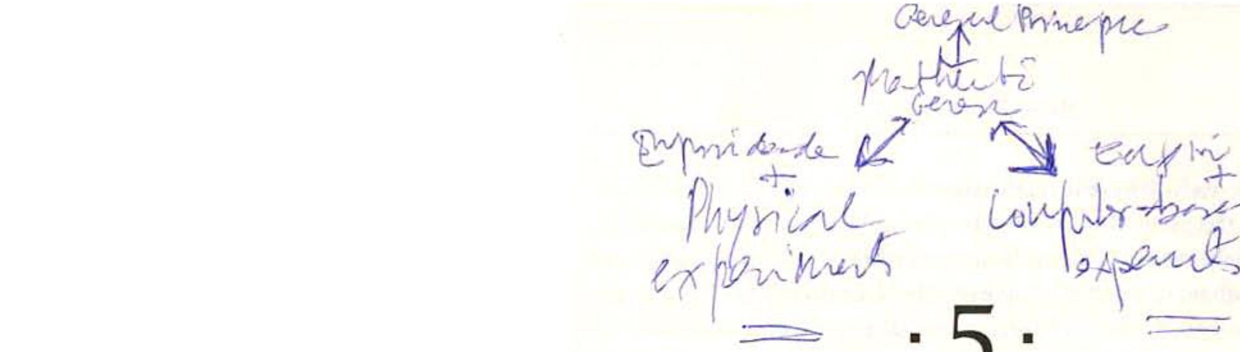
"*A monumental achievement...*" —**Douglas Hofstadter**

JOHN H. HOLLAND

HIDDEN ORDER

How Adaptation Builds Complexity



· 5 ·

Toward Theory

ALMOST ALL OF OUR EFFORT to this point has been spent in getting to, and designing, the halfway house represented by Echo. Now we look to the destination—general principles. Although that destination is still on the horizon, there are useful landmarks, and those of us who have been studying *cas* at the Santa Fe Institute are optimistic about the way ahead. We believe that there *are* general principles that will deepen our understanding of *all* complex adaptive systems. At present we can only see fragments of those principles, and the focus shifts from time to time; but we can see outlines, and we can make useful conjectures. Just what can we see and imagine?

Mathematics is our sine qua non on this part of the journey. Fortunately, we need not delve into the details to describe the form of the mathematics and what it can contribute; the details will probably change anyhow, as we close in on the destination. Mathematics has a critical role because it alone enables us to formulate *rigorous* generalizations, or principles. Neither physical experiments nor computer-based experiments, on their own, can provide such generalizations. Physical experiments usually are limited to supplying input and constraints for rigorous models, because the experiments themselves are rarely described in a language that permits deductive exploration. Computer-based experiments have rigorous descriptions, but they deal only in specifics. A well-designed mathematical model, on the other hand, generalizes the particulars revealed by physical experiments, computer-based models, and interdisciplinary comparisons. Furthermore, the tools of mathematics provide rigorous derivations and predictions applicable to all *cas*. Only mathematics can take us the full distance.

The Separation between Observation and Theory

To see more clearly the distance between observation and theory for *cas*, let's look again at an example—this time concerning sustainability.

Early in this century the supposedly inexhaustible forests of the Upper Peninsula of Michigan were cut down, reducing most of the area to a barren stumpland. Then, during the depression of the 1930's, the Civilian Conservation Corps (CCC) was formed to reduce the devastating effects of unemployment in the cities. Over several years, at a surprisingly low cost to the government, the CCC (many of whose members in this region were drawn from Detroit) planted seedlings throughout vast tracts of the Upper Peninsula. Now, half a century later, the land is once again forested, to the great benefit of tourism and the lumber industry (more cautious this time around). Extensive interviews of former CCC members several decades later show that almost all of them look on this period as a turning point in their lives.

We would seem to have here a prime example of a lever point in a political-economic context. But questions abound. Would this procedure be repeatable, at least in outline, if we replaced Detroit and the Upper Peninsula with Los Angeles and the forests of the Northwest? Is this an example of a broader class of symbiotic solutions coupling inner-city problems with resource sustainability? More generally, what combined circumstances in economics and politics make such long-horizon investments possible? Must they always be centered on some disaster, as in our earlier example of the San Francisco earthquake and public transport? Why do those working with renewable resources, such as forests and fish, exhaust those resources when they know (as they do) that the action destroys their livelihood? Is this somehow connected with the downside of the Prisoner's Dilemma?

The last two of these questions have anecdotal answers. We talk of the "tragedy of the commons," where some common resource is overrapidly exploited by everyone, because each person mistrusts the moderation of others. That is indeed reminiscent of the defect-defect solution of the Prisoner's Dilemma. And we talk of the "mobility of capital," where the investors in an industry are distinct from the "locals" (the workers and owners), so the investors simply reinvest in some other industry when the local industry collapses. The investors don't suffer the consequences of the collapse, at least in the short run, so they show little concern. These answers have more substance than, say, the pundits' reasons for the rise or fall of today's stock market, but we have no firm basis for knowing when, or if, they apply.

We could, with substantial effort, model situations like this in Echo. A flight simulator version would be particularly helpful, letting the politician or economist observe the short-term and long-term outcomes of policies they consider feasible. Still, that is not really enough. We would do much better with guidelines that suggest where to look. We need some way of searching beyond familiar policies, which may offer little or may be caught in a legislative deadlock. The space of possible policies is large, and there may be some that exploit lever points, if we can just uncover them. But lever points, at least in our examples, are often obscure and not easily located by trial-and-error exploration. In these cases, theoretical guidelines relating lever points to specifics of the problem would be an invaluable help.

Two-Tiered Models

The first step in moving toward an appropriate theory is, once more, careful selection of mechanisms and properties from a multitude of possibilities. It is helpful to recast the problem in a framework, such as Echo, that relies on selected mechanisms common to all *cas*. It is particularly helpful if the model is kept simple, while retaining salient features of the problem that aim at thought experiments rather than a full flight simulator. We can still keep looking toward theory, favoring elements that can be mathematicized, where this can be done without jeopardizing relevance.

Consider the CCC example. A major part of the simulation in Echo would center on the action of one set of agents (inner-city workers) as catalysts for the recovery of another set of agents (the trees), after the first set had moved from one site (Detroit) to another (the Upper Peninsula). Here we are dealing with the consequences of *flows* (Chapter 1). We are also dealing with differing timescales. The workers move and act on one timescale, call it a "fast dynamic," while the trees recover on a much longer timescale, a "slow dynamic."

With the help of Echo, we can recast the problem in terms of flows of resources between different kinds of agents, as is true of most *cas* problems. We can make solid contact with mathematical models if we make two simplifying assumptions: (1) the agents can be usefully aggregated into species or kinds, and (2) there is a rapid mixing of resources among agents of like kind. With respect to the first assumption, the hierarchical organization typical of *cas* usually makes aggregation easy and natural. (See, for example, the discussion of default hierarchies in Chapter 2.) The second assumption assures that the consequences of interactions are rapidly distributed within each aggregate. Rapid distribution, in turn, assures that we can assign average resource levels to aggregates at each instant, without being stymied by nonlinear effects *within* the aggregate. Under these assumptions we can treat Echo-based models (and complex adaptive systems) in a kind of two-tiered format.

THE LOWER TIER

The lower tier concerns itself with the flow of resources between agents of different kinds. The combination of rapid mixing within each kind, and random contact between kinds, makes possible a mathematical model much like the billiard ball model discussed in the first chapter. That is, we can treat each *kind* of agent as a *kind* of billiard ball, and for each pair we can determine a reaction rate. The rate is directly determined by the exchange condition and the exchange scoring mechanism specified for each agent in Echo (see model 2 in Chapter 3). The result is an array of reaction rates (see *Nonlinearity* in Chapter 1).

Once this array has been computed, we are close to having a mathematical model that describes changes in flow over time. In particular, we are close to describing mathematically the change in the proportion of each kind of agent at a site, as time elapses. The relevant vehicle is the version of the Lotka-Volterra equations discussed in the nonlinearity example. Those equations let us determine the changes in *proportion* of each agent-kind by using the reaction rates for various possible pairs. However, we face a problem. The flow model gives the *total* resources held by each agent-kind, but the equations require the *proportion* of each agent-kind. Different kinds of agents use different amounts of resources in their structures, so aggregate resource totals do not directly determine agent-kind proportions. To derive the proportions, we must divide the aggregate resource totals by the amounts of each resource required to make a copy of that kind of agent.

The rapid mixing assumption now lets us treat the resource totals as equally shared by the individuals in each aggregate. Specifically, the rapid mixing assumption ensures that all reservoirs in the aggregate hold similar amounts of each resource. Knowing this, we can determine the number of agents in the aggregate by dividing the total resources held by the number of each kind of resource required to build that agent's chromosome. Then, knowing the *number* of individuals of each kind, we can determine their *proportions* in the total of all individuals. Having determined the proportion, we can use the Lotka-Volterra equations as a mathematical description of the changing resource flows mediated by the agents.

Even at this preliminary level, some theoretical progress can be made concerning lever points. Because agents can have surpluses of some resources, only certain resources held by the aggregate "count" toward the number of any given agent-kind. The notion of a "bottleneck resource" emerges. A close look at the flow model shows that a change in the bottleneck resource—say a new interaction greatly increases its level—can have much the effect of a mutation. It can open a cascade of new interactions. Changes in a bottleneck resource often give rise to effects far out of proportion to the change.

To adopt a term from physics, the lower tier gives us a mathematical model of the fast dynamics of the system.

The Upper Tier

For a mathematical theory of *cas* to be effective, the fast dynamics of the flows must be successfully coupled to the slow dynamics of long-term adaptation and evolution. In this two-tiered model, it is the upper tier that specifies the evolution of the agents. It uses a genetic algorithm to change the structures of offspring, as described at the end of Chapter 2. In Echo the resulting agent structures precisely determine the amounts of resource exchanged, so the reaction rates of the lower tier are directly coupled to the results of actions in the upper tier. Note that a change in the definition of the agent-kinds (aggregations) used in the lower tier will result in different couplings to the upper tier.

In selecting the aggregations and couplings to the lower tier, we want to make it easy to see how the network changes when the genetic algorithm causes given building blocks (schemata) to spread and recombine. One extreme would be to allow one node in the network for every distinct agent. Then the lower tier would be formally correct, but the patterns of change would be spread over large numbers of nodes. At best, the patterns would be difficult to discern. The lower tier only becomes useful, both computationally and theoretically, when we can aggregate agents into kinds based on the presence or absence of the chosen building blocks. Then the patterns of change relative to these building blocks will be manifest. This is the burden of the earlier "useful aggregation" assumption (look back again at Chapter 1).

Aggregation of agents, however, raises a problem similar to our earlier difficulty with aggregation of resources. For a given pair of agents, we can directly determine a flow of resources and a reaction rate (as detailed in Chapter 3). However, this is not necessarily an appropriate reaction rate for the pair of *aggregates* to which these agents belong. Agents of a given kind will not generally exchange resources in identical fashion; after all, we only collected them into a common kind because they had *some* building blocks in common. So two agents of the same kind may have different associated reaction rates. This puts us squarely into the difficulty discussed under the topic of nonlinearity in Chapter 1. We cannot simply average the reaction rates of individuals of a given kind to get a reaction rate for the aggregate agent-kind. That is, reaction rates associated with the flow network are not simply related to reaction rates associated with agent pairs.

We *can* determine a useful reaction rate for an agent-kind if the constituent agents are not too different from one another relative to their ability to exchange the resources of interest. In this instance the individual reaction rates are close to one another, so that the flow calculated with the average rate will not differ greatly from the actual flow. (The actual flow is determined by summing the individual flows of individual agents.) At worst, we can establish that no agent has a reaction rate slower (larger) than a determined amount, allowing us to determine bounds on the flow, rates of reproduction, and the like.

Keeping the individual reaction rates in an aggregate close to one another actually is largely under the control of the theorist setting up the two-tiered model. That person selects the characteristics that group the agents into aggregates. By selecting appropriate characteristics, the theorist can limit the variation in the individual reaction rates within each aggregate. The building blocks of the exchange conditions and the interaction tags are central to this purpose. By aggregating agents with the same alleles for these building blocks, the theorist can assure closeness of reaction rates, while benefiting from a simplified lower tier.

In sum, one way to generate a useful coupling of the upper tier to the lower tier is to aggregate agents with similar building blocks in the parts of the chromosome devoted to the offense tag, the defense tag, and the exchange condition. If we further constrain these aggregates by conditional replication, we achieve something much like biological speciation. Patterns should be sharpened because aggregates cannot blend into one another. In any case, the upper tier has the effect of continually changing the flow network of the lower tier, as the agents evolve and adapt under the genetic algorithm.

A Theory of Two Tiers

The relevant theory for the upper tier starts with the schema theorem for genetic algorithms because that theorem tells us about the spread and decline of building blocks. However, the version of the theorem given at the end of Chapter 3 is only a beginning. We need a version of the schema theorem that holds for the implicit fitness of the Echo models. And the theorem should tell us about the spread of schemata across kinds, with particular attention to the effects of selective mating. This element is important if we are to understand the spread of building blocks in real *cas*, such as the spread of the Krebs energy transformation cycle throughout the vast range of aerobic organisms or the spread of computer chips throughout machines ranging from automotive engines to cameras.

Given the perpetual novelty of agents in the Echo models, we need still more from a satisfactory theory. The unfolding development of an Echo world is a trajectory through a space of multiple possibilities; we need to know something of the form of this trajectory, particularly because *cas* rarely reach end points or equilibria. We are likely to understand a *cas* process only if we know what the trajectory looks like along the way.

It will be difficult, perhaps impossible, to predict details of the trajectory, but surely it is far from a random walk. At worst, we may face a phenomenon similar to the day-to-day, month-to-month changes in weather, though I think *cas* are more predictable than that. Even with the weather, there are building blocks—fronts, highs and lows, jet streams, and so on—and our overall understanding of changes in weather has been much advanced by theory based on those building blocks. It is still difficult to predict detailed weather changes, particularly over an extended period. Nonetheless, theory provides guidelines that lead us through the complexity of atmospheric phenomena. We understand the larger patterns and (many of) their causes, though the detailed trajectory through the space of weather possibilities is perpetually novel. As a result, we can do far better than the old standby: predict that "tomorrow's weather will be like today's" and you stand a 60 percent probability of being correct. A relevant theory for *cas* should do at least as well.

Complex adaptive systems exhibit more regularities than weather for at least two reasons. First, there is the persistence of favored building blocks. (In biological systems, the Krebs cycle is pervasive in both space and time; in economies, taxes too are pervasive in space and time.) Second, there is the phenomenon known in biology as *convergence*, which imposes further predictable regularities. Convergence in this sense should not be confused with the attainment of end points (fixed points), the subject of mathematical convergence. Here convergence refers to the similarity of agents occupying similar niches. With some knowledge of the niche, we can say something of the form of the agent that will occupy it. As an example, biologists recently discovered a tropical flower with a throat of unprecedented depth, a flower belonging to a genus invariably pollinated by moths. The niche provided by this flower led the scientists confidently to predict the existence of a moth, yet to be found, with a proboscis of equally unprecedented length.

The regularities provided by building blocks and (biological) convergence imply regularities in the development of the flow network. These, in turn, imply that agents attain high concentrations at certain kinds of nodes. New variants are most likely to arise where there are many agents; more samples mean more possibilities for variation. Accordingly, the generation of new agent-kinds (nodes) should center on these well-populated nodes, a kind of *adaptive radiation*. So we have some hints about how the network would grow. If the fast dynamic is modeled by a set of equations of the Lotka-Volterra form, this growth means adding new equations to the set. The added equations produce corresponding changes in the dynamics. To couple this growth to the upper tier, we need a version of the schema theorem that takes selective mating into account, while using only endogenous fitness. Such a theorem would let us determine something of the form of the trajectory through the space of lower-tier flow networks. It could give us some idea of what convergence means in this general setting, a setting that holds for all complex adaptive systems.

A Broader View

This two-tiered model undoubtedly captures a substantial portion of what is going on in *cas*. Yet we are only starting to give it the precision required for mathematical theory. Two advances in mathematics would help provide a theory of this two-tiered model. One is an organized theory of a dynamics based on *sets* of equations that change in number (cardinality) over time. Another is a theory that relates generators (building blocks) to hierarchical structure (for example, default hierarchies), strategies (*classes* of moves in games), and the "values" associated with those strategies (game payoff).

Now an aside, for those conversant with mathematics. Such a mathematics would resemble the use of generating functions to estimate parameters of stochastic processes (see Feller, 1950). Its combinatorial aspect would have the flavor of the work on "automatic" (automaton) groups (see Baumslag, 1994). The stochastic aspect can be studied with the help of Markov processes, but the usual treatment of such processes, which concentrates on eigenvectors and fixed points, will *not* be of much help. Instead, we need to know what happens to aggregates during the transient part of the process. Aggregation of states of the full process encounters the usual difficulties with nonlinearities; still, there are ways around this that may enable us to deal with perpetual novelty (see, for example, Holland, 1986). A successful approach combining generating functions, automatic groups, and a revised use of Markov processes should characterize some of the persistent features of the far-from-equilibrium, evolutionary trajectories generated by recombination.

Whatever our mathematical approach to *cas*, the objective remains to determine common causes of common characteristics. When we embarked, I listed three mechanisms—tags, internal models, and building blocks—and four properties—aggregation, nonlinearity, flows, and diversity—that have become the prime candidates for causes and characters in my own search. Other researchers will have other candidates. Nevertheless, at the Santa Fe Institute I think we would all agree on the following broad requirements for a successful approach to theory:

1. *Interdisciplinarity*. Different *cas* show different characteristics of the class to advantage, so that clues come from different *cas* in different disciplines. In this exposition we've seen many comparisons and the uses to which they can be put.
2. *Computer-based thought experiments*. Computer-based models allow complex explorations not possible with the real system. I have pointed out that it is no more feasible to isolate and repeatedly restart parts of a real *cas* than it is to test flameouts on a real jet airplane carrying passengers. Computer-based models make counterpart experiments possible. Such models can provide existence proofs, which show that given mechanisms are sufficient to generate a given phenomenon. They can also suggest critical patterns and interesting hypotheses to the prepared observer, such as conditions for the existence of lever points.
3. *A correspondence principle*. Bohr's famous principle, translated to *cas*, means that our models should encompass standard models from prior studies in relevant disciplines. Two advantages accrue. Bohr's principle assures relevance of the resulting *cas* theory by requiring it to incorporate hard-won distillations and abstractions from well-established disciplines. It also forestalls what I call "eye of the beholder" errors. Those errors occur when the mapping between a simulation and the phenomena being investigated is insufficiently constrained, allowing the researcher too much freedom in assigning labels to what are, after all, simply number streams in a computer. Standard models from well-established disciplines constrain this freedom because they have been developed with a standard mapping in mind.
4. *A mathematics of competitive processes based on recombination*. Ultimately, we need rigorous generalizations that define the trajectories produced by the interaction of competition and recombination, something computer-based experiments cannot provide on their own. An appropriate mathematics must depart from traditional approaches to emphasize persistent

features of the far-from-equilibrium evolutionary trajectories generated by recombination.

I believe this amalgam, appropriately compounded, offers hope for a unified approach to the difficult problems of complex adaptive systems that stretch our resources and place our world in jeopardy. It is an effort that can hardly fail. At worst, it will disclose new sights and perspectives. At best, it will reveal the general principles we seek.