SESSION I

PRESIDENTIAL ADDRESS

Connectionist and rule-based representations of expert knowledge

EARL HUNT

University of Washington, Seattle, Washington

Research in computer science has led to the development of two broad classes of models of knowledge representation: rule-based and connectionist systems. Both techniques solve the same abstract problem, that is, the assignment of cases to classes. Connectionist modeling techniques have been applied to three classification situations in which one would expect rule-based models to be applicable. Two of the situations involved the diagnosis of problems in and operation of a fictitious power plant. The third situation involved the classification of misconceptions held by elementary physics students. Connectionist modeling developed adequate simulations of behavior in all three cases.

Homo Sapiens rely on knowledge to survive. How is knowledge represented in our minds? This is a fundamental question for psychology. A similar question has been asked in computer science, since useful computer programs, no less than useful humans, manipulate internal models of the external world. Much of modern cognitive psychology rests on the assumption that computer models of thinking do indeed provide useful analogies for human thought. However, there are a lot of computer programs. What sorts of programs are reasonable models of human cognition?

In a speech that heralded the development of computer-based *expert systems*, Edward Feigenbaum (1977) described several programs that used knowledge bases to solve nontrivial problems in chemistry and biology. These programs utilized a *rule-based* approach to knowledge representation. They manipulated explicitly coded rules for governing behavior. For instance, the rule-based approach to driving requires an explicit rule that says, "If the light is red, then stop," and another that says, "If a pedestrian is crossing in front of you, stop." (These rules apply for areas in which civilized practices are observed. Such locations can be found in rural sections of Britain and North America.) The idea behind rule-based

representations is that expert reasoning is determined by the interaction between rules and a problem description. Most of the highly publicized effort to develop practical expert system programs can be thought of as an elaboration of the ideas in Feigenbaum's paper. These ideas are also reflected in many of our theories of how human expertise operates.

A quite different approach to knowledge utilization has been taken in the development of *connectionist systems*. These are programs that acquire knowledge from experience, in a manner that is at least not incompatible with plausible neurological models of learning (Rumelhart & McClelland, 1986). In the connectionist approach, rules are never explicitly written into a program or a data base. Rather, these programs record associations between external problem descriptors and internal "nodes" that accumulate knowledge about statistical co-occurrences of features and data. An internal node is activated by the presence of certain patterns of features. The system also observes the result that occurs when a particular response is made, based upon the activation of one of the internal nodes. This information is used to change the weighting system until an optimal stimulus-response mapping is achieved.

Internal nodes are not equivalent to rules, because the same descriptors may be associated with many different internal nodes, and a single internal node may be associated, in different ways, with many different responses. The response made in a given situation will be determined by the total pattern of activation of nodes by descriptors, and responses by nodes, rather than by the activation of a single explicit rule. Knowledge is distributed across a

This paper is based on the Presidential Address given to the Society for Computers in Psychology, at the Society's 1988 Annual Meeting in Chicago, IL. The research supported here relevant to instruction in physics was supported by a grant from the McDonnell Foundation to the Mercer Island School District and to the University of Washington. Address correspondence to Earl Hunt, Department of Psychology, University of Washington, NI-25, Seattle, WA 98195.

network of descriptors and nodes, rather than being encapsulated in understandable rules. Because of the distributed nature of knowledge, and because the descriptor-node-response associations are computed (conceptually) in parallel, connectionist systems are sometimes referred to as *parallel distributed processing* (PDP) systems.

Models of human knowledge use have been developed using both the rule-based and the connectionist approach. A commitment to one approach or the other carries with it a commitment to conduct fundamentally different sorts of data analysis. When a model builder constructs a rule-based system, the principle source of data is an expert's description of his or her own behavior. There is no question at all but that this is a rich source of data. Who should teach the machine but the expert practitioner? In fact, people who are interested in commercial applications of artificial intelligence (AI) technology have developed a semi-profession called *knowledge engineering*. Knowledge engineers are supposed to translate the verbal statements of doctors, lawyers, or other experts into appropriately coded rules for a computer program. However, it is not at all clear to what extent practitioners of this art are guided by our knowledge of the pros and cons of analyzing verbal protocols.

The connectionist approach to model building is quite different. The connectionist focuses on overt behavior; what subject-matter experts do rather than what they say they do. In connectionist modeling, a "suitably coded" representation of a situation is presented to a learning system. System output, again suitably coded, is compared to a similar coding of a human's response to the same situation. The pattern of associations between nodes is adjusted until human and system responses agree. This certainly seems more objective than the analysis of verbal protocols, at least to those who value B. F. Skinner's view of psychology. However, there are problems. In many situations, connectionist learning algorithms do not work, or work very poorly, unless the associations that the system can generate are restricted in some way. Study after study has shown that the choice of an appropriate suitable code for representing stimuli and responses is crucial. Just how are configuration restrictions and coding schemes to be chosen? This is at least as much of a black art as is knowledge engineering.

There has been considerable debate over whether rule-based and connectionist systems are fundamentally different, or just alternative ways of implementing computationally equivalent models (Rumelhart & McClelland, 1986; Fodor & Pylyshyn, 1988). Whatever the outcome of the debate, it is certainly true that, on the surface, one proceeds in a very different way when trying to implement a rule-based or a connectionist program. In the past, the choice of approach appears to have been dictated largely by the biases of the theorist. People have clearly wanted to develop rule-based or connectionist programs, and have chosen problems to match. However, it is possible, and possibly more useful, to look at the situation from the point of view of a person interested in finding

an answer for a question, rather than from the point of view of a person interested in finding a question for an answer. What are the circumstances that would dictate the use of a rule-based or a connectionist approach?

My colleagues and I have conducted some exploratory studies comparing the utility of rule-based and connectionist approaches to modeling the way people use knowledge in different situations. Our purpose was to see if we could find situations that favored one or the other approach. However, the results were somewhat surprising. To see why, let us consider each study in turn.

A STUDY OF DIAGNOSTIC REASONING

The first study to be considered is a study of diagnostic reasoning. In a diagnosis problem, a person observes a case, represented by a set of symptoms. The task is to assign the case to a pre-defined set, that is, to make a diagnosis. (Formally, the set membership of an instance is inferred from its description.) Diagnostic reasoning has been extensively explored, both as an experimental task in psychological research and as a practical problem-solving task.

In his PhD dissertation, James Lundell (1988) studied an artificial diagnostic situation that is at once more complex than those typically studied in the experimental laboratory, and more controlled than real-world diagnostic problem solving. Lundell asked people to imagine that they were learning to troubleshoot an imaginary, but realistic, internal combustion power plant. He then presented readings of instruments that would be available to the operator of an actual power plant, for example, coolant temperature and fuel consumption. The task was to diagnose a malfunction in the system. Figure 1 shows a typical display and the diagnoses that might be associated with it.

University students (who were not mechanics) observed 500 displays each. They were asked to provide a diagnosis

Diagnosis

1 - radiator	3 - generator
2 - air filter	4 - gasket

Figure 1. A typical display shown in the diagnostic task study.

sis for each display, and were then told the correct diagnosis. As would be expected, performance improved dramatically. The students moved from 25% correct (chance) diagnoses at the outset to 75% correct diagnoses at the end of training. Clearly, the students learned something about the display task. How should it be represented?

Lundell first developed a rule-based system for representing student knowledge. This was done by analyzing the results of a structured interview about the diagnosis task. Following the learning phase, the students were asked to infer a diagnosis from a single dial reading. They were also asked to specify typical configurations for particular diagnoses. Examples of the questions are shown in Figure 2. Finally, the students were asked to specify one or more prototypical examples of each diagnostic category. This information was used to develop rule-based expert systems tailored to the (asserted) knowledge of each student. In addition to asking knowledge-acquisition questions during the knowledge-acquisition stage, Lundell asked students to classify statistically defined typical cases for each diagnosis. The examples used will be referred to as the *first test sequence*. No feedback was given for responses to the first test sequence.

Page from the direct rule questionnaire

An item from the Decomposed Rule Questionnaire

Figure 2. Displays used to elicit rules in the diagnostic task study. The direct rule questionnaire was used to allow participants to state configurations they thought were typical of a diagnosis. The decomposed rule questionnaire was used to determine the diagnoses suggested to participants by a single item of information.

Figure 3. The architecture of the connectionist system used in the diagnostic task study. Dial readings constituted the outermost units. They fed activation forward to the hidden units. The hidden units then activated response units. The response units corresponded to diagnoses, as in the two cases shown in the figure. Corrective feedback was always provided.

After the knowledge-elicitation phase, the students were tested on new cases, the *second test sequence*. They were again correct 75% of the time. The new test cases were also diagnosed by the expert system programs Lundell had developed. The programs produced the correct diagnosis for 55% of the test cases; well above chance but substantially below human performance. Perhaps even more telling, the system extracted from a given student was no more predictive of that student's performance than were systems extracted from the responses of other students. While the rule-based expert systems were not irrelevant to diagnosis, they were suspect as models of psychological performance.

Lundell also constructed connectionist models for each student. The models were constructed by using a feed-forward connectionist model with a single layer of hidden units intervening between the stimulus and response units. Figure 3 shows the configuration of the network. Weights were established by "training" the network on the first test sequence, augmented by the prototype items for each diagnosis, as generated by each student. The programs were trained by combining error correction with the back propagation algorithm for weight modification (Rumelhart & McClelland, 1986). Program training was continued until the programs agreed with the responses of the student whose knowledge was being modeled.

The connectionist systems were then used to classify the items in the second test sequence. Seventy-two percent of the test patterns were correctly classified, averaged over simulations. This compares to 75% correct classifications, averaged over students. Furthermore, on a system-by-system basis, there was a higher correlation between the responses made by a system and the responses made by the student from whose data the system had been generated, than between system responses and the responses of other students. The correlations were generally in the .4 range, which is not as high as one might wish, but there was no question of statistical significance.

Quite clearly, the connectionist models more accurately portrayed student behavior than did the rule-based models. This was true in spite of the fact that the rule-based systems contained the rules that the students themselves had generated. The method of interviewing (i.e., the use of the forms shown in Figure 2) seems a reasonable assurance that the failure of the rule-based system was not due to the knowledge engineer's (Lundell's) ability to understand what the students were saying. Furthermore, the rules themselves were quite reasonable guides for diagnostic performance. If Lundell's participants had been real experts, responsible for the maintenance of a real power plant, most professional knowledge engineers would have been able to generate a rule-based system from their responses. Knowledge engineering did not produce a satisfactory model, but connectionist training did.

A SYSTEM-CONTROL STUDY

A second study, by Merrill McSpadden (1989), has extended Lundell's paradigm to a *system-control* situation. The difference between system-control and diagnosis is in the form of feedback provided to the learner. In diagnostic studies, the learner is told, explicitly, what the classification response should have been. For instance, on each trial in the training phase of Lundell's study, the participants were told what their response should have been. In a system-control study, the learner knows what he or she did, and what happened, but does not know what the optimal response would have been. Intuitively, the system-control situation seems to be at least as common as the diagnostic situation.

The basic design of McSpadden's study was quite similar to Lundell's. University students learned to play a computer game, in which they had to manage the power of a simulated nuclear power plant.¹ The general design of the plant was explained to them prior to the beginning of the game. The game itself consists of a series of "episodes," consisting of 30 or fewer "trials." A trial is supposed to represent the operation of the plant for a single day. Participants are shown five variables describing the system state (e.g., power output, primary coolant level). The participant then sets the system-control variables (e.g., primary coolant flow). Table 1 lists the system and control variables displayed. The program then calculates the values of the system state variables for the next trial, based on equations that are not available to the participant, but are not inconsistent with the information about the plant that the participant has been given. An episode consists of a sequence of 30 trials, unless the system state variables enter a catastrophic condition, such as a simulated reactor meltdown. The goal of the game is to maximize power output during an episode without entering a catastrophic state.

University student participants played the game for a total of 10 h, over a 5-day period. At the beginning, no one was able to complete an episode without creating a

Table 1
System State Variables and Control Variables
Manipulated in the System-Control Task

System State Variables	
1. Average power generated	5. Power generated this day
2. Reactor temperature	6. Emergency fluid level
3. Exchange unit temperature	7. Primary fluid level
4. Cooling tower temperature	8. Secondary fluid level
System-Control Variables	
1. Fuel level	3. Primary fluid flow
2. Emergency fluid flow	4. Secondary fluid flow

Note—The task is to maximize average power generated per day without entering a catastrophic state. The participant sets the system-control variables in order to do this.

catastrophe. By the final session, participants routinely completed episodes, although many of them did not maximize power output. Participants had access to written diagrams explaining plant operation throughout the experiment. At regular intervals, half of the participants were asked to explain the rules they were using. The other half of the participants were never queried. This manipulation was crossed with testing. Half of the students were periodically tested for their knowledge of the likely results of their actions, and half were not.

The testing and explanation manipulations were intended to encourage careful development of efficient rules. However, the manipulations made absolutely no difference to behavior. People learned to perform the task just as quickly in the condition when they just worked on the problem as when they were asked to explain their rules, or when they were tested during learning. We concluded from this that the explicit statement of verbal rules had rather little to do with learning how to control the system. This conclusion was reinforced by our second observation. Behavior could be modeled by using a connectionist learning procedure that did not contain any representation of rules.

McSpadden developed a variation of the exemplar-learning model proposed by Hintzman (1986),² specialized to apply to his task. McSpadden's and Hintzman's models are based on the idea that long-term memory contains a coded representation of all cases that have been experienced. In the situation considered here, a case consists of the system variables at the outset of a trial, the response variable settings that were used on that trial, and the system variable settings that resulted. The following steps were used to convert the memory model to a model of learning in the system-control situation.³

1. On each trial a *probe* was constructed. The probe consisted of the current state of the system variables.

2. The probe was compared to all items in memory. Using the methods described by Hintzman (1986), an *echo* was constructed by calculating a weighted average of all vectors in memory, weighted by their closeness to the probe, considering only those variables in the probe. The probe would then contain control variables and a predicted "result," that is, the variables in the probe correspond-

ing to the next state of the system variables. The predicted response was simply the value of the control variables in the echo.

These steps led to a crude form of problem solving, in which the model would try to do whatever it did in previously encountered situations similar to the current one. In theory, such a strategy might simply have replicated disasters. In practice, disaster replication was unlikely to occur, because responses that led to disaster states rapidly terminated the episode. Therefore, they had few representations in memory. By contrast, as soon as the learner tried out something that worked, successful responding prolonged the episode, and successful system-control (or, if you prefer, stimulus-response) memories began to populate the memory.

In spite of its simplicity, the model made surprisingly accurate predictions about behavior. Figure 4 shows the settings that two people produced for the variable *fuel level* during an episode after considerable learning had taken place. These responses are compared to the range of settings predicted by the model extracted from each individual's data.⁴ Although some people's responses were clearly more accurately modeled than were others', there was a substantial amount of agreement overall. In particular, when unusual situations were encountered, both the person and the model began to change control variables widely. People tended to make somewhat more extreme changes than did the model. However, both people and the model agreed about when radical changes in settings were needed, and about the direction of those settings.

The machine diagnosis and system-control experiments are examples of situations in which people ought to follow rules, but we seem unable to state the rules that they are following. We look next at a situation in which one of the criteria for successful analysis is the ability to state a rule. Would connectionist modeling have anything to do with this sort of problem solving?

THE CLASSIFICATION OF PHYSICS STUDENTS

When students enroll in high school physics classes, their minds are not *tabulae rasae* with respect to the physical world. Everyone has experience with things that bump other things, and everyone knows that things fall down rather than up. The theory that accounts for these facts, Newtonian mechanics, is a quite sophisticated explanation that appeals to concepts rather outside our common sense. If this were not so, someone would have thought of Newtonian mechanics well before Newton's time.

One way to think of naive theories of physics is to regard them as rules of thumb that are useful in some situations, but that are not coherently related to each other. For instance, a person might believe that when objects are in contact, (1) larger objects exert more force than smaller objects, and (2) only moving objects exert force. These rules explain how a massive linebacker "forces" a quarterback to the ground in football, and why a baseball can "force" its way through a window. However,

SIMULATION PREDICTIONS VERSUS SUBJECT PERFORMANCE FOR FUEL LEVEL

SIMULATION PREDICTIONS VERSUS SUBJECT PERFORMANCE FOR FUEL LEVEL

Figure 4. Comparison of participant actions to predicted performance in the system-control task. The two panels depict data from two different participants. The solid line indicates the participant's setting across trials. The dashed lines indicate the maximum and minimum settings predicted by the connectionist-expert system model.

the rules make conflicting predictions when a baseball is thrown against the side of a building.

James Minstrell, an experienced high school physics instructor, is developing a procedure for teaching physics that uses knowledge about student rules to guide instruction. The idea is that an instructor should identify in-

consistent rules and construct demonstrations that accentuate their conflict. The example in the preceding paragraph, of a baseball being thrown against the side of a building, is an illustration of such a conflict situation. The instructor then leads the student toward a general (Newtonian) rule that resolves the conflict. Obviously, Minstrell's approach relies on being able to identify the misconceptions that an individual student has. In collaboration with Minstrell, Colene McKee and I have been working on this problem.

We first needed to identify plausible rules of thumb. Attendees at a conference on physics instruction were asked to describe misconceptions they had observed while teaching Newtonian mechanics. The attendees included experienced physics instructors and cognitive scientists who had done extensive research on the topic.⁵ Following the conference, Minstrell reviewed the suggestions the attendees made and compiled a final list of rules. Henceforth, each rule will be called a *facet*. For instance, one quite common facet is the misconception "Larger objects exert more force than smaller objects." Thirteen facets were generated, including a "physicist" facet equivalent to Newton's laws of force.

Facets were operationalized by reference to the University of Massachusetts Diagnostic Test (UMDT). This test consists of a set of questions about the distributions of forces. For example, Question 1 is,

Jim buys a new floor lamp and leaves it standing in the corner of his room. Which of the following do you think is true?

1. The floor exerts an upward force on the lamp.
2. The floor does not exert an upward force on the lamp.

One of the facets the educators generated was a belief that objects above exert force on objects below, but not vice versa. This facet predicts that on Question 1, Answer 2 will be chosen. The physicist facet, of course, predicts that Answer 1 will be chosen.

More generally, each of the facets could be defined in terms of the answer expected to the question on the UMDT. Note, though, that no one facet except physicist was relevant to every question. Various combinations of facets could be used to generate answers to each question on the UMDT. However, a student could produce a pattern of answers that could not be completely accounted for by any combination of the facets. In fact, though, such patterns were relatively rare. In one class that we examined, 75% of the students' responses to the UMDT could be explained as examples of the operation of one or more facet. Furthermore, most of the unexplained responses occurred on only 4 of the 18 questions on the test. It is clear that the physics instructors' rule-based explanations for student behavior could be used to classify students.

The physics instructors claimed that they generated their explanations by considering systematic distortions of the laws of physics. But is this what they really did? To answer this question, we developed a connectionist model for grouping students into clusters. Would this model, which did not contain any knowledge of physics, iden-

tify clusters of students that held particular misconceptions in common?

The model used was based on McClelland and Rumelhart's (1988, p. 188) competitive learning model for cluster development.⁶ This model was chosen because it is closely related to the exemplar-based learning model that had proven useful in the system-control study. Each student's answer pattern was treated as a vector of binary features, that is, the features of having marked particular answers on the answer sheet. The PDP model organized itself so that each vector would be mapped to one of k hidden units. Every time a vector was successfully mapped to a hidden unit, a learning algorithm was applied. This had the effect of increasing the correlation between the vector of features and the vector of weights associated with the relevant hidden unit. The vectors were presented and classified several times. The classification process was halted when the same classifications were repeated. (At this point the weight vectors stabilized.) In a geometric interpretation of the clustering procedure, answer patterns were grouped into clusters of points on the unit hypersphere. Each cluster was defined by its associated weight vector.

Adequate statistical techniques for evaluating the reliability of a clustering solution do not exist. Therefore, the clustering procedure was replicated on two independent groups, consisting of 75 and 70 student answer sheets, respectively. Very similar results were obtained from each sample. In each case, there was a single large cluster, comprising about two-thirds of the students, a smaller cluster containing about one-fifth of the sample points, and a few outliers. A second stage analysis in terms of facets, described below, produced essentially the same results for each sample. Therefore, the results will be discussed in terms of the group as a whole.

Was there a relation between the clustering procedure and the ways in which students conceived (or misconceived) the physical world? In order to answer this question, two analyses were carried out. The weight vectors associated with the hidden unit defining each of the clusters were correlated with the patterns of answers expected by each facet. Only the answers relevant to a particular facet were considered in computing the correlation. None of the correlations were as high as .5. The majority were in the .2-.3 range, and there did not seem to be any clear difference in the pattern of correlations between weights and facets, across clusters. This suggests that the PDP clustering did not capture the students' reliance on misconceptions.⁷ However, this may not be the correct comparison. The weights used for the internal decision making of a PDP model are constrained to produce a nonlinear transformation of any statistical regularities that may exist in the stimuli presented to the system. If one is interested in those regularities, the stimulus vectors should be examined directly.

The centroids of the vectors of answers within each of the two clusters provide a measure of the frequency with which different patterns of answers were chosen within a cluster. Each centroid was compared to the vector of

answers expected if a student answered exactly in accordance with a particular misconception, as had been done for the weight vectors. There was considerably more variation in the pattern of correlations than there had been for the correlations involving the weight vector. Correlations ranged from the .01 range to as high as .85. The patterns of correlations showed some clear differences across clusters. This led to the following conclusions.

The naive conceptions of the high school students bore little resemblance to Newtonian ideas of force. Only about 25% of the answers within the two major clusters corresponded to those in the physicist vector. There were some outliers, not included in the clusters, who did answer as a physicist would. It should be remembered that the students were examined at the start of the course, and had not yet received formal instruction on Newtonian concepts. All that this proves is that the Newtonian model is not obvious to the naive observer. Newton really was a genius.

The misconceptions themselves were not used consistently by students within a cluster. If they had been, the correlation involving a facet and at least one cluster would have been one, and this never happened. The two most frequent misconceptions, apparent in both clusters, were that moving objects exert more force than static objects, and that larger or stronger objects exert more force than smaller or weaker objects. However, these misconceptions were used in only about half of the cases where they might have been used.

The major difference between the clusters was in the frequency of use of a third misconception. The larger group (roughly two-thirds of the students) tended to use the rule of thumb that gravitational forces only act downward, that is, that an object on top of another exerts a downward force, but there is no countervailing upward force. The students in the large cluster regarded this rule of thumb as being on a par with the others, and used it about 40% of the time. The smaller group (slightly less than one-third of the students) were much more consistent in their belief. In about 75% of the cases in which the "gravity downward" rule was appropriate, these students used it.

These results provide information about the power of the PDP clustering method. The method did reveal consistencies in data patterns that suggested the use of rule-governed behavior by the students. However, the consistencies were not revealed by the weight vectors used internally by the PDP model. In fact, in this study, the examination of the weight vector was actually misleading. The consistencies were revealed, rather clearly, by examining the response vectors that the PDP clustering method assigned to different groups.

These results cannot be used to provide information about how physics students reason about physics, since no attempt was made to use a PDP model to analyze individual problems. What the program models is the way in which physics teachers may identify patterns in student answers. I speculate that the teachers first notice that

some patterns of answers seem to reoccur, which is what the connectionist model noted, and that the teachers then examine these patterns and use their knowledge of physics in order to explain them. The latter activities would correspond to the statistical analysis of the clusters produced by the connectionist model that has been reported here.

SPECULATIVE REMARKS

In a field as diverse as cognition, sweeping conclusions should not be drawn from only three studies. Furthermore, both the system-control and physics student cases require further analysis. Perhaps most important, we need to confirm our analyses in independent studies. This is particularly true for the conclusions based on the analysis of physics student responses, since this situation makes the most explicit use of rule-based behavior.

A particularly interesting question is whether rule-based or connectionist approaches provide better predictors of what people will do when they have to respond directly to a situation, rather than when they have to explain it. To continue with the physics example, suppose students were asked to take an action to affect a situation (e.g., changing the flight of a ball), instead of being asked to analyze force relations. Would the response be better predicted by a connectionist or a rule-based expert system? Similarly, what about medical diagnoses? Do physicians really use all those rules in an expert system, or do they cluster cases on a more intuitive (connectionist-modelable?) basis and then use their knowledge of medicine to justify their clustering? Or, to take a situation in which my colleagues and I have already done some preliminary work, what about the law? It is supposed to be a rule-governed system. To what extent can judicial decisions be based on connectionist models? And if they are, what does that tell us about how the decisions are being made?

When my colleagues and I embarked on these studies, we had expected that they would show the need for rule-based models of complex decision making. Instead, rather simple connectionist models produced adequate simulations in three situations where current AI wisdom would recommend the use of rule-based expert system technology. As always, we need to push our studies into other fields, in order to establish limits on our conclusions. We also need to think carefully about what these results mean for cognitive psychology.

REFERENCES

- FEIGENBAUM, E. A. (1977). The art of artificial intelligence. *Proceedings of the International Joint Conference on Artificial Intelligence*, 5, 1014-1029.
- FODOR, J. A., & PYLYSHYN, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-77.
- HINTZMAN, D. L. (1986). "Schema abstraction" in a multiple trace model. *Psychological Review*, 93, 411-428.
- LUNDELL, J. W. (1988). *Knowledge extraction and the modelling of expertise in a diagnostic task*. Unpublished doctoral dissertation, University of Washington, Seattle, WA.

- MCCLELLAND, J. L., & RUMELHART, D. E. (1988). *Explorations in parallel distributed processing*. Cambridge, MA: MIT Press.
- MCSPADDEN, M. (1989). *Remembering the future*. Unpublished manuscript.
- RUMELHART, D. E., & MCCLELLAND, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, MA: MIT Press.

NOTES

1. The program was a modification of the commercially available computer game MELTDOWN.
2. Hintzman's model assumes a situation in which objects are coded by binary vectors, where each variable in the vector indicates the presence or absence of a property. McSpadden extended Hintzman's coding to represent the values of positive real variables, by using a coding scheme in which the Euclidean distance between two real-valued vectors is approximated by the Hamming distance between the binary-valued vectors representing the real-valued ones. Hintzman's model was not explicitly presented as a connectionist model, but its formal mathematical statement is clearly related to the connectionist approach, and in particular, to McClelland and Rumelhart's (1988) explicitly connectionist "Jets and Sharks" instance-based model of memory. The Hintzman model has

the advantage of being supported by explicit comparisons to data, while Rumelhart and McClelland's model is primarily a heuristic device for illustrating the connectionist approach. For the purpose of the points being made in this article, the differences between the two models can be disregarded. What is new in our work is the use of exemplar-based models as problem-solving models.

3. This technique closely follows the algorithm presented by Hintzman (1986) for modeling memory retrieval experiments.

4. The model predicts that a control variable will fall into an interval, while the person sets it to a point value. This is a property of the coding scheme used to map the continuous variables of the problem into the binary variables required by Hintzman's model.

5. We thank John R. Clement, Andrea di Sessa, and James Minstrell for their participation in this stage of the research.

6. There was only one difference between our algorithm and McClelland and Rumelhart's. They required that the sum of the elements in the vector of external-to-hidden unit weights equal one. We required that the sum of squares of the elements equal one, that is, that all vectors map to a hypersphere of radius one, centered at the origin of the space defined by the weights from external-to-hidden units.

7. In fact, this conclusion was reported at the Society for Computers in Psychology 1988 Annual Meeting. The conclusion was considerably mitigated by the subsequent analysis.