PART VI

CONCLUSION

CHAPTER 26

Reflections on Cognition and Parallel Distributed Processing

D. A. NORMAN

This book presents a novel approach to the study of cognitive processes. The overall proposal is for a major change that alters the fundamental mechanisms of human information processing. Still, despite the impressive results, it is not always obvious just what impact this work should have upon the study of cognitive mechanisms. Yes, the underlying components might change, but what is the effect on models of cognition? Why should cognitive scientists care about the class of models discussed in this book—parallel distributed processing models: PDP. What difference do they make? What really has been learned about cognition? Anything? And what has been left out: Where are the strengths and the weaknesses? Where do we go from here? This chapter addresses these issues.

I start with the issue of level, for I can already hear my colleagues muttering—those who are not part of the enterprise, of course:

> So what? So you have this work on low-level processing structures. Why should it make any difference to our work, or to cognitive science in general? After all, the study of cognition is a high-level pursuit and this PDP stuff is awfully low level—all this talk of weights and synapses and connections between low-level units. Oh sure, it is all very nice and even significant, but it has nothing to do with cognitive science: It has no effect on anything we are doing. We care about language and thought,

> problem-solving and memory, not about neural mechanisms and Hebbian rules for adjusting synaptic weights.[1]

Well, in many ways those concerns are correct. Most of what is presented in this book differs radically from the kind of studies that are called "cognitive." Most are what would normally be called "low-level," concerned with "implementation details," with the hardware and the neurophysiology. Why should it be of concern to the normal, everyday cognitive scientist? And, equally important, how does this work fit with the substantive concerns of the cognitive scientist? What is useful, what is questionable, and what has been left out?

It is still too early to give firm answers to all these questions But I can say where I stand. I am an in-between cognitive scientist. I have watched the developments reported in this book with great interest. Much of my work in the last ten years has been in the spirit of these models, some of it, of course, done with my long-time collaborator, Dave Rumelhart. I have long thought this philosophy to be powerful and important. But I have also disagreed with some aspects, for I have had some problems with the work. It is probably useful to discuss these here, at the end of the book. Some of these problems may reflect inherent, fundamental difficulties. Some may just be temporary problems that persist simply because they haven't been worked on. And some may be pseudoproblems, resulting either from my own misunderstandings or from the current point of view, but that will disappear when the "proper" approach is adopted. Let me start off with the good things: What does this approach—and this book—deliver? Then I will turn to the problems.

Before discussing "the PDP approach," I must remind you that there really isn't a single PDP approach. No simple listing of either strengths or weaknesses is really fair, for there are many different contributors, each with a somewhat different approach with its own set of virtues and vices. Although much of the work has a common thread, the details and even the operative philosophies are oftentimes quite different. This level of disagreement is healthy for a developing science, although it may cause confusion to the unwary onlooker.

---

[1] I wrote this paragraph without realizing that my prophecy had already been confirmed: See the essay by Broadbent (1985) and the partially responsive answer by Rumelhart and McClelland (1985), recapped in Chapter 4.

WHAT IS NEW: THE PDP APPROACH TO THE STUDY OF COGNITION

"This is a difficult book, difficult but important." With this sentence, Rumelhart and I opened our introduction to the book *Parallel Models of Associative Memory,* the earlier work on the study of parallel, distributed models of psychological processing edited by Hinton and J. A. Anderson (1981). The sentence still applies, perhaps with a bit less emphasis on the difficult, a bit more on the important.

This book starts to put it all together, to make significant advances in the mathematics and underlying bases of the approaches and to explore their implications for the study of cognition. The result is a new form of model for understanding the mechanisms of human (and animal) information processing, models that have several important aspects:

- A major rethinking of the levels at which we must think of cognitive processes, involving, among other things, an intimate relationship between neuroscience and psychology.

- A new form of computational model involving large numbers of highly interconnected parallel computational units, a form that takes into account constraints from studies of the nervous system, constraints from psychological data, and a deep understanding of computational requirements.

- New conceptualizations of cognition that change the emphasis from symbolic processing to states that reflect the quality of the match between the data to be described and the possible configurations of the underlying processing structures. These lead to new modes of thinking of cognition, modes that use metaphors such as "stability," "harmony," "local" and "global minima," and "temperature."

- An emphasis on learning—the mechanisms are built around continual adjustment and learning.

Everyone agrees that full understanding of cognition requires a combination of efforts that cut across disciplinary boundaries. Indeed, this is the basic premise underlying the establishment of *cognitive science.* Nonetheless, even relatively recently, there has been little combination of efforts. Scientists in one field tended to think work in other fields interesting, but irrelevant to their own. Thus, cognitive psychologists tended to think that studies of the brain were interesting but irrelevant;

neuroscientists acted as if models of cognition or psychological data were too speculative and unrelated to neural activity (amusing pastimes, but not science); those in artificial intelligence thought the emphasis should be on control structures and representation, and paid little attention to other disciplines. This has changed. There is now a spirit of genuine cooperation. The research reported in this book shows the results. Neurological data offer strong constraints and invaluable guides to the model building efforts: The work reported here both builds upon what is known of the nervous system and asks specific questions to be decided by future research. Psychological evidence is taken seriously: The models must be able to reproduce this form of evidence. And the models must reflect the real demands of computational plausibility, which means that the models must work, must produce answers in appropriate times and with appropriate degrees of accuracy. Each field pushes the other. Each field suggests the data that should be provided by the other, suggests questions to be examined, and sometimes even suggests in which directions future efforts should be directed. The overall effort reflects the appropriate interplay of theory and data, with genuine interaction among the disciplines: a good model for a cognitive science.

On the Computer Metaphor

It has always bothered me that models of psychological processing were thought to be inspired by our understanding of the computer. The statement has always been false. Indeed, the architecture of the modern digital computer—the so-called Von Neumann architecture—was heavily influenced by people's (naive) view of how the mind operated.[2] Now that whole debate can finally be put aside: The work presented here in no way can be interpreted as growing from our metaphor of the modern computer. Here, we are talking about a new form of computation, one clearly based upon principles that have heretofore not had any counterpart in computers. These models are highly parallel, with thousands or millions of elements interacting primarily

[2] Perhaps I had better document this. Simply read the work on cybernetics and thought in the 1940s and 1950s prior to the development of the digital computer. The group of workers included people from all disciplines: See the Macy Conferences on Cybernetics, or "Her Majesty's Conference on Thought Processes." Read the preface to Wiener's book on cybernetics. Everyone who was working together—engineers, physicists, mathematicians, psychologists, neuroscientists (not yet named)—consciously and deliberately claimed to be modeling brain processes.

through activation and inhibition of one another's activity. Each element is highly interconnected, with perhaps tens of thousands of interconnections. Although the processing speed of each element is slow—measured in milliseconds rather than the picoseconds common in today's computers—the resulting computations are fast, faster than is possible with even the largest and fastest of today's machines. Parallel computation means that a sequence that requires millions of cycles in a conventional, serial computer can be done in a few cycles when the mechanism has hundreds of thousands of highly interconnected processors. These neurologically inspired computational processes pose new requirements on our understanding of computation, suggest novel theoretical explanations of psychological phenomena; and suggest powerful new architectures for the machines of the future.

Although the models reported here are new, they carry on a tradition that has long existed, a tradition of building models of neurological processes. Work in neural networks, in cybernetics, and in brain theory has preceded the present efforts. The studies of perceptrons, a computational device invented by Frank Rosenblatt in the 1950s clearly anticipated many of the results in use today (see Rosenblatt, 1962). The critique of perceptrons by Minsky and Papert (1969) was widely misinterpreted as destroying their credibility, whereas the work simply (but elegantly) showed limitations on the power of the most limited class of perceptron-like mechanisms, and said nothing about more powerful, multiple layer models. Grossberg developed many of the ideas in popularity today with his mathematical models of brain structures, starting in the late 1960s through today (see Grossberg, 1982).

Strengths of the PDP Approach

For me, the work has been important for several reasons. In my personal opinion, many students of cognition have boxed themselves into corners, forcing themselves to develop ever more complex rule structures, and ever more complex algorithms for dealing with the behavior of people. Each new experimental finding seems to require a new theory. Clearly, some new approach has been needed. But what?

The PDP approach offers a very distinctive counterapproach. Basically, here we have an adaptive system, continually trying to configure itself so as to match the arriving data. It works automatically, prewired if you will, to adjust its own parameters so as to accommodate the input presented to it. It is a system that is flexible, yet rigid. That is, although it is always trying to mirror the arriving data, it does so by means of existing knowledge, existing configurations. It never expects to make a perfect match, but instead simply tries to get the best match possible at any time: The better the match, the more stable the system. The system works by storing particular events, but the results of its operations are to form generalizations of these particular instances, even though the generalizations are never stored directly. The result, as has been illustrated throughout the chapters of the book, is that although the system develops neither rules of classification nor generalizations, it *acts* as if it had these rules. Thus, the system really mirrors experience; the regularities of its operation result partially from the regularities of the world, partially from the interpretations of the beholder. It is a system that exhibits intelligence and logic, yet that nowhere has explicit rules of intelligence or logic.

These ideas have particular charm for many of us. The attempt always to find a best fit to existing data by maximizing harmony (or minimizing energy or discrepancy) gives mechanism to a common hope among many modern cognitive theorists (e.g., Shepard, 1984). The heavy dependence upon particular instances and the lack of explicit generalizations fit a reasonable body of evidence about the importance of specific experiences (e.g., see Kahneman & D. T. Miller, in press; Medin & Schaffer, 1978). The way by which these systems try to accommodate themselves to the data by minimizing energy or maximizing harmony results in preferred states or interpretations, where the preferences reflect the particular events the system has experienced. This leads to categorization and classification of the input signals, but in a flexible manner—categorization by distance from prototypes, a mode that is consistent with much recent evidence. Here is a system that incorporates learning as a fundamental, essential aspect of its behavior, that makes no attempt to form categories or rules, yet that acts as if it were a prototype-matching, categorization system that has rules and strategies.

The schema. What about the schema, that omnipresent, powerful tool of modern cognitive theory? Well, as Rumelhart, Smolensky, McClelland, and Hinton show in Chapter 14, the schema still exists, but in somewhat different form than that which schema theorists have thought of it (schema theorists such as Rumelhart and I, I might add). I like this new formulation. More than that, I think this an important step in the understanding of thought. Schemas are not fixed structures. Schemas are flexible configurations, mirroring the regularities of experience, providing automatic completion of missing components, automatically generalizing from the past, but also continually in modification, continually adapting to reflect the current state of affairs. Schemas are not fixed, immutable data structures. Schemas are flexible interpretive states that reflect the mixture of past experience and present circumstances.

Because the schema is in reality the theorist's interpretation of the system configuration, and because the system configures itself differently according to the sum of all the numerous influences upon it, each new invocation of a schema may differ from the previous invocations. Thus, the system behaves as if there were prototypical schemas, but where the prototype is constructed anew for each occasion by combining past experiences with the biases and activation levels resulting from the current experience and the context in which it occurs.

Essential properties of human information processing. Some years ago, Bobrow and I listed some properties we felt were essential components of the human cognitive system (Bobrow & Norman, 1975; Norman & Bobrow, 1975, 1976, 1979). We constructed our lists through observations of human behavior and reflection upon the sort of processing structures that would be required to yield that behavior. We concluded that the system must be robust, relatively insensitive to missing or erroneous data and to damage to its parts. Human cognition appears to work well in the face of ambiguity, incompleteness, and false information. On the one hand, the system is not only robust, it is also flexible and creative. On the other hand, the system continually makes errors. Speech is riddled with incomplete sentences, with erroneous words and false starts. Actions are riddled with "slips" and mistakes. People have learned to deal with these problems, in part by elaborate (although mostly subconscious) correcting mechanisms in language, in part through humor and tolerance for the slips that characterize everyday behavior. Nonetheless, these characteristics of human cognition seem to provide important clues about the nature of the processing mechanisms. We argued that the system had to work by descriptions rather than precise specifications, by partial information rather than complete information, and by competition among competing interpretations.

We argued for a set of essential properties: graceful degradation of performance, content-addressable storage, continually available output, and an iterative retrieval process that worked by description rather than by more traditional search. Although these requirements seemed to us to be both necessary and sensible, supported by the evidence, a common complaint was that these were hand-waving speculations, dreams, that there was no method for implementing these concepts. And if they couldn't be built, then they didn't exist—the ideas were wrong. Well, the PDP mechanisms described in this book have exactly the properties we required. Hurrah for our side.

New metaphors for thought. A very useful result of the PDP work is to provide us with new metaphors with which to discuss cognitive activity. These metaphors are more important than might be realized. They accommodate the transfer of knowledge from one application domain to another. The metaphors of PDP are of weights and goodness-of-fit, of stability, of local and global minima or maxima, and of the means to move from one stable state (a local minimum) to another (perhaps a global minimum) by "shaking up the system" or "raising its temperature." We see why systems have what psychologists call set or functional fixity: The systems are stable, even if in only a local minimum; it takes much less evidence to get into a stable state than it takes to get out of it. To change a state is not necessarily easy. You have to shake up the system, heat up the temperature. Don't let it freeze into position. New interpretations suddenly arise, with no necessary conscious experience of how they came about: a moment of nothing, and then clarity, as the system heats up, bounces out of one stable configuration and falls into a new configuration. The descriptions flow smoothly from the tongue. Good metaphors and comfortable descriptions are not scientific criteria. But they make a large difference in the psychological acceptance of the system and in the general utility of the ideas.

Levels of description. And what about the most common complaint of all: Who cares about this low-level stuff? I think that the answer is clear: all of us. Now, be careful about how this is interpreted. I, personally, am not particularly excited by much of the discussions of synaptic adjustment rules and the details of the operation of single neurons (and worse, single synapses). But careful attention to these low-level details is essential in understanding the appropriate primitive structures of cognition.

The whole point for the cognitive scientist is to understand cognition: perception, language, thought, social interaction, action, emotion. . . . To do so, we insist on explanation of the internal processing structures that give rise to cognitive activities. This is why we have spoken of representation, of the mechanisms of memory and attention, of the rule-structures of language, and of the plans, schemas, and beliefs of people. But in developing models of internal activity, we must adopt an appropriate set of primitive descriptors. So far, the primitives have been based upon symbols: symbol manipulating rules, production systems, control structures, and representational systems. How do we know these are the appropriate descriptors? We don't. The choice of primitive building blocks depends upon the choice of computational device. Selection of the PDP computational system changes things dramatically: New conceptualizations emerge, new primitive levels of description are required, new explanations of cognition are possible.

By attending to the seemingly uninteresting level of the adjustment of synaptic weights, we end up with a whole new conceptualization of how concepts are encoded. The point is that some low-level stuff makes large differences in the way that the entire system operates, which in turn makes huge differences in the types of high-level descriptions that are most important for the cognitive scientist. Who would have thought that a person interested in problem solving or thinking should be interested in energy states or measures of "harmony" among units at a "neural-like" level? But guess what? They turn out to be important. A focus upon these levels turns out to change the way we view cognition. The PDP approach forces a re-evaluation of the importance of what would otherwise have been thought to be low-level concepts.

Weaknesses of the PDP Approach

The approach has weaknesses. In part, it is hard to see exactly how to apply it to many of the difficult issues in the study of cognition. In general, the closer to perception or to motor output, the easier it is to see how the work applies. Thus, there seems to be no question about the significance for pattern recognition, for vision or speech understanding, or even for categorization. There have been a few studies that show promise for the control of motor movements (Hinton, 1984; Rumelhart & Norman, 1982). The question arises whether there are fundamental limitations in these models or whether the lack of examples simply represents the early stage of development and the initial concentration of effort upon the mathematics, the modeling techniques, and examples drawn from more restricted domains. The answers, of course, are not known.[3]

The type-token problem. Among the problems that have long bothered me about the PDP approach are some technical issues in

[3] Yes, it is possible to make a PDP network balance a pole (Barto, Sutton, & C. W. Anderson, 1983), but this is really an exercise in demonstrating learning, and the perceptual input and motor output required are, well, trivial. The study isn't trivial—it is a very nice demonstration—but the theme is the adjustment of weights, not perception, not motor control, and not decision or thought. In this section I examine several of these problems: the type-token problem, the problem with variables, the need for an extra, evaluative structure capable of overseeing and passing judgment on others, and then some general issues in learning and consciousness.

computation. There is a list of them, but the two most significant are the type-token distinction and the handling of variables. The type-token problem is to be able to handle different instances of the same concept, sometimes at the same time. Thus, if the system has the knowledge that "John eats a sandwich" and that "Helene eats a sandwich," the system has to treat these as different sandwiches. This capability is not easy for PDP systems. PDP networks are very good at representing general properties, properties that apply to classes of objects. This is where their power to generalize, to generate default values automatically, arises. But the complementary skill of keeping individual instances separate seems much harder.

The issue is addressed in the book; it is discussed in Chapter 3, "Distributed Representations," by Hinton, McClelland, and Rumelhart; in the McClelland and Rumelhart chapters on distributed memory and amnesia; and in McClelland's chapter on "The Programmable Blackboard Model of Reading" (and probably elsewhere that I missed). All this discussion is somewhat symptomatic of the difficulties: In traditional symbolic representations, for example, semantic network and frame-based approaches to representation, the issue is straightforward—the solution is natural. Here, however, considerable complications must be introduced to handle the issues, and even then it is not clear that the problem is entirely solved.

One question that should be raised, however, is whether we really wish the problem to be solved. It is interesting to note that in our typing model, Rumelhart and I concluded that the kinds of errors people make in typing require a system in which there is no type-token distinction at the level of control for the typing of individual letters (Rumelhart & Norman, 1982). The lack of distinction gave rise to some difficulties, but we argued that these are exactly the difficulties that humans face. McClelland makes a similar argument in his chapter on the blackboard model of reading. Thus, it may turn out that the problems faced by PDP systems in handling the type-token distinction are actually virtues with regard to their ability to model human cognition, especially if people have the same problems.

Variables. Once upon a time I worried about how the nervous system could have evolved to do all the things computational systems seemed required to do. Among them was the ability to call subroutines. The problem with calling subroutines was that it required variables: not only the values of the actions to be performed, but the name or symbol of the thing that had requested the subroutine—this is how the system would know where to resume operations once the subroutine was completed. Symbol processing systems are fundamentally variable-interpretation processes, and I couldn't see how they could have evolved.[4] Well, now along comes the PDP approach: In one sense, my problem is solved—there are no variables, save for activation values. All that is represented is particular instances. No need to represent variables: The variable is represented by the configuration of units that are active. But I still am not satisfied.

Now I worry about how PDP units can get away without representing variables. Aren't variables necessary? What about thought? If I solve a problem mentally don't I have to postulate hypothetical situations in my head, evaluate them, and make decisions? How do I plan my day, deciding which activities should be done first, which last, which require others to precede them, and which can be postponed? How do I compose music, or do science? Don't I need to have mental variables, symbols that I manipulate? My answer is "yes." I think this lack is a major deficiency in the PDP approach. I believe this problem can be solved by having several levels of systems, each specialized for different things. The PDP system is fine for perception and motor control, fine for categorization. It is possibly exactly the sort of system required for all of our automatic, subconscious reasoning. But I think that more is required—either more levels of PDP structures or other kinds of systems—to handle the problems of conscious, deliberate thought, planning, and problem solving.

The need for an extra, evaluative structure. A problem with the PDP models presented in this book is that they are too specialized, so concerned with solving the problem of the moment that they do not ask how the whole might fit together. The various chapters present us with different versions of a single, homogeneous structure, perfectly well-suited for doing its task, but not sufficient, in my opinion, at doing the whole task. One structure can't do the job: There do have to be several parts to the system that do different things, sometimes communicating with one another, sometimes not.

The argument seems especially relevant in considering learning. Though many of the learning rules are self-correcting, and therefore tend to converge in one way or another on optimal performance, they seem to be insufficiently sensitive to the global goals and evaluations made by the organism in which they are implemented. While it is now well accepted that the intention to learn, per se, is not a major determinant of learning, this intention mobilizes cognitive activities that result in better learning. Not much is said about these intentions, their

[4] Of course, I can't figure out how the bones and impedance-matching structures of the ears could have evolved, or the complex tendons that control the tens of bones in the hands could have evolved either: My lack of imagination is not restricted to information processing.

source, or the ways in which they influence the system's learning and performance.

When it comes to learning, it is frequently the case that something has to watch over the operations and act as a trainer. But this trainer is separate from the learning mechanisms. And it has to be able to evaluate the quality of performance. How does this take place? What is this second, overseeing mechanism? And how did it get started? How did the trainer know what task it was to train, and when? And how did it acquire the knowledge of what was good performance if it was a task that the person had never before performed? Even in the competitive learning environment (see Chapter 5 by Rumelhart & Zipser) in which learning can take place without an overseeing, evaluatory mechanism, it is often advantageous to "train" the system by careful presentation of the items that are to be classified. Thus, Rumelhart and Zipser "trained" their system on various stimulus situations—"teaching" is what they called it. But how does this come about? Who does the "teaching" or "training"? In some cases, it is another person, but not always.

It is often the case that complex behavior has to be monitored to see if things are proceeding well. This requires some sort of evaluation of the output of the system. But what mechanism does this evaluation? And what mechanism decides what the desired output is to be? I believe this requires an "overlooking" system, one that compares expectations with outcomes. This second, evaluative, system might very well also be a PDP structure, but it must be separate from the first, overlooking the first and operating upon it.

More than one system at a time is required. A second problem is that because any single PDP network can only settle into one state at a time, it is probably the case that multiple systems are required: modules of PDP systems, perhaps. Now, this is not really a criticism of the PDP model. In fact, the property that a PDP network reaches a single state at any moment is both interesting and important: It gives rise to some of the strong points of the model. The point is that although the system is highly parallel and very fast when viewed at the level of computational operations, it is highly serial and relatively slow when viewed at the "higher level" of interpreting and analyzing the resulting state changes. This dual perspective is a strong virtue for the modeling of human cognition, for this duality reflects current understanding. People interpret the world rapidly, effortlessly. But the development of new ideas, or evaluation of current thoughts proceeds slowly, serially, deliberately. People do seem to have at least two modes of operation, one rapid, efficient, subconscious, the other slow, serial, and conscious. The problem, however, is that people can do multiple activities at the same time, some of them quite unrelated to one another. So, in my opinion, a PDP model of the entire human information processing system is going to require multiple units. That is, the complete model requires that the brain consist of several (tens? hundreds? thousands?) of independent PDP-like systems, each of which can only settle into a single state at a time.

Learning and Consciousness

The two problems discussed in the previous section come together into a general statement of the problems of learning and consciousness. To be skilled at something requires that the requisite actions be performed smoothly, effortlessly, with minimum conscious thought or effort. In many ways, skilled performance consists of the execution of automated routines. The PDP structures seem like ideal candidates for mechanisms that produce these automated, subconscious behaviors. But how do these structures come about? What are the procedures for learning? This view suggests that we consider "learning" to be the programming of automated routines.

Under this view, when we learn some new task or skill, two different systems are involved. One involves conscious control and awareness of the task and the contingencies. Call this system DCC (for *deliberate conscious control*—after Norman & Shallice, in press). The other involves a PDP mechanism, a pre-established, preprogrammed set of procedures that get followed with minimum need for conscious control or even awareness.

The basic idea is this: Learning involves the setting of weights on the associative network structures of a PDP system. However, this weight setting requires some evaluative mechanism that determines when things are going well and when they are not, and perhaps other kinds of judgments that can be used to affect the weight setting and establishment of appropriate connections. This is one role for DCC. In addition, during the time while the weights are being established, before the PDP structure is properly set up, if the activity is to be performed at all, it must be through DCC. Hence, two roles for DCC are to control activities when the final "programming" has not yet been completed, and also to control the "programming." Norman and Shallice suggested that DCC only controls activities by influencing activation values. Among other things, this means that the control by PDP and DCC mechanisms are compatible, so that both may operate at the same time. Thus, partially learned activities may be under control of both modes. Conscious control of activities acts by activating or inhibiting the PDP networks, attempting to direct them into appropriate states, but without full power to cause such a result: Conscious control is simply one more source of activation and inhibition that adds and subtracts from the other influences on total activation values.

These views of learning and of direct conscious control are consistent with a number of qualitative phenomena:

- Control by DCC is slow and serial, hence not very effective for some kinds of skills. If DCC control is exerted after appropriate PDP mechanisms exist, then conscious control of activity will be inferior to subconscious control through the PDP mechanisms. Performance improves if done subconsciously, blocking conscious control. This is indeed the procedure taught by a number of athletic training methods: "Stop thinking about your actions and performance will improve."

- Introspection results from the interpretive attempts of deliberate conscious control—by the requirements of the "evaluator" or "trainer" for learning. Introspection is therefore essential for learning, but it is based upon observation of the outputs of a subconscious (PDP) system. As a result, introspections are only capable of accurate descriptions of system states. Because there is no information available about how that state was reached, introspection cannot give reasons for the resulting states. People usually do give reasons however, but these reasons are of necessity constructions and self-hypotheses: We have no conscious awareness of how PDP mechanisms settle into states and no vocabulary for describing the weights and synaptic interactions even were we to have that awareness.

- Consider the role of meaning and understanding in learning. Much has been made of the need for the learner to understand the task and activity. Yet, understanding does not seem relevant to the way that skilled performers operate. Moreover, many things we learn have arbitrary relations: the relation between the name of an object and the object, the order of the letters of the alphabet, the fingering of musical instruments, command sequences of computers. I believe the point is that PDP mechanisms can set up almost any arbitrary relationship. Hence, to the expert, once a skill has been acquired, meaningfulness of the relationships is irrelevant. However, the conscious, supervisory mechanisms require meaningful and understandable relationships: The DCC mechanisms must watch over actions and evaluate them, and for this, understandable and intelligible relationships are important. Understanding is also essential for troubleshooting, but in many ways, this is what the learning process is all about: troubleshooting the performance in order to establish the appropriate connections.

TOWARD A NEW UNDERSTANDING OF HUMAN INFORMATION PROCESSING

The work presented in this book provides the framework for new directions in our understanding of the human information processing system. In the older view, processing was done by separate components that communicated by passing messages—symbols—among themselves. The perceptual system communicated with the memory system. Memory communicated with perception and problem solving. Language systems communicated with the others. The communications consisted of symbolic messages, interpreted through a representational system, that implied a correspondence between the symbol and the referent and, therefore, an interpretive mechanism (see Rumelhart & Norman, in press). The emphasis was on symbolic representation, and hence the emphasis was on the rules of thought, problem solving, and memory. This led to the development of frame-based schema systems, with slots, fillers, default values, and inference mechanisms as well as to production-theory descriptions of human problem solving.

The new view is quite different. Under this new view, processing is done by PDP networks that configure themselves to match the arriving data with minimum conflict or discrepancy. The systems are always tuning themselves (adjusting their weights). Learning is continuous, natural, and fundamental to the operation. New conceptualizations are reflected by qualitatively different state configurations. Information is passed among the units, not by messages, but by activation values, by scalars not symbols. The interpretation of the processing is *not* in terms of the messages being sent but rather by what states are active. Thus, it is what units are active that is important, not what messages are sent. In the conventional system, learning takes place through changes in the representational structures, in the information contained in memory. In this new approach, learning takes place by changes in the system itself. Existing connections are modified, new connections are formed, old ones are weakened. In the conventional system, we distinguished between the information being processed and the processing structures. In the PDP system, they are the same: The information is reflected in the very shape, form, and operation of the processing structures.

A large number of issues are now naturally addressed that were difficult to deal with before. In turn, a large number of things that were easy to do before now are difficult. This is a combination of evolution and revolution. Many of our old ideas are still applicable. Many of the essential, critical problems are still problems. But we have a new set of tools, powerful tools that can help make major advances on a number of fronts. These new mechanisms are fundamentally subconscious ones. Learning results from the natural adjustments to states: It results naturally from the operation of the system. But in the process of examining what extra mechanisms are necessary to guide the learning and the overall behavior of the system, we appear to be led directly to consideration of the interacting roles of processing mechanisms and evaluative mechanisms, of subconscious and conscious processes. A nice fallout.

I still have concerns. I do not know how to map much of my understanding of cognition onto these new structures. I believe the type-token distinction is critical, that the cognitive system can and does manipulate variables, that there must be levels of processes, some of which oversee others. None of these problems seem insurmountable to me. Some may go away, being problems forced upon us (me) only because of old habits of thinking, old views of computation. Some may be solved by new developments. Some may get handled by an amalgamation of methods, combining the best of the more conventional mechanisms of human information processing with these newer views of information processing.

Some old questions are put to rest. To ask "why" a particular action was performed or "why" a particular interpretation was placed on an event means to ask "why" a given internal stable state was reached. But there is no simple answer to these questions. To force the questions is to risk confabulation: answers made up after the fact with no necessary connection to reality. In general, *there is no single reason why any given cognitive state occurs.* The system is multiply connected, multiply controlled. The states are a result of many different factors all impinging upon one another: The state is simply the best match to all the sources of information. This provides a rather nice resolution of many of the arguments that have taken place over the years within cognitive psychology.

ACKNOWLEDGMENTS

I thank Jay McClelland for thoughtful comments on the manuscript.