

# chapter SIX

## Thinking Is Social

The linkage between social science simulations and computer programs for engineering and other applications is not always obvious. Neural networks, simulated annealing, cultural algorithms, ant colony optimization, and evolutionary algorithms are several instances where psychological, physical, and biological theories have influenced the development of computational methods for problem solving.

This chapter takes a simulation from the social sciences and shows how it can be modified slightly to perform combinatorial optimization. The adaptive culture model is extremely straightforward and easy to understand; further, it contains most of the ingredients that will be used in the more sophisticated particle swarm algorithms studied in the rest of the book. ■

## Introduction

The problem with the story about the blind men and the elephant is that it assumes they were also deaf. The famous tale, set in verse in the 19th century by John Godfrey Saxe (1869/1936), describes the experiences of six blind men who, converging upon an elephant, each come to believe that the entire elephant resembles the particular part he has encountered by chance:

The First approached the Elephant,  
And happening to fall  
Against his broad and sturdy side,  
At once began to bawl:  
"God bless me! but the Elephant  
Is very like a wall!"

The second blind man, feeling a tusk, cries that the elephant is like a spear, and so on. Of course the moral of the story is that people derive incomplete beliefs from their incomplete experiences in the world. In reality, if a group of blind men took turns announcing a description of his particular part of the elephant, it is clear that by listening to each other all the blind men would come to a rather complete and correct understanding of the heterogeneous qualities that make up an elephant. All members of the group would know that the creature has a side like a wall, tusks like spears, legs like trees, and so on. Through discussion they might even figure out how the parts are connected and how they function together. The point of this literary critique is that societies are able to benefit from the sharing of individuals' partial knowledge, resulting in a body of facts and strategies that far exceeds what any individual could have ever acquired independently.

The theme of the present chapter is that thinking is a social activity; human culture and cognition are aspects of a single process. People learn from one another not only facts but methods for processing those facts. The theme is not new: Bandura, for one, has explicitly theorized about the learning that occurs when individuals observe one another's behaviors:

If knowledge could be acquired only through the effects of one's own actions, the process of cognitive and social development would be greatly retarded, not to mention exceedingly tedious . . . Fortunately,

most human behavior is learned through modeling . . . The capacity to learn by observation enables people to expand their knowledge and skills on the basis of information exhibited and authored by others (Bandura, 1986, p. 47).

### **Adaptation on Three Levels**

Not only do people learn from one another, but as knowledge and skills spread from person to person, the population converges on optimal processes. The present chapter describes a system that operates simultaneously on three levels:

- Individuals learn locally from their neighbors. People are aware of interacting with their neighbors, glean insights from them, and sharing their own insights in turn, and local social learning is an easily measured and well-documented phenomenon.
- The spread of knowledge through social learning results in emergent group-level processes. This sociological, economic, or political level of phenomenon is seen as regularities in beliefs, attitudes, behaviors, and other attributes across individuals within a population. A society is a self-organizing system with global properties that cannot be predicted from the properties of the individuals who make it up.
- Culture optimizes cognition. Though all interactions are local, insights and innovations are transported by culture from the originator to distant individuals; further, combination of various innovations results in even more improved methods. This global effect is largely transparent to actors in the system who benefit from it.

### **The Adaptive Culture Model**

In an earlier chapter we discussed Robert Axelrod's influential work in the 1980s with the prisoners' dilemma and the "evolution of cooperation." In 1997, this same Axelrod published a volume titled *The Complexity of Cooperation*, in which he proposed a computational model of the dissemination of culture. The present chapter extends Axelrod's computer simulation of the spread of features through a culture to demonstrate how social interaction might comprise a natural computation

method that results in coherent and intelligent human thought, opinion, and action. We are considering the Axelrod model and the version we derive from it to contain the fundamental principles of the swarm algorithms that will be discussed in the following chapters. The current model has a kind of simplicity and elegance that makes it ideal for introducing the more challenging algorithms.

It has been theorized elsewhere that collaboration among individuals can result in cognitive optimization, that is, that social interaction may enable individuals to arrive at effective structures and strategies for the management of knowledge. The view has been supported in part by a paradigm called *particle swarm adaptation*, computer simulations of populations interacting in a multivariate real-number space; through imitation of successful others, individuals evolve optimal weights or activation values in connectionist and other cognitive models.

While the particle swarm algorithm has been found effective for optimization of continuous and binary phenomena, many cognitive theories include discrete variables. A major instance occurs in symbol-processing models, where discrete elements represent symbolic entities. Further, many connectionist models, especially those based on Hopfield's earlier network, are essentially combinatorial in nature; even if activations range between lower and upper limits during optimization, when a stable state is attained, activation values are usually seen to have settled in the corners of the hypercube; that is, they approach their limits of +1 and 0 (or -1, depending on the specification of the model), and variables are essentially discrete.

Axelrod's recent simulation of the spread of culture provides insights into the effects of social interaction and gives us a starting point for demonstrating that a small number of exceedingly simple principles can cause an artificial system to behave remarkably like a complex human society. Axelrod's culture model and the particle swarm are two branches of the same tree, with the culture model simulating societies in terms of discrete variables and the particle swarm algorithm operating on continuous or binary ones. The two approaches complement one another and point the way to a theory of culture and cognition emerging from interaction. In the present chapter, an expansion of Axelrod's algorithm called the Adaptive Culture Model (ACM) is shown to be able to optimize complex functions, and it is suggested that cognitions, attitudes, and other arrays of psychological phenomena are optimized by interaction among individuals. A series of experiments are reported that test the capabilities of the algorithm. Several cognitive theoretical models are simulated, including a new kind of multivalued combinatorial constraint satisfaction network.

## Axelrod's Culture Model

Axelrod has theorized that similarity between pairs of individuals can result in the spread of culture. In his simulations, individuals are represented as strings of symbols called "features"; the number and length of the strings and the universe of symbols available to them are parameters of the system. For example, if individuals comprise five features, and these are defined as numerals in  $\{0, 1 \dots, 9\}$ , then one individual may be represented as 42237 and another as 99217. A two-dimensional matrix of individuals is initialized into the simulation, and they are allowed to interact.

Axelrod postulates that the probability of human interaction is a function of the similarity of two individuals: "The basic idea is that agents who are similar to each other are likely to interact and then become even more similar" (Axelrod, 1997). For instance, in the example above, the two individuals are similar in the third and fifth positions; with 40 percent similarity they have a 0.40 probability of interacting.

An interaction in ACM occurs when one individual adopts a non-matching feature of the other. An individual and one of its horizontally or vertically adjacent neighbors are selected at random. If stochastic similarity criteria are met, the selected individual will change one of its elements to match the symbol in the same position of the neighbor's string. For instance, if the two individuals above interacted, 42237 could take the 1 from its neighbor, and become 42217. The element changed is selected stochastically from the elements that are different. (This way of selecting elements to change is the same for the rest of the examples in the chapter.)

As a simulation iterates, neighbors are observed to begin resembling one another, until regions of the matrix contain identical strings. Axelrod focuses largely on the group-level effects of various parameters on the formation of regions, for instance, the numbers and sizes of regions as a function of number of features and population size. Figures 6.1 and 6.2 show "before" and "after" snapshots of a population of five-featured individuals. In this run and the others following, the beginnings and ends of rows and columns are considered adjacent, in a torus field, so that all individuals have four neighbors with whom they may interact. Note that fields do not usually wrap in Axelrod's simulations; this may result in some differences in results between the two implementations, but these do not affect the theoretical implications of the findings. Because similarity sets the threshold for instigating changes, boundaries between regions containing strings with no matching members

27217	74924	31157	53671	22660	37316	07959	57666	33206	92725
66219	08226	26707	45600	48767	39481	62784	89859	27792	35492
37262	66163	89178	60968	91098	19937	62103	07562	03500	13864
87746	66209	94122	72784	03593	16647	19776	87819	22160	48185
16880	09713	76057	30843	92125	41152	74156	98801	64760	00144
86287	66161	23271	46773	53014	44442	25424	98309	32553	16678
90624	65685	68785	32385	90770	24676	68806	25347	16640	30602
98681	11402	57304	68003	16943	01041	44693	63237	76040	61075
52249	30617	91425	92780	82342	30467	19721	84117	96595	55215
79949	70851	29089	89311	19176	67653	95954	64805	51332	74301

**Figure 6.1** Initial random start for a simulation of a 10 × 10 population of individuals made up of strings of five features represented by numerals ranging from zero to nine.

*Sim. by 12/20/13*

22233	22233	22233	22233	22233	22233	22233	22233	22233	22233
22233	22233	22233	22233	22233	22233	22233	22233	22233	22233
22233	65955	65955	22233	22233	22233	22233	33588	33588	33588
22233	22233	22233	22233	22233	22233	22233	33588	33588	33588
22233	22233	22233	22233	22233	22233	22233	22233	22233	22233
22233	22233	22233	22233	22233	22233	22233	22233	22233	22233
22233	22233	22233	22233	22233	22233	22233	13157	22233	22233
22233	22233	22233	22233	22233	22233	22233	13157	22233	22233
22233	22233	22233	22233	22233	22233	22233	22233	22233	22233
22233	22233	22233	22233	22233	22233	22233	22233	22233	22233

**Figure 6.2** Result of simulation where interaction is a stochastic function of similarity (Axelrod's paradigm).

*Sim. by 12/20/13*

eventually become fixed, and change in the system stops. This kind of polarization is very similar to that noted in the simulations of Nowak, Szamrej, and Latané's 1990 paper and other social impact studies.

The present chapter remains conceptually close to Axelrod's original, insofar as the simulations model populations of individuals within a society who interact by exchanging features. "Features" in both models may comprise a variety of phenomena: "Although beliefs, attitudes, and

behaviors cover a wide range indeed, there are even more things over which interpersonal influence extends, such as language, art, technical standards, and social norms." Axelrod moves between levels of analogy when he simultaneously describes the symbol strings as persons and as "sites," which might be thought of as neighboring villages. The model is described in this chapter on the level of individual persons interacting with their neighbors, which does appear to be Axelrod's primary focus as well.

### Experiment One: Similarity in Axelrod's Model

Axelrod theorizes that similarity is a precondition for social interaction and subsequent exchange of cultural features. In the simulations that instantiate his theory, the probability of interaction depends on similarity, and culture is seen to spread and finally stabilize through links between similar individuals. According to Axelrod and others, people become more similar as they interact; an apparent paradox, however, is that populations do not converge on unanimity. Instead, subgroups tend to become more homogeneous over time and more different from one another.

The "birds of a feather" hypothesis—that people are attracted to and influenced by others who are similar to them—has been supported by much research, especially in the attitude similarity paradigm promulgated by Byrne. On the other hand, Wetzel and Insko have shown that people are attracted to others who resemble their ideal, rather than actual, selves. Over six carefully conducted experiments, when the effects of "ideal similarity" were statistically removed from data, the effect of "self-similarity" was nonsignificant. Wetzel and Insko's conclusion was that people strive to attain their ideals; in the meantime they find others who approximate their ideals to be attractive and persuasive. At the least, there is evidence to suggest that the role of similarity is not as important as Axelrod theorizes; as he notes, few models of social influence give the factor much weight. As Axelrod purports to simulate the social influence process, that is, the propagation of beliefs, attitudes, and other features from one individual to another, it is prudent to question the causal position of similarity in the paradigm.

In the first experiment, Axelrod's model was altered slightly: the effect of similarity as a causal influence was deleted from the model. This was accomplished easily by setting the probability of interaction to 1.0

```

70971 70971 70971 70971 70971 70971 70971 70971 70971 70971
70971 70971 70971 70971 70971 70971 70971 70971 70971 70971
70971 70971 70971 70971 70971 70971 70971 70971 70971 70971
70971 70971 70971 70971 03593 70971 70971 70971 70971 70971
70971 70971 70971 70971 70971 70971 70971 70971 70971 70971
70971 70971 70971 70971 70971 70971 70971 70971 70971 70971
70971 70971 70971 70971 70971 70971 70971 70971 70971 70971
70971 70971 70971 70971 70971 70971 70971 70971 70971 70971
70971 70971 70971 70971 70971 70971 70971 70971 70971 70971
70971 70971 70971 70971 70971 70971 70971 70971 70971 70971

```

**Figure 6.3** Result of a simulation of the culture model with no criterion for interaction.

for all selected pairs. Thus when an individual and its neighbor were selected they interacted, regardless of their similarity, with the individual changing one of its nonmatching elements to be the same as the neighbor's. These trials resulted in unanimity (see Figure 6.3). All 20 trials run to stability in a  $10 \times 10$  population resulted in uniform populations of individuals with identical features.

Interestingly, *the effect of similarity as a causal influence in Axelrod's model is to introduce polarization*: dissimilarity creates boundaries between cultural regions. Interindividual similarities do not facilitate convergence, but rather, when individuals contain no matching features, the probability of interaction is defined as 0.0, and cultural differences become insurmountable. Interaction occurs, and the population converges, in the absence of any similarity criterion, but polarization was not seen; thus the effect of similarity is negative, in that its absence creates the conditions for impassable group boundaries to form.

## Experiment Two: Optimization of an Arbitrary Function

It was hypothesized that the culture model might belong to a larger class of general function optimizers and that in Axelrod's implementation the function that is optimized is, in fact, similarity. Axelrod's placement of similarity as a cause in the simulations makes it essentially an objective function; if a test is passed, then a feature is adopted from a

neighbor, and in the end the population maximizes the criterion. Thus, in the original versions of the model, the change rule is "if ( $\text{rand} < S/N$ ) then interact," where  $\text{rand}$  is a random number between zero and one,  $S$  represents the number of similar or matching elements, and  $N$  is the number of features in a string. The following experiments substitute new terms into the parentheses on the left side of the change rule.

Experiment Two substituted a simple arbitrary function for the similarity test previously used. Rather than testing the similarity of two neighbors, the algorithm was modified so that the numerals comprising an individual's features were summed: this sum was a performance metric. The change rule became "if (the neighbor's sum is larger than the targeted individual's sum) then interact." As before, an interaction comprised the taking of a nonmatching feature from a neighbor. The question was whether the algorithm would maximize the sums of numerals comprising the individuals. 1978

It was seen that a randomly initialized population does indeed converge on the maximum; in 20 trials of the paradigm, the population converged on the global optimum every time. Though the string "99999" was not seen in the initial population, interaction resulted in the adaptive discovery of that optimal set of features and its spread through the population. ACM is capable of optimizing a simple numerical function.

### Experiment Three: A Slightly Harder and More Interesting Function

Minimizing and maximizing an entire string of digits is perhaps the simplest optimization exercise conceivable. A second task was programmed, which is at once more difficult and more interesting as a social science metaphor. The task in the next experiment was to find a set of five numbers—the individual's feature string—within which the sum of the first three numbers equaled the sum of the last two. Thus, the string 34153 would successfully accomplish the objective, since the sum of  $3 + 4 + 1$ , the first three characters in the string, equals the sum of  $5 + 3$ , the last two. In the program, the difference was calculated between the first and second sums, and if the neighbor's difference was smaller (the sums were more nearly equal), the target adopted a feature from the neighbor.

This task is interesting for two reasons. First, unlike the previous example, in which a string of nines or of zeroes satisfies the maximization or minimization constraints, the equal-sums task has a great number of

17769	13967	13967	13967	03764	49094	83193	83193	03434	07979
17769	18137	13967	13967	49094	49094	49094	24381	23454	07979
17567	14537	14537	12737	49094	49094	49094	13453	13453	13453
32739	34539	85599	72999	49094	49094	49094	13453	13453	13866
82789	84598	85599	72999	72081	04581	40581	40581	82459	13866
95169	85168	62127	62127	09081	04581	40141	40141	90191	92193
95187	85168	62127	62127	09081	09081	00110	92193	92193	92193
68178	19239	19239	01230	01230	21223	00110	92193	92193	92193
10955	19239	92579	02727	02727	21223	27283	92193	92193	10955
17769	13509	13509	13509	02754	23124	83193	83193	83139	07979

**Figure 6.4** Result of a simulation where interaction occurred when the difference between the sums of the first three numbers and the last two was smaller for the neighbor than for the individual. Note that in all cases the sum of the first three numbers equals the sum of the last two.

perfect solutions. In the current paradigm, assuming that the algorithm would optimize the problem, it was impossible to predict whether homogeneous regions would develop or individuals would evolve idiosyncratic vectors.

Second, the task requires the complex coordination of the entire vector of elements. A “2” in the fifth position is only successful if it and the fourth element contribute together to a sum that is predicted by the first three elements. In a psychological sense, this is analogous to a model of cognitive or attitudinal consistency. Thus the task supports conceptualization of the paradigm as a model of individuals in a society, each trying to acquire and maintain a cognitive set that meets the requirements of the situation.

As seen in Figure 6.4, all individuals in the population solved the problem, and parts of solutions were distributed through contiguous regions of the population; in 20 trials, this pattern of results was seen every time. Even though the paradigm required coordination of all elements and the interaction operated on individual elements, the method successfully found solutions to this rather difficult problem. It was as if an individual picked up a hairstyle from one friend, a style of jacket from another, slang from another—and made them all fit together. It was, in other words, a depiction of a social process in which no individual

embodied every single aspect of the stereotypical culture, but rather individuals adopted particular aspects of the cultural features prevalent in their area.

These results have an obvious analogue in human society. A string in the simulation may be seen as a set of features, attitudes, or beliefs held by an individual, which must be internally consistent in order to become stable. The features are also constrained to be *externally* consistent; that is, individuals strive to resemble their neighbors, at least when the neighbors are relatively successful at attaining a "good" set of features.

These experiments have taken Axelrod's model of the spread of features through a culture and modified it somewhat. In the ACM paradigm, an individual takes a feature from a randomly selected neighbor if a criterion is met. In Axelrod's writings the criterion is similarity; the present chapter substitutes other criteria and shows that the spread of culture can optimize other functions, resulting, by the way, in similarity among proximal individuals. Similarity, which was a cause in Axelrod's simulations, is now an effect.

### Experiment Four: A Hard Function

A problem is considered intractable if the amount of time required to solve it increases at a faster-than-polynomial rate as the size of the problem increases. For instance, the traveling salesman problem (TSP) requires finding the shortest path through a set of nodes, or cities, without passing through any node twice, and ending up at the starting point. With each additional city the number of possible solutions grows exponentially; with  $N$  nodes there are  $N^N$  possible combinations of nodes and  $N!$  "legal" tours. The TSP is a type of problem called "NP-complete" by computer scientists. The NP problems are intractable, as there is no known deterministic solution to them that reduces the search to polynomial time.

The TSP requires the ordering of discrete elements. As such it is used here to represent analogous cognitive tasks, such as determining the sequence of steps necessary to solve a complex problem, arranging conscious thoughts such that no contradictory beliefs are juxtaposed, and so on. For instance, Thagard and Verbeurgt (1995) have shown that constraint satisfaction in connectionist networks is an NP-complete problem. Rich has suggested that the essence of artificial intelligence (AI) is

CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB  
 CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB  
 CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB  
 CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB  
 CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB  
 CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB  
 CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB  
 CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB  
 HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA CDEFGHAB CDEFGHAB CDEFGHAB HGFEDCBA  
 HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA  
 HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA  
 HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA  
 HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA  
 HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA  
 HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA  
 HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA  
 HGFEDCBA CDEFGHAB HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA HGFEDCBA  
 CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB CDEFGHAB

**Figure 6.5** Result of a simulation of TSP.

in solving these problems: “. . . one way of describing AI is that it is an attempt to solve NP-complete problems in polynomial time” (Rich, 1983, p. 104).

The present experiment implemented an eight-city tour. Thus, there are  $8^8$ , or 16,777,216, possible combinations of eight cities, allowing for cities to be visited more than once, and there are  $8!$ , or 40,320, legitimate tours, in which each city is visited once. Of those tours, there are 16, one starting in each of the eight cities, and going either direction around the tour, that are globally optimal, or provide the shortest possible path. Some heuristic algorithms have been devised to find relatively good tours without testing all possible permutations (for instance, terminate a tour without completing it when it exceeds the shortest distance found so far), but the problem is considered hard by all standards.

The test was set up with a set of cities defined as two-dimensional Cartesian coordinates, which were contrived so that the best tour was

known to the researcher. The problem was set up so that the optimal tour was "ABCDEFGH," of course starting on any letter and going in either direction. A typical result is seen in Figure 6.5. The adaptive culture algorithm is able to optimize combinatorial functions. A penalty was added to the length of a tour if it went to a node more than once. The algorithm was run 20 times with an  $18 \times 8$  population of individuals: the population converged on the globally optimal tour, which had a distance of 7.483 units, on 11 of those trials ( $16/16,777,216 \approx 0.00000095$  probability of finding the optimum by chance). The mean tour was 7.909 units in length; when suboptimal tours dominated, these were seen in all cases to differ from the optimal by at most one element. Either a city was repeated (perhaps a higher penalty would have prevented this) or two neighboring cities were reversed (e.g., BDCE instead of BCDE). Note that this trial resulted in convergence on two globally optimal solutions to the problem.

In a test of the propensity for the algorithm to find multiple optima, the TSP program was run for a second set of 20 trials. Nine of these trials resulted in the globally optimal tour of length 7.483; of these, two trials resulted in convergence on two different optimal patterns, and one trial found five different series of cities that produced the shortest possible route. Other successful trials converged on a single optimum. Thus polarization was sometimes seen to occur in this paradigm.

### Experiment Five: Parallel Constraint Satisfaction

ACM models the spread of features through a community. These can be beliefs and attitudes as well as tangible phenomena. A string of features then could represent a kind of cognitive system, with string symbols representing cognitive elements, using some model of cognitive goodness to optimize the vector.

Connectionist networks often implement binary coding of variables. A set of mutually inconsistent units or nodes are coded with inhibitory connections between each pair. An optimized network contains a value of one for one of the beliefs and zeroes for the others. For instance, Rumelhart, Smolensky, McClelland, and Hinton present an example of a constraint satisfaction network comprising features of rooms. Given some subset of features, such as toaster and refrigerator, the network is able to compute other features of the room that are likely to be present, such as table and sink, while deactivating features such as bed and couch

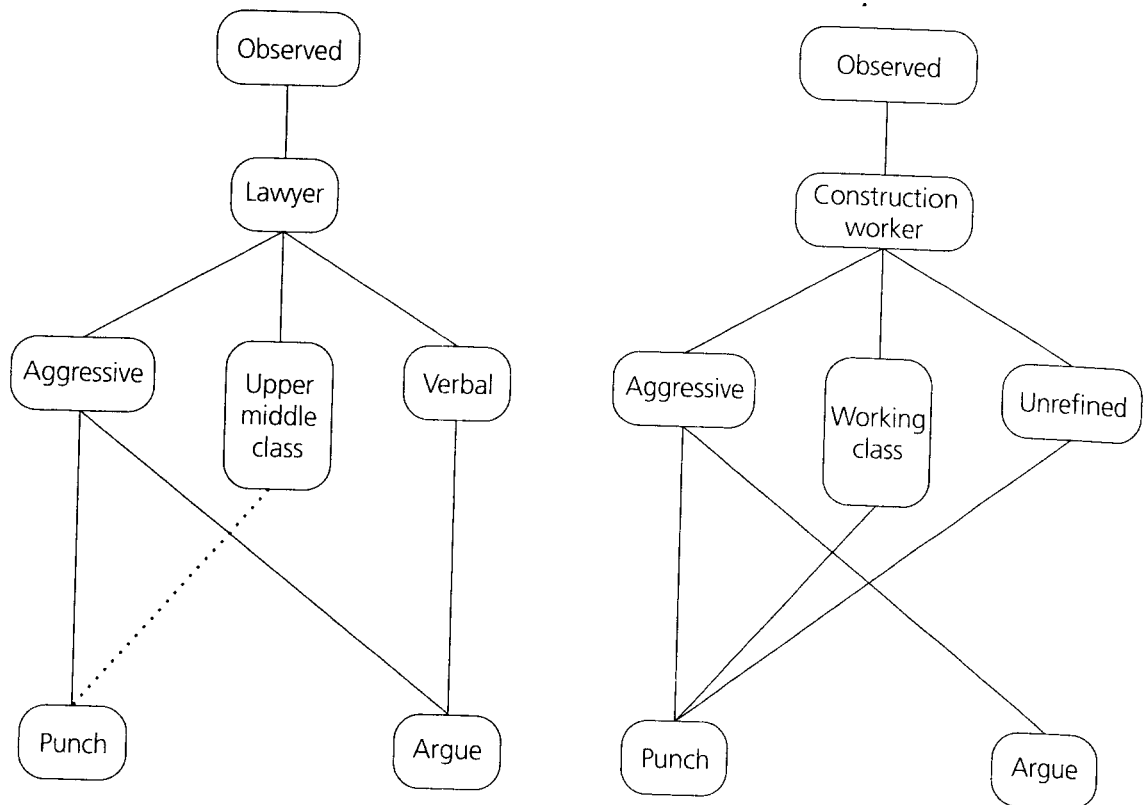
that are likely to be absent. Rumelhart et al. make the point that schemata have variables, which they call "slots." As they write:

In some cases, there are sets of units that are mutually inhibitory so that only one can be active at a time, but any of which could be combined with most other units . . . Perhaps the best example from our current data base is what might be called the size slot. In this case, the *very-large*, *large*, *medium*, *small*, and *very-small* units are all mutually inhibitory . . . (Rumelhart et al., 1986, pp. 33–34).

It does not seem reasonable to suppose that a person who thinks about a large room must actually inhibit thoughts about small, medium, and very large rooms. But because the network is coded in binary terms, it is necessary to code a large number of variables, and an even larger number of connections, to describe a single "slot." A primary reason for implementing binary constraints is the feasibility of optimizing the activation pattern; the Hopfield techniques work on binary nodes. ACM, however, offers a method for optimizing multiple-valued discrete nodes, as well as a sound social-psychological premise for how it could be done in reality. In the present paradigm, the "size slot" can be coded as a single node that can take on six values—the five size classes and a zero for absent or irrelevant. The model is essentially unchanged, except that it is now more comprehensible, realistic, and parsimonious.

For demonstration a parallel constraint satisfaction network was taken from a recent *Psychological Review* paper by Kunda and Thagard. The model simulates the effect of stereotypical information on a concept, in this case the descriptor "aggressive." Kunda and Thagard hypothesized that individuals are more likely to expect a stereotypical construction worker to punch someone and a lawyer to argue with someone, given that both targets are labeled "aggressive." In their paper this was demonstrated using two networks, as shown in Figure 6.6. The first network shows the constraints resulting from factors associated with a lawyer, that is, the target is stereotypically expected to be Upper middle class and Verbal, and if labeled "Aggressive," is expected to Argue rather than Punch. This effect results largely from the negative connection between Upper middle class and Punch. The second network represents the stereotype of a construction worker, i.e., Working class, Unrefined, and more likely to Punch than to Argue.

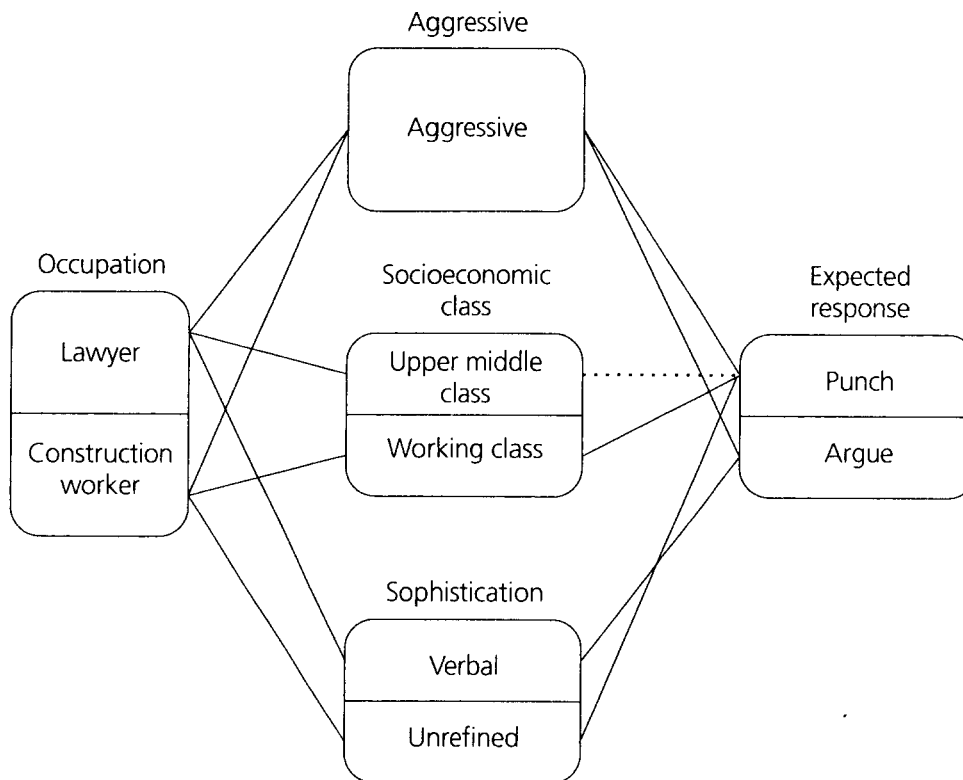
One advantage of using multivalued nodes is that comparisons can be implemented in a single network. In the example, both of Kunda and Thagard's networks have identical binary nodes representing Aggressive,



**Figure 6.6** Parallel constraint satisfaction examples. Solid lines represent positive links, and the dotted line represents a negative link. (From Kunda and Thagard, 1996.)

Punch, and Argue; they differ in the other three nodes. It is clear, however, that Lawyer and Construction worker are levels of the same variable, as are Upper middle class and Working class, and Verbal and Unrefined are implicitly exclusive of one another. Further, as the question is whether Punch or Verbal satisfies the constraints better, it is possible to code them as levels of a variable called "Response." Thus, the two networks can be conceptually collapsed into one, as seen in Figure 6.7, with nodes labeled "Occupation," "Aggressive," "Socioeconomic class," "Sophistication" (for want of a better term—as usual, all labels are arbitrary), and "Expected response."

An experimental trial clamped, that is, held constant, a value of Occupation: a population was initialized with random values for the nodes and with one occupation fixed, for instance, Occupation clamped to the value "Lawyer." Individuals in the population generated patterns of activation, evaluated these, compared their own evaluations to their neighbors', and adopted a feature from the neighbor if their pattern was better.



**Figure 6.7** The two binary networks shown in Figure 6.6 are combined into a single multivalued network. Note that each node can also take on a zero value.

A string is composed of the variables in the following order:

1. Occupation (A = Lawyer, B = Construction worker)
2. Aggressive (value is A or 0)
3. Socioeconomic class (A = Upper middle class, B = Working class)
4. Sophistication (A = Verbal, B = Unrefined)
5. Expected response (A = Punch, B = Argue)

Thus a string such as "AABBA" represents a Lawyer who is Aggressive, Working class, Unrefined, and more likely to Punch than to Argue. In cases where a node is not connected, or the total inputs to all levels of it are zero or less, the node takes on the value of zero, indicating it is irrelevant, nonsalient, or unnecessary.

The network was coded as a series of rules of five elements. Negative connections were coded with a minus sign on the node value or on the weight, and a period represented no connection. Note that coding such a system in a two-dimensional matrix would remove all advantage and make it identical to a binary network, and thus a series of rule statements was employed.

For instance, the following was the rule set for Figure 6.7, with "Construction worker" clamped on:

B	.	.	.	.	
A	A	A	A	.	
B	A	B	B	.	
.	A	.	.	#	
.	.	A	.	-A	
.	.	B	.	A	
.	.	.	A	B	
.	.	.	B	A	

The first rule is special; it determines values to be clamped on throughout the trial. In this set the first line tells the system to assign a "B" for the first node; the dots in that row mean the other nodes are free to vary. Rules are designed so that the first column with a symbol in it represents the focal node, and other columns are nodes connected to it. The second line says that state "A" of the first node is connected to "A" for the second, third, and fourth nodes and is not connected to the fifth. The "#" symbol in the fourth rule means that the second node (the second column is the first with a symbol in it) connects to *any* value of the fifth node. A minus sign means that the element should not be present; it represents a negative or inhibitory connection.

The rule set may be accompanied by a set of weights for each element in each row, or a default weight of 1 can be used for everything. The program compares a proposed solution, for instance, "BABAB"—meaning that a Construction worker is Aggressive, Working class, Verbal, and likely to Argue—to each rule, and sums up the weights of the items that match. In the present example, the first "B" is clamped on, so all population members will have that state for their first node. The rule in the second line (A A A A .) fails to match in the first specified position, so the rest of the rule is not evaluated. The next rule (B A B B .) does match in the first specified position, so the weights for the rest of the line are summed, with a penalty for mismatches. Assuming all weights = 1.0, the sum is 1 for the matching "A" in the second position, plus 1 for the matching "B"

in the third position, minus a penalty for the mismatch between the “A” in the problem solution and the “B” in the rule, minus another penalty for the final “B” in the solution, which corresponds with a “.” or no connection in the rule.

The next rule says that a state of “A” for the second node corresponds with anything in the fifth node. Since the proposed solution does meet the criterion, the sum is added for a match in this row. Thus the program goes through, comparing the symbol string to the rules. A measure of goodness is defined by the sum of the weights of all constraints that are satisfied by a pattern of values:

$$G = \sum_r \sum_n W_m$$

where  $r$  is rules,  $n$  is nodes, and  $W_m$  is the weight of a relation or connection that matches, that is, is identical in the test and the rule strings.

Note that some theorists code binary node activations in  $\{0, 1\}$  and some in  $\{-1, +1\}$ . The difference between these two implementations is seen in the effect of a node when it is turned off; a node with zero value has no effect on the units it is connected to, while a node value of  $-1$  actively tends to inhibit them. This results in some paradoxical cases, where in order to achieve the desired result positive relationships must actually be coded using negative connections. The present model emulates connectionist networks with  $\{0, 1\}$  activations and as such avoids that anomalous instance.

When the program is executed, a population of random symbol strings is generated. A string is evaluated by comparing it to each rule in the rule file. Weighted matches between the test string and the rules are summed through the rule set. A larger total indicates that more constraints were satisfied. The ACM algorithm is applied by selecting an individual and a neighbor, comparing their evaluation totals, and interacting when the neighbor's total is greater than the individual's.

This network model using a  $10 \times 10$  population was tested 20 times with Lawyer clamped on and 20 times with Construction worker clamped on. All 40 trials resulted in the population converging on the correct stereotypical conclusion, that is, Lawyers would be thought more likely to Argue than to Punch, and Construction workers would be expected to be more likely to Punch than to Argue.

What does this approach to constraint satisfaction buy us? First, the use of multivalued nodes is a more efficient and theoretically sound way to encode slots, or sets of mutually exclusive variables. Second, the

population dynamics allow important insights into the cognitive operations involved. Part of the definition, indeed, of a stereotype is that it is a belief shared by a group about members of another group. ACM shows the development of stereotyped thinking as it spreads through a population. A set of commonly held beliefs is arranged in various ways until the best explanations are found. The search is shared by the population, and the successful results spread to all members.

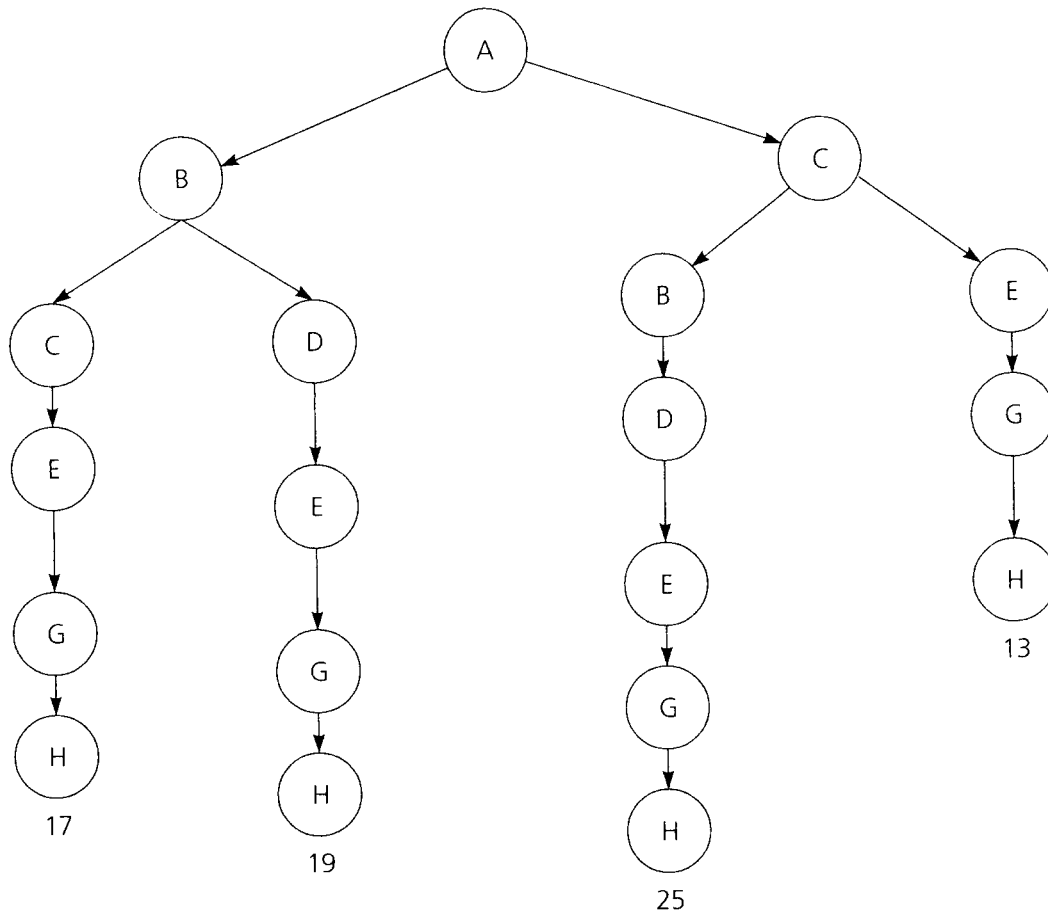
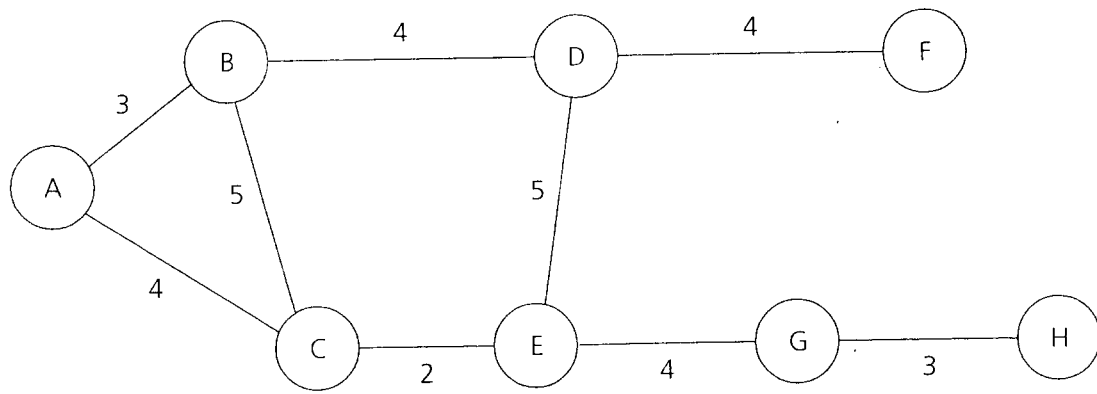
## Experiment Six: Symbol Processing

Many kinds of problems in conventional AI are conceptualized as a search for the shortest path through some symbolic nodes. The example described here is taken from Chapter 4 of Patrick Henry Winston's classic textbook, *Artificial Intelligence* (Winston, 1992). Winston gives the example in terms of finding the shortest path from city A to city H. Symbolically this kind of problem is the same as rebuilding a carburetor, cooking an omelet, preparing for an exam—a sequence of steps must be followed, they must be in correct order, and no steps must be skipped. Many decision trees have a number of alternative successful routes, some of which are superior to others, in which case the investigator decides whether to accept a good-enough solution or to hold out for the global optimum.

Figure 6.8 depicts a network of paths, which can be transformed into a set of hierarchical paths that can be searched without looping back to a previous point. The usual discussion of this kind of problem describes strategies that can be used to reduce unproductive searches. The culture model completely ignores these rational approaches. As usual, a population of individuals is initialized, with each individual comprising a string of symbols, for instance, ADCHBEAF, in an eight-node problem (the first node will always be A).

As before, each string is evaluated, that is, the distance from the initial A to the first occurrence of the goal H is calculated. Since only a small proportion of nodes are legitimately connected to one another, a penalty is added to the length of the path for illegal connections. A neighbor is selected, the two are compared, and a feature of the neighbor is taken if the neighbor's trip is shorter.

Because trips through the graph vary in the number of nodes they pass through to reach the goal, an adjustment in the performance measure was necessary. If simple tour length were used as a performance



**Figure 6.8** The shortest path problem. A network of nodes is transformed into a hierarchical tree structure. (Adapted from Winston, 1992.)

measure, the system would converge quickly on a shorter, more probable path, even though it is suboptimal, simply because it will be found sooner. In order to equalize the chances of finding longer, less probable, sequences of graph nodes, total distance was weighted by

$$\left(\frac{N-1}{N}\right)^{L-2} \times \left(\frac{1}{N}\right)$$

where  $N$  is the number of nodes in the graph and  $L$  is the ordinal position of the goal, for comparisons of individuals and their neighbors. The first term of this weight represents the probability that elements preceding the goal in the series will not be the goal value, and the second term is the probability that the final element will be the goal value.

For example, in the example given above, the best possible tour is ACEGH, with a distance of 13 units and the target in the fifth ordinal position of the string. The penalty for invalid sequences, that is, for going from one node to another when there is no connection between them, can be set by the user; let us assume it has been set equal to 10 units of distance added to the total for each violation. An individual then that tried the invalid tour AH would have a distance of 10 units—better than the global optimum! Thus cheating would be rewarded if it was not prevented. Weighting by the formula above, though, transforms the optimal tour into

$$13 \times \left(\frac{7}{8}\right)^{5-2} \times \frac{1}{8}$$

which evaluates to approximately 1.088, while the “cheater” solution becomes

$$10 \times \left(\frac{7}{8}\right)^{2-2} \times \frac{1}{8}$$

which is 1.25 transformed units in distance. Thus the incentive to take shortcuts is reduced. This approach enables operations on solution strings of various lengths.

In 20 trials of the Winston decision tree a  $12 \times 8$  population converged on the global optimum 14 times. While it is not argued here that this trial-and-error algorithm is more efficient than traditional search procedures, it is suggested that in fact many real problems requiring the sequencing of elements are solved socially through collaboration,

with each individual contributing and recombining pieces to result in improvement.

### Discussion

An ACM system iterates, with individuals repeatedly interacting, until it reaches a stable point where change ceases. The simulations reported in this chapter have been seen to stop in three kinds of states: states where the population is uniform, with all feature strings identical; states with two or more regions of strings that are identical within regions and different between them; and states with a population of unique strings that satisfy the change criterion equally well.

The question of what causes groups to diverge is of course profound, important, and difficult. In Axelrod's simulations change stopped when the similarity criterion was not met anywhere in the matrix; the result was a population with one or more homogeneous regions. As was seen in the first experiment above, when the similarity criterion was deleted from the paradigm, the system evolved until no individual had a neighbor with a nonmatching feature—the entire population became homogeneous. Thus it was concluded that the similarity rule was responsible for producing boundaries between distinct groups.

In the other experiments reported here, the change rule stated that interaction occurred only when the neighbor's performance exceeded the selected individual's. In one experiment (Experiment Three), very many global optima existed, and individual feature strings tended to be correlated but unique. In others, for example, the maximization and stereotype experiments, a single global optimum existed, and the population converged on it. When a limited number of optima existed, the population would converge on as many as were discovered; when the function was hard and global optima few in number (e.g., the TSP), only one solution was found in most cases, but in some trials multiple best solutions were discovered by the population. Where two or more global optima were more easily found, regions of the population tended more frequently to converge on separate patterns. It is interesting to note the formation of "multiethnic" populations out of benign conditions, that is, individuals converged on different combinations of features; group divergence occurred even in the absence of self-categorization, as in Henri Tajfel's theory, or competition for resources, as theorized by Donald Campbell and others.

In sum, culture and cognition are seen from three simultaneous levels of phenomena. First, individuals searching for solutions learn from the experiences of others. The "problems" addressed by the present theoretical viewpoint may be epistemological, logical, ethical, aesthetic, or metaphysical; they may be emotional, political, physical, or sexual. At the social learning level of culture and cognition, individuals learn from their neighbors. This is the level that is most easily measured by social scientists, and importantly it is the level at which the system is programmed.

Second, an observer looking at a population as a whole perceives phenomena of which individual people are the parts. Opinions, beliefs, and behaviors correlate with geographical regions, as well as with ethnic, religious, socioeconomic, and other cultural dimensions. People who interact frequently become similar to one another in many ways. Nothing in the computer programs specified that regions should become homogeneous or that borders should form between groups; these global effects emerged from simple local interactions.

Third, culture affects the performance of the individuals who comprise it. Individuals gain benefit by imitating their neighbors, good patterns spread through the population, and the best patterns spread most widely through the group, so that individuals benefit from improvements that might have been discovered in quite remote regions. Further, as patterns move through a population and become integrated with existing patterns of features, the probability of even greater improvement increases. Again, nothing in the programs specified that individuals would solve problems, but only that they would imitate others who performed better than themselves.

This latter global phenomenon is largely invisible to participants in the system, for two reasons. First, an individual does not necessarily realize where an idea came from, if it originated beyond the horizon. Second, individuals are incapable of seeing the effect of culture because they are it: there is no background against which the figure can be seen, as the perceiver is an element in the perceived field.

This model appears to give individuals very little credit. Thinking, and in fact *hard* thinking, is depicted here with no assertions about, or reliance on, the intelligence of individuals. A human processing unit in these simulations functions mainly through adaptive imitation. Obviously an individual human processes a great deal of information. The present view would suggest that a relatively large proportion of cognition is concerned with evaluation and comparison of self and others. The strings of symbols processed in the current examples are highly oversimplified tokens of the multidimensional, dynamic arrays that are

processed by human societies—and these experiments suggest that *societies* process information.

### Summary

The ACM is able to find some combinatorial optima, but is not designed for that purpose—it was really devised only to show that a simple principle, which could be called “imitation of your betters,” is able to find its way through a complex search space. The following chapters present and elaborate a paradigm that capitalizes on the ability of social interaction to result in optimization of hard problems.



