

- Charness, N., & Campbell, J. I. D. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. *Journal of Experimental Psychology: General*, 117, 115-129.
- Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141-189). Hillsdale, NJ: Erlbaum.
- Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. *American Psychologist*, 27, 1017-1032.
- Gentner, D. R. (1988). Expertise in typewriting. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 1-21). Hillsdale, NJ: Erlbaum.
- Kliegl, R., Smith, J., & Baltes, P. B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology*, 25, 247-256.
- Kliegl, R., Smith, J., Heckhausen, J., & Baltes, P. B. (1987). Mnemonic training for the acquisition of skilled digit memory. *Cognition and Instruction*, 4, 203-223.
- Salthouse, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General*, 113, 345-371.
- Salthouse, T. A. (1985). Anticipatory processing in transcription typing. *Journal of Applied Psychology*, 70, 264-271.
- Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 71-128). Hillsdale, NJ: Erlbaum.

In K.A. Ericsson & J. Smith (Eds.),  
Toward a general theory of expertise.  
 Cambridge: Cambridge University Press (1991).

## 12 *Symbolic connectionism: toward third-generation theories of expertise*

KEITH J. HOLYOAK

In discussing literary expertise, Scardamalia and Bereiter (chapter 7, this volume) mention an intellectual who was accused of not actually reading books, but of raiding them. I must confess that I have similarly raided the chapters in this book, and other sources as well, looking for clues to the directions that new theories of expertise are likely to take. Furthermore, given my entering biases, I no doubt was primed to value some clues more than others. In any case, I found that certain conjectures began to coalesce around the themes linking the diverse approaches to expertise reflected in this volume. In this chapter I shall review the themes that caught my attention and suggest what our current knowledge of expertise implies about the likely course of future theory development in this area. I realize this is a dangerous enterprise, for without a reliable crystal ball, today's predictions can easily turn into tomorrow's embarrassments. Still, the sweeping scope of this collection of reports on the state of the art in expertise research, delivered as they are at the turn of a decade and in the midst of theoretical ferment in cognitive science, surely justifies the risk.

### THE FIRST TWO GENERATIONS

Theories of expertise have now passed through two generations. The first generation centered on the early insights of Newell and Simon (1972; Newell, Shaw, & Simon, 1958): their conceptualization of problem solving as search, and their specification of a small number of heuristic methods for serial search (e.g., means-ends analysis and hill climbing) that could be applied across an indefinitely broad range of domains, with minimal knowledge about the specific content of any particular domain. The first fruits of work on artificial intelligence – the "Logic Theorist" and the "General Problem Solver" – were based on general methods for heuristic search. The obvious first conjecture about expertise was that an expert was someone particularly skilled at general heuristic search.

That conjecture was short-lived. First in chess (Chase & Simon, 1973; de Groot, 1965), then in physics (Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980), and then in myriad other domains it became

apparent that expertise depended crucially on detailed domain knowledge, reflected in specialized memory abilities and inference patterns. Heuristic search methods were general but weak, characteristic of novice rather than expert performance. "Knowledge is power" was the slogan that captured the essence of the second generation of theories of expertise, which dominated both cognitive psychology and artificial intelligence throughout the 1970s and 1980s; for a capsule history, see Feigenbaum (1989).

The second generation of expertise theories transformed the study of cognition, bringing the study of high-level problem solving – the "prototype" area for expertise research – to a prominence it had not enjoyed since the era of Gestalt psychology early in this century. Complex problem solving was seen as a particularly worthy proving ground for cognitive theories, as it necessarily required integration of assumptions about the basic component processes of memory, attention, and reasoning. Complex problem solving has real-world importance and ecological validity as an area of research. Expertise obviously depends on learning how to do something well; hence, the study of procedural learning (rather than only declarative memory) became a crucial area for research. The sophisticated methodology of protocol analysis (Ericsson & Simon, 1980) was developed in conjunction with the second generation of theories. Perhaps most important, second-generation theories were based on a particular canonical cognitive architecture – serial production systems (Newell, 1973) – that created direct ties between cognitive psychology and artificial intelligence and hence contributed to the rise of interdisciplinary cognitive science. In artificial intelligence, production systems became the basis for expert systems, the first major commercial applications of the field. In cognitive psychology, production systems provided the core of Anderson's evolving ACT theory (Anderson, 1976, 1983, 1987), which became the first grand, overarching theory of cognition since the Hull-Spence theories of the 1940s and 1950s.

Second-generation theories of expertise provide a fundamentally simple picture of the development of expertise, which has been most clearly articulated and empirically motivated in Anderson's (1983, 1987) theory of knowledge compilation. Similar principles provide the basis for the Rosenbloom and Newell (1986) theory of chunking. The central idea of knowledge compilation is that operator sequences that yield a successful solution to a problem can be "cached" as new, specialized production rules that will lead to more efficient solutions of similar problems in the future. Compilation can be viewed as an instantiation of Chase and Simon's (1973) hypothesis that expertise involves the acquisition of large integrated "chunks" of knowledge. In knowledge compilation, chunks take the form of larger, more detailed conditions and actions of production rules. Larger conditions provide more precise specification of the circumstances under which the action is appropriate; larger actions allow more to be accomplished by a single "rule-firing." In addition, compilation involves a reduction in the need to access declarative memory, as well as speeded-up rule-firing due to increases in the strengths of

rules with each successful application. Knowledge compilation is closely related to Shiffrin and Schneider's (1977) theory of the development of automaticity with practice at a consistent task. The general picture is that a novice first solves problems by weak methods (often working backward from the goal), and successful solutions result in automatic generation of specialized productions (often allowing forward progress from the initial problem state toward the goal). Relative to the novice, the expert is able to reach the correct solution more quickly and efficiently.

#### WHY A THIRD GENERATION?

As Kuhn (1962) and others have told us, two great pressures drive scientific change: problems encountered in using current theories to explain empirical findings, and the rise of theoretical alternatives. Both of these conditions are evident in the field of expertise research. The chapters in this volume, and other recent papers, describe findings about expert performance that are unexplained and in some cases anomalous from the perspective of second-generation theories. Many authors express dissatisfaction with the current theoretical understanding (e.g., Lesgold, 1989), and Patel and Groen (chapter 4, this volume) and Sloboda (chapter 6, this volume) allude to the potential of an alternative theoretical paradigm – connectionism – that has invigorated other areas of cognitive science in recent years (Feldman & Ballard, 1982; Rumelhart, McClelland, & PDP Research Group, 1986b). I shall first survey the reasons for unease about our current understanding of expertise, and then sketch an emerging framework that may yield a third generation of theories of expertise.

#### *Empirical inconsistencies and theoretical anomalies of expertise*

The canonical second-generation view of expert performance suggests some major uniformities in the nature of expert performance and its acquisition across different domains. Among the supposed commonalities widely cited in textbook treatments are the following: (1) experts perform complex tasks in their domains much more accurately than do novices; (2) experts solve problems in their domains with greater ease than do novices; (3) expertise develops from knowledge initially acquired by weak methods, such as means-ends analysis; (4) expertise is based on the automatic evocation of actions by conditions; (5) experts have superior memory for information related to their domains; (6) experts are better at perceiving patterns among task-related cues; (7) expert problem-solvers search forward from given information rather than backward from goals; (8) one's degree of expertise increases steadily with practice; (9) learning requires specific goals and clear feedback; (10) expertise is highly domain-specific; (11) teaching expert rules results in expertise; (12) performances of experts can be predicted accurately from knowledge of the rules they claim to use. These predicted characteristics

of expertise have received empirical support in varying degrees; nonetheless, surveys of expertise research (e.g., Waldmann & Weinert, 1990) reveal that none provides a universal characterization of expert performance. Although not all of these inconsistencies are incompatible with second-generation theories, none seems clearly illuminated by them. Let us examine some exceptions that have been reported.

**Experts sometimes achieve mediocrity.** Camerer and Johnson, in their review of research on expert clinical decision making (chapter 8, this volume), provide the following summary: Expert decision-makers appear to do remarkably well, "generating hypotheses and inducing complex decision rules. The result is a more efficient search of the available information directed by goals and aided by the experts' superior store of knowledge. Unfortunately, their knowledge and rules have little impact on experts' performances. Sometimes experts are more accurate than novices (though not always), but they are rarely better than simple statistical models."

**Experts sometimes feel more pain.** Scardamalia and Bereiter (chapter 7, this volume) point out that studies of writing provide important exceptions to the idea that experts always accomplish with ease what novices do only with difficulty. As they summarize the field, "expert writers generally are found to work harder at the same assigned tasks than are nonexperts, engaging in more planning and problem solving, more revision of goals and methods, and in general more agonizing over the task." The reason, as Scardamalia and Bereiter point out, is that writing tasks are inherently ill-defined problems. The result is that expert writers tend to define the task in such a way that it is problematic, so that it cannot be accomplished by routine application of available skills, but instead requires them to work at the edge of their competence. This situation, of course, is not unique to expert writers; laborious extended efforts have been documented through the notebooks of scientists (e.g., Tweney's, 1985, analysis of the work of Faraday) and in the verbal protocols of physicists attempting to solve nonroutine problems (Clement, 1989).

**Means-ends analysis can impair learning.** According to second-generation theories, problem solutions initially attained by weak methods, most notably means-ends analysis, provide the grist for knowledge compilation and hence expertise development. But the work of Sweller and his colleagues (Owen & Sweller, 1985; Sweller, Mawer, & Ward, 1983) has shown that having subjects solve algebra-word problems by means-ends analysis actually impairs their performances on subsequent transfer tests. A more effective initial learning strategy involves free forward search from the given information in the absence of an explicit goal.

**Conditions and actions sometimes can be flexibly recoupled.** In second-generation theories, expertise is viewed as the result of automatic evocation of

specialized actions in response to specialized conditions, a connection typically formalized with production rules. However, Allard and Starkes (chapter 5, this volume) report a series of studies of motor performance in which subjects displayed striking abilities to adjust to altered condition-action links. Their evidence indicates that the greater the skill level of the performer, the less the performance decrement resulting from being forced to alter the action required by a given condition – opposite to the prediction that would seem to follow from the hypothesis that expertise involves increasingly automated firing of condition-action rules. Such evidence is reminiscent of classic demonstrations that visual perception adapts quite rapidly to the effects of distorting prisms, with minimal aftereffects when the prisms are removed (Stratton, 1897). In a more cognitive procedural task, learning to use a text editor, Anderson (1987) reported that subjects were able to switch from one text editor to another with relative ease.

**Expertise sometimes can be decoupled from memory performance.** Ever since the seminal work of de Groot (1965) and Chase and Simon (1973) on chess expertise, a standard finding has been that experts have superior memory for stimuli related to their domains. This is, of course, especially the case when the domain of expertise is actually memory performance, as documented in the work of Ericsson and Staszewski (1989) on skilled memory. Nonetheless, expertise and memory performance sometimes are decoupled. Perhaps the most striking example was provided by a study of computer programmers by Adelson (1984), in which she found that novices actually had better memory for details of code than did experts. The reason appeared to be that experts attended more to the overall goal structure of the programming task, rather than to actual code. The experts found it easier to solve a programming task again rather than memorize a detailed solution, whereas the reverse was the case for novices.

Other dissociations of problem-solving and memory performances are reported by Patel and Groen (chapter 4, this volume) for medical diagnosis and by Charness (chapter 2, this volume) for chess. Patel and Groen report that memory for clinical cases does not always increase (and may even be non-monotonic) with medical expertise. In studies of chess, Holding and Reynolds (1982) found that skilled players chose better moves for disorganized but legal chess positions, even though they showed no recall advantage for such positions; and Charness (1981) found that older players had poorer memory for board positions than did equivalently skilled younger players. Such exceptions call into question the common assumption that domain-specific memory skill is directly related to expert problem solving.

**Expertise sometimes can be decoupled from pattern perception.** Closely related to the typically superior memory performances of experts is their greater ability to perceive patterns in stimuli drawn from their domains (e.g., the chess expert can more quickly detect a potential fork). But Allard and Starkes

(chapter 5, this volume) report an exception for expert volleyball players (in contrast to experts in dance, basketball, and hockey, who show the typical expertise advantage in processing structured stimuli). Better volleyball players do not show a consistent advantage over weaker players in perceiving offensive volleyball patterns; however, they do show an advantage in sheer speed of detecting a volleyball (but not a referee!) in photographic slides showing game positions. Allard and Starkes argue that this exception arises because in volleyball, offensive positions typically are designed to deceive the defenders and hence are best ignored in favor of continuous focus on the ball.

**Expert search strategies are extremely varied.** Work on solving routine physics problems indicates that acquisition of expertise is accompanied by a shift from backward search to forward search. In computer programming, however, both novices and experts emphasize backward search from goals (Anderson, Farrell, & Sauers, 1984; Jeffries, Turner, Polson, & Atwood, 1981). The reason appears to be that in computer programming, unlike routine physics problem solving, the initial state places few constraints on the solution path. The search processes are not simply identical for novice and expert programmers; experts do a kind of breadth-first search for a global program design, whereas novices tend to get lost in depth-first searches. Even in physics, nonroutine problems evoke backward search by experts (Tweney, 1985). More generally, expertise in complex tasks often is distinguished not by some single canonical search strategy but by flexible switching among alternative strategies (Dörner & Schölkopf, chapter 9, this volume).

**Performance may not show continuous improvement with practice.** Performances on many tasks seem to improve smoothly with practice, typically following a power function. Yet exceptions are common for complex tasks. As Scardamalia and Bereiter (chapter 7, this volume) put it, "vague notions of 'experience' and 'practice' obscure what is undoubtedly the socially most significant issue in the study of expertise, the issue of why there are such great differences in competence among people with equivalent amounts of experience and practice. No one is disturbed by the fact that experienced physicians are better at diagnosis than interns. We are all disturbed by the possibility that our health may fall into the hands of physicians whose diagnostic expertise has not kept pace with their years of experience."

The acquisition of expertise, even when it does not prematurely "asymptote," does not always follow a smooth path. Ericsson and Staszewski (1989) describe the development of memory skill in their subject SF, who was able to dramatically improve his short-term memory ability by using specialized strategies. His learning curve exhibited flat periods followed by dramatic increments corresponding to qualitative changes in his evolving strategy. Higher levels of expertise may sometimes require not just greater speed and efficiency in processing but a more radical restructuring of the task itself (Cheng, 1985). The development of typing skill, for example, depends on a shift from

serial to parallel planning and execution of finger movements (D. R. Gentner, 1983; Salthouse, chapter 11, this volume).

**Learning need not require goals or feedback.** The canonical second-generation account of skill acquisition emphasizes that learning depends on clear feedback about the success or failure of attempts to achieve goals (Anderson, 1987), and in many contexts there is good evidence in favor of this view. However, studies of the acquisition of musical expertise (Sloboda, chapter 6, this volume) suggest that children typically learn the basic chordal structure of music by age 7 simply from exposure to music and that premature stress on achievement of goals in musical performance may actually be detrimental. Such complex forms of perceptual learning seem to lie outside the scope of second-generation theories. In a much more explicit problem-solving context, Koedinger and Anderson (1989) found that subjects skilled in geometry solved proof problems on the basis of perceptual chunks related to canonical diagrams. The subjects planned solutions in considerably fewer steps than were actually required to execute the proofs, and the nature of the abbreviated planning phase was inconsistent with standard models of knowledge compilation.

**Knowledge can be transferred across domains.** A central tenet of second-generation theories has been that high levels of performance reflect specialized domain knowledge that by its very nature is of little or no use in performing tasks in other domains (or even novel tasks within the same domain). And, indeed, demonstrations of failure to achieve transfer of solution methods across domains are commonplace in the problem-solving literature; for a recent review, see Gick and Holyoak (1987). Nonetheless, there is a growing body of evidence to indicate that with appropriate instruction, knowledge often can be transferred effectively to novel problems (e.g., Brown, Kane, & Echols, 1986; Catrambone & Holyoak, 1989; Gick & Holyoak, 1983). Dörner and Schölkopf (chapter 9, this volume) report that experienced executives were more successful than college students in coping with an unfamiliar problem involving management of a complex dynamic environment. This finding is consistent with other evidence that abstract types of reasoning skills acquired through systematic training can be applied in contexts quite different from that in which training occurred. Nisbett and his colleagues (Cheng, Holyoak, Nisbett, & Oliver, 1986; Fong, Krantz, & Nisbett, 1986; Nisbett, Fong, Lehman, & Cheng, 1987) have found that training in statistics or in everyday deductive reasoning can improve performances on problems with novel content. Theorists such as D. Gentner (1983) and Holyoak (1985) have emphasized the power of analogical thinking as a tool for transfer of knowledge across domains.

Scardamalia and Bereiter (chapter 7, this volume) raise the possibility that expert writing is really a kind of expert thinking, which could have a direct impact on performances at the frontiers of an indefinitely wide range of disci-

plines. Another candidate source of broad expertise is knowledge of mathematics, which provides formal representations of wide potential applicability. Bassok and Holyoak (1989) found training in algebraic sequences (by direct instruction in the relevant concepts and equations, coupled with practice in solving example word problems) was sufficient to allow most students to recognize immediately the relevance of the learned procedures to isomorphic problems with novel content (constant-acceleration problems in physics). Novick (1988; Novick & Holyoak, 1991) found that high levels of expertise in mathematics (as indexed by high scores on the quantitative section of the Scholastic Aptitude Test) was predictive of successful analogical transfer of a novel mathematical procedure. Anzai (chapter 3, this volume) describes a transfer study in which a subject was able to apply general procedures for constructing diagrams to novel types of physics problems. All of these demonstrations of relatively flexible transfer seem to require explanations that go beyond a characterization of expertise as the product of increasingly specialized domain knowledge.

**Teaching expert rules may not yield expertise.** If expert knowledge can be fundamentally represented as a set of production rules, as second-generation theories assume, then the most direct way to improve students' expertise would seem to be to teach them the experts' rules. The "overlay" paradigm for building intelligent tutoring systems assumes that a student's state of knowledge at any time is a subset of that of the expert and that a tutor should incrementally add expert rules to the student's knowledge base (Carr & Goldstein, 1977). In fact, however, several researchers in the area of automated tutoring systems have argued that the overlay paradigm is inadequate (e.g., Clancey, 1986; Wenger, 1987).

**Rules elicited from experts may not predict their performance.** An even more obvious prediction of the view that expertise can be represented by a set of production rules is that if we know what rules experts are using to perform a task, we should be able to predict their performances. This prediction has been challenged by the findings in a study by Lundell (1988; Hunt, 1989): University students were exposed to five hundred displays representing possible readings of instruments for an imaginary but realistic power plant; they were asked to provide a diagnosis for each display and were then told the correct diagnosis. By the end of the session the subjects were accurate on 75% of trials (as opposed to initial chance performance of 25%). Using structured interviews based on techniques for knowledge engineering, Lundell developed a rule-based system to represent each subject's knowledge. The rule-based system for each student was then used to predict the student's performance on a set of new transfer cases. The programs produced the correct diagnoses for just 55% of the new cases, whereas the students were again correct 75% of the time; but worse, the system tailored to an individual student was no more predictive of that student's performance than were systems tailored to other students. In contrast, Lundell found that connectionist

networks constructed for each student by an incremental error-correction algorithm classified 72% of the transfer trials correctly, and each student's performance tended to be better predicted by his or her own network than by someone else's.

Caution is clearly called for in generalizing from Lundell's results, as the amount of training used was modest, and the validity of the method used to extract rules from subjects could be questioned. Nonetheless, the greater predictive success of the connectionist networks is at least suggestive.

**Summary.** When we survey the overall field of expertise research, we find what is surely a disconcerting lack of constancy in the correlates of expertise. There appears to be no single "expert way" to perform all tasks. Perhaps the most apt general characterization of expert performance is that suggested by Dörner and Schölkopf (chapter 8, this volume): An expert is someone capable of doing the right thing at the right time. This characterization is, of course, nearly vacuous; nonetheless, it does suggest a way of understanding some of the variations noted earlier. In general, an expert will have succeeded in adapting to the inherent constraints of the task. If the task can be done most efficiently by forward search, the expert will search forward; if backward search is better, the expert will search backward. If certain patterns of cues are crucial to performing the task well, the expert likely will perceive and remember them; if patterns are not so important, the expert will not selectively process them. The tendency of experts to adapt to task constraints would account for the fact that whereas novices differ widely in the way they organize domain-relevant concepts, experts tend to resemble each other (and differ from novices) in their conceptual organizations (McKeithen, Reitman, Rueter, & Hirtle, 1981; Olson & Biolsi, chapter 10, this volume).

Given the importance of task constraints, as emphasized many years ago by Simon (1969), it might be useful to analyze expertise systematically in terms of the kind of "rational analysis" proposed by Anderson (1990), which attempts to eliminate the need to specify process models. But even this general approach must confront the unfortunate experts at clinical diagnosis, whose adaptation to their task fell short of that afforded by simple linear regression models. It may well be the case that their failures can be explained as the products of generally useful learning strategies that have been confounded by the inherent randomness and poor feedback associated with their target task. But if so, a complete model of expertise acquisition will necessarily require a clear account of human learning mechanisms and their processing limitations, rather than rational analysis alone.

Second-generation theories certainly are capable of explaining some of the diversity in expert performance surveyed earlier. However, to do so they must be elaborated to incorporate learning mechanisms other than knowledge compilation and its variants. A number of researchers have suggested that expertise depends largely on the induction, retrieval, and instantiation of schematic knowledge structures (e.g., Gick & Holyoak, 1983; Koedinger & Anderson,

1989; Schank, 1982; Sweller et al., 1983). Such processes seem quite different from the acquisition and use of specialized production rules. In addition, a great deal of evidence suggests that skilled performance in tasks ranging from the perceptuomotor level (e.g., word recognition and typing) to higher cognitive levels (e.g., discourse comprehension and analogical reasoning) depends on the parallel integration of multiple sources of information. This style of processing cannot be gracefully implemented in serial production systems, the typical architecture for second-generation models.

### *Routine versus adaptive expertise*

The diversity of expert learning and performance suggests the importance of distinguishing qualitatively different varieties of expertise. As Ericsson and Smith (chapter 1, this volume) have argued, it is likely that "research on superior expert performance is benefited more by the development of a taxonomy of different types of mechanisms acquired through different types of learning and adaptation processes than by restricting the definition of expertise to a specific type of acquisition through learning."

A broad distinction between two classes of expertise is suggested by two tentative definitions of expertise raised by Sloboda (chapter 6, this volume). One possible definition is that expert performance involves "the reliable attainment of specific goals within a specific domain." A more demanding definition is that "an expert is someone who can make an appropriate response to a situation that contains a degree of unpredictability." These alternatives correspond to the distinction drawn by Hatano and Inagaki (1986; Hatano, 1988) between *routine* expertise and *adaptive* expertise. (Salomon & Perkins, 1989, elucidated a related distinction between "low-road" and "high-road" mechanisms of transfer.) Whereas routine experts are able to solve familiar types of problems quickly and accurately, they have only modest capabilities in dealing with novel types of problems. Adaptive experts, on the other hand, may be able to invent new procedures derived from their expert knowledge. Hatano and Inagaki (1986) suggested that the key to adaptive expertise is the development of deeper conceptual understanding of the target domain. Such understanding, they argued, is heavily dependent on the conditions under which learning takes place. Understanding is more likely to result when the task is variable and in some degree unpredictable, rather than stereotyped, and when the task is explored freely without heavy pressure to achieve an immediate goal. Understanding can result from sensitivity to internally generated feedback, such as surprise at a predictive failure, perplexity at noticing alternative explanations for a phenomenon, and discoordination due to lack of explanatory links between pieces of knowledge that apparently should be related. Understanding is also fostered by social support and encouragement of deeper comprehension, and by efforts to explain a task to others.

Hatano (1988) exemplified the distinction between routine and adaptive expertise with a cross-cultural contrast between two forms of mathematical calculation skills: use of the abacus in Japan and other Asian cultures (e.g.,

Hatano & Osawa, 1983) and the "street math" of Brazilian children working as vendors. Expertise in use of the abacus leads to extremely rapid calculations and to increased digit span; however, such knowledge cannot readily be generalized to repair "buggy" pencil-and-paper arithmetic procedures (Amaiwa, 1987) or to use nonconventional abacuses with a different base value. In contrast, unschooled Brazilian children who acquire arithmetic skills in the context of selling merchandise on the street can adapt general components of their procedures, such as decomposition and regrouping, to solve novel problems both on the street and in classroom mathematics (Carraher, Carraher, & Schliemann, 1987; Saxe, 1988). The primary difference between the two skills, according to Hatano, is that representations of number relations on the abacus are impoverished in meaning, whereas those used in street math are semantically transparent, analogous to the wider range of activities involving goods and money. In addition, abacus use is basically a solitary skill in which speed and accuracy are the dominant goals, whereas street math is a social enterprise in which transparency to the customer, rather than speed, is crucial.

Other researchers have also noted that learning directed toward understanding is associated with more adaptive forms of expertise. The advantage of learning through free forward search, rather than through goal-dominated means-ends analysis (Sweller et al., 1983), is consistent with this pattern. Both scientific discovery (Clement, 1989) and advanced skills in writing (Scardamalia & Bereiter, chapter 7, this volume) emphasize understanding as an overarching goal. The kind of cognitive knowledge acquired in "open" motor skills (Allard & Starkes, chapter 5, this volume) presumably reflects the inherent variety of the performances through which learning takes place. Skill at jazz improvisation (Sloboda, chapter 6, this volume) seems to be acquired in the context of supportive social interaction involving free exploration rather than fixation on a precise goal.

Hatano (1988) emphasized cases in which an entire skill lent itself more to the acquisition of routine (abacus) or adaptive (street math) expertise. However, it is quite likely that individual differences in the acquisition of a given basic skill may reflect differences in learning styles. For example, Chi, Bassok, Lewis, Reimann, and Glaser (1989) found that better students of physics took a more active approach to learning from worked examples of word problems than did weaker students. The better students continually tried to explain *why* the steps of the illustrated solutions were required. As other investigators have argued, motivation and ability to monitor one's own comprehension seem to be crucial to the acquisition of flexible expertise (Dörner & Schölkopf, chapter 9, this volume; Scardamalia & Bereiter, chapter 7, this volume).

The second-generation theories of expertise, with their emphasis on the acquisition of more specialized production rules through knowledge compilation, can be characterized as attempts to explain routine expertise. As Ericsson and Smith (chapter 1, this volume) point out, most empirical and theoretical work has been directed at accounting for stable superior performance on representative tasks, for which reproductive methods and specific



knowledge are in fact central. Indeed, such theories typically are described as models of "skill acquisition," which, as Wenger (1987) has pointed out, is not coextensive with expertise: "Whereas skill acquisition can be tested by straightforward performance measures, expertise is a much more subtle notion. . . . [It] must also be evaluated by the capacity to handle novel situations, to reconsider and explain the validity of rules, and to reason about the domain from first principles" (p. 302). In Hatano's terminology, skill acquisition results in routine expertise; adaptive expertise requires something else. The second generation of expertise theories was born of the hope that domain-specific knowledge, built up on top of a foundation of weak methods for serial heuristic search, would have the power to fully model human expertise. For some researchers in the area of expertise that hope has now faded, and their loss of innocence is accompanied by theoretical quandaries and increasing openness to new directions. A full account of expertise, it seems, will require a new generation of theories.

### THE THIRD GENERATION: SYMBOLIC CONNECTIONISM

I am, of course, trying merely to predict the future, not to describe a present reality; thus, I intend to sketch not a new theory of expertise, or even a framework for a theory, but simply an evolving paradigm within which I conjecture that new theories of expertise will eventually emerge. The name I give to this paradigm is "symbolic connectionism." The knowledgeable reader may find the label internally contradictory; there has been much discussion of whether or not connectionism will allow cognitive science to do away with symbolic representations altogether. In agreement with the severest critics of connectionism (Fodor & Pylyshyn, 1988; Pinker & Prince, 1988), I believe that reports of the demise of symbols are premature. In particular, adaptive expertise in tasks requiring high-level reasoning appears to require representations that by any reasonable definition are inherently symbolic, just as does the ability to speak and comprehend a human language. But unlike its severest critics, I believe that connectionism offers important new insights into information processing that will sufficiently change the character of cognitive theories that it will be reasonable to speak of a generational change. As was the case in the second generation, these third-generation theories will first arise as models of aspects of the human cognitive architecture, rather than of expertise per se, but then will be tested in part by their ability to account for various forms of expertise.

Symbolic connectionism, as the name implies, will be based on the integration of theoretical ideas drawn from symbolic models (including second-generation models of expertise) and connectionist models. I shall first describe connectionism and its apparent implications for understanding expertise and then provide some reasons why symbolic representations are also required. I shall then briefly describe some early examples of symbolic connectionist models that may point the way toward new treatments of expertise.

### The connectionist view of expertise

Connectionist representations consist of networks of relatively simple processing units connected by links. Processing involves a series of cycles in which in each cycle the units take on new states of activation as a function of their own prior activations, the prior activations of units to which they are connected, and the weights on the interconnecting links. Weights can be either excitatory (tending to make the receiving unit active when the sending unit is active) or inhibitory (tending to make the receiving unit inactive when the sending unit is active). Connectionist models embody three central ideas. First, decision making is based on *parallel constraint satisfaction*: A cycle of processing tends to converge on an activation pattern over units that best satisfies the constraints embodied in the weights on links. The units with the highest asymptotic activations will tend to support each other and to inhibit their competitors. Second, knowledge is, to varying extents, *distributed* over sets of units, rather than identified with single units. Third, learning consists in *incremental revision of weights* on the basis of internally or externally generated feedback concerning the performance of the network. For detailed introductions to connectionism, see Feldman and Ballard (1982) and Rumelhart et al. (1986b).

The parallelism of connectionist networks supports a style of knowledge representation in which decisions are based not on individual rules with large conditions and actions, as is suggested by the notion of knowledge compilation, but rather on interactions between multiple, simpler connections. An example of a connectionist network is depicted in Figure 12.1, which shows the representation of knowledge used in a model of the generation of musical expectations (Bharucha, 1987a, 1987b). The network consists of layers of units representing different types of musical units, such as tones and chords, densely interconnected by links.

There are general advantages to having smaller units of knowledge operating in parallel relative to having larger units of knowledge (such as compiled rules) operating individually. A rule with multiple clauses in its condition is likely to be "brittle," providing no information on how to behave in slightly different situations. And because a rule with a highly specific condition will be less likely to be tested than a rule (or connection) with a more general trigger, any validity estimate will tend to be less reliable for the more specific rule, simply because of the smaller associated sample size (Camerer & Johnson, chapter 8, this volume). Complex rules of the sort produced by knowledge compilation may indeed be useful for performing routine tasks efficiently (see Miyata, 1989, for a connectionist version of knowledge compilation); however, such rules are not likely to provide the key to adaptive expertise.

The connectionist perspective offers a number of possible insights into the nature of expertise (Rumelhart, Smolensky, McClelland, & Hinton, 1986c; Smolensky, 1986). Parallel constraint satisfaction can in principle capture the most striking aspect of human expert performance: Experts tend to arrive

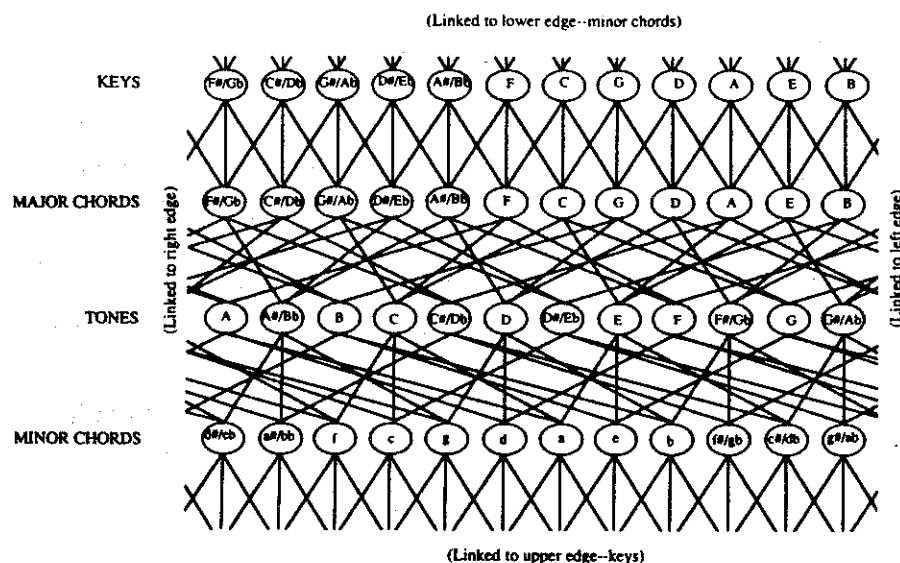


Figure 12.1. A network representing relationships among tones, chords, and keys. Links between units reflect the memberships of tones in chords and of chords in keys. A musical context activates tone units, and the activation spreads through the network, reverberating until it settles into a state of equilibrium. The pattern of activation at equilibrium represents the array of chordal expectations and key implications and influences the consonance and recognition of events that follow. From Bharucha (1987a). Copyright by the Cognitive Science Society Incorporated. Used by permission.

quickly at a small number (sometimes one) of the best solutions to a problem, without serial search through alternative possibilities (de Groot, 1965; Patel & Groen, chapter 4, this volume). Fundamentally, expertise is the product of the acquisition of task-appropriate constraints. A "chunk" of knowledge will correspond to a tightly connected excitatory subnet of mutually supportive units. In contrast to the standard conception of chunking as an all-or-none phenomenon, from the connectionist perspective it may be viewed as a matter of degree of excitatory connectivity. Research has shown that multiple, possibly overlapping chunks appear to form the cores of expert representations of board positions in chess and other games (Chase & Simon, 1973; Reitman, 1976). It is natural, from the connectionist view, for a unit to participate in multiple chunks, and several chunks may be active simultaneously to represent a complex, possibly unique, problem situation. If units representing board positions are connected to units representing possible moves, then the weights on links connecting units for chunks with units for moves will provide constraints that can be used to generate plausible moves in response to particular board configurations.

Connectionist principles provide mechanisms for implementing a parallel,

content-addressable memory-retrieval system (Hinton & Anderson, 1981). Thus, any component of the representation of a problem situation can potentially provide access to similar structures in memory that may provide information relevant to a solution. The more links that connect problem cues to representations of relevant prior knowledge, the more likely it is that such knowledge will be activated. In addition, connectionist processing provides automatic pattern completion, reflecting top-down processing and its interactions with bottom-up processing. Thus, partial cues in the problem situation may activate a cluster of units representing one or more schemes for interpreting the situation, which will in turn provide a richer interpretation of the problem.

A number of connectionist learning schemes have been devised to adaptively modify the weights on links. These learning rules, in effect, pick out statistical regularities among clusters of inputs (Rumelhart & Zipser, 1986) or among reinforced input-output relations (Rumelhart, Hinton, & Williams, 1986a). Schemes of this nature may play an important role in the basic processes of perceptual learning (e.g., the process by which children learn musical patterns from exposure to music; Bharucha, in press), motor learning (e.g., the gradual shift from serial to parallel movement planning in typing; Miyata, 1989), and more central associative learning.

The idea that states of networks have varying degrees of "harmony" or coherence (Smolensky, 1986) suggests that networks can be augmented with mechanisms that allow sensitivity to internal states of the system, thus generating internal feedback. For example, a network may preactivate units representing potential perceptual inputs; deviations of the internally generated activation pattern from actual perceptual inputs may trigger a "surprise" reaction. An asymptotic activation pattern in which active units are inhibiting each other is a sign that contradictory interpretations are simultaneously supported, triggering "perplexity." Such information about the state of the system could potentially contribute to learning by understanding.

### *The need for symbolic representations*

The reader will doubtless have noticed that the foregoing sketch of connectionism as an approach to understanding expertise is laden with promissory notes. There are as yet no serious connectionist models of expertise in chess, physics, or any other domain involving high-level cognition. Furthermore, there are significant impediments to the development of such models. Connectionist theorists are grappling with difficult questions concerning the representational adequacy of their networks. The problems they face have been articulated by critics ranging from the sympathetic (Dyer, 1991; Norman, 1986) to the highly skeptical (Fodor & Pylyshyn, 1988).

The central difficulties all hinge on the need to represent various types of knowledge that are inherently *relational*. To take a simple example, one might suppose that the proposition "the dog chased the cat" could be represented by



the simultaneous activation of units representing the concepts "dog," "chase," and "cat." However, this pattern would be indistinguishable from that activated by the proposition "the cat chased the dog," or for that matter by the word list "chase, cat, dog." Keeping variable bindings straight, as is required to use simple inference rules, poses similar problems. For example, if we know the rule that "if a seller sells a possession to a buyer, then the buyer comes to own the possession," and find out that Bob sold a bicycle to Helen, we can readily conclude that Bob is the seller, the bicycle is the possession, Helen is the buyer, and therefore it is Helen who ends up owning the bicycle. However, this inference requires more than simply activating the relevant concepts; in addition, roles must be represented and bound to the appropriate individuals. Although there have been serious attempts to deal with relational knowledge in connectionist terms (e.g., Hinton, 1981; Smolensky, 1987; Touretsky & Hinton, 1988), fully satisfactory solutions have yet to emerge.

The absence of such solutions makes it difficult for connectionist models to provide accounts of many cognitive abilities that are linked to expertise, such as learning from verbal instructions, representing goal hierarchies (Anderson, 1987; Newell & Simon, 1972), computing analogical mappings (D. Gentner, 1983), assessing similarity of relational structures (Goldstone, Gentner, & Medin, 1989), and accounting for the role of relations in memory retrieval (Ratcliff & McKoon, 1989). A general account of expertise, especially if it is to account for cross-domain transfer of knowledge, will require such inherently relational constructs as goals, types of solution methods, abstract inference rules, and metacognitive procedures.

These representational problems are related to the basic questions of what a unit in a connectionist network can represent. The fact that knowledge is distributed over a set of units does not in itself constrain the "grain size" of what can be represented as a unit. Connectionist theorists have often been extremely vague in defining what a unit can represent. For example, McClelland and Rumelhart (1986) suggested that "a unit may correspond to a neuron, a cluster of neurons, or a conceptual entity related in a complex way to actual neurons" (p. 329). Although radical proponents of parallel distributed processing (PDP) often stress that units represent subsymbolic "microfeatures," many of the most successful connectionist cognitive models have included units representing such complex elements as lexical entries, concepts, or propositions. For example, Bharucha's (1987a) model of musical expectations (Figure 12.1) includes units for abstract musical structures (chords and keys), which are inherently relational in nature, in addition to units for simple tones. It has for some time been recognized that the most serious issues of representational adequacy arise in networks in which concepts that must have a complex internal structure are represented by diffuse, overlapping sets of units (Feldman & Ballard, 1982). Localist connectionist networks, of the sort proposed by Feldman and Ballard, represent individual concepts by a small number of units (a dozen or less, and often just one). Such networks can more readily perform symbolic functions.

The lack of constraint placed on what a unit can represent might be interpreted as a weakness of the connectionist research program; however, it might instead be taken as a clue that representational and processing issues can usefully be separated. As I have noted elsewhere,

for many purposes it is useful to extract the essential properties of connectionist models from their metaphorical neural trappings. In general terms, units represent hypotheses, and connections capture inferential dependencies among hypotheses. Thus if one unit has an excitatory connection to another, this indicates that support for the first hypothesis provides some degree of positive evidence for the second. . . . Summation of activation at each unit serves to integrate multiple sources of converging or contradictory evidence regarding a hypothesis. . . . [M]any of the processing principles embodied in PDP models can be readily incorporated into models that choose to represent hypotheses as symbol structures rather than primitive units. [Holyoak, 1987, p. 994]

The symbolic connectionist paradigm simply accepts that units can potentially represent "hypotheses" with substantial internal complexity. Standard techniques for symbolic representation (or more connectionist-style techniques, as these are developed) can be used to represent relational information; at the same time, connectionist processing techniques can be used to manipulate the units to accomplish such cognitive tasks as memory retrieval and decision making. Such models attempt to capitalize on the complementary strengths of symbolic representation and connectionist processing. As we shall see, this integrated approach proves to be especially useful in deriving inferences from complex relational knowledge. Symbolic connectionist models can make inferences that standard symbolic systems often are too brittle to derive, using knowledge that diffuse connectionist systems cannot readily represent.

### *Symbolic connectionism: case studies*

A number of artificial-intelligence models have been proposed to reflect various hybridizations of symbolic representations and connectionist processing (e.g., Ajjanagadde & Shastri, 1989; Cottrell, 1985; Dolan & Dyer, 1988; Dolan & Smolensky, 1988; Hendler, 1989; Lange & Dyer, 1989; Shastri, 1988; Shastri & Ajjanagadde, 1990; Touretsky & Hinton, 1988). These models are aimed at the development of new, connectionist-style techniques for symbol processing. Another group of models, which I shall sketch here, can combine standard symbolic representations with connectionist constraint-satisfaction procedures to account for psychological data concerning human performance of high-level cognitive tasks, such as discourse comprehension, analogical thinking, and evaluation of explanations. These models suggest the potential breadth of the domains to which symbolic connectionism can be applied, as well as some important commonalities in theoretical mechanisms.

**Discourse comprehension.** Most of us have some degree of expertise at understanding spoken or written discourse. Comprehension typically succeeds with little apparent effort, despite the fact that discourse is fraught with lexical and syntactic ambiguities, as well as lacunae that must be filled by inference processes. Traditional symbolic models typically have depended on the use of parsing and inference rules carefully tailored to produce "correct" interpretations, leading just as typically to various forms of nonhumanlike brittleness. Kintsch (1987) has developed a symbolic connectionist model that appears to circumvent some of these difficulties. His "construction-integration" model has four main components: (1) initial parallel activation of memory concepts corresponding to words in the text, together with formation of propositions by parsing rules; (2) spreading of activation to a small number of close associates of the text concepts; (3) inferring additional propositions by inference rules; and (4) creating excitatory and inhibitory links, with associated weights, between units representing activated concepts and propositions, and allowing the network to settle. The entire process is iterative. A small portion of text is processed, the units active after the settling process are maintained, and then the cycle is repeated with the next portion of text.

The central characteristic of the model is that it allows the parsing and inference rules to apply in a loose, error-prone fashion, overgenerating concepts and propositions that initially form an incoherent representation of the discourse. For example, a parser is given this text: "The lawyer discussed the case with the judge. He said 'I shall send the defendant to prison.'" The parser will create rival propositions that respectively will represent the judge and the lawyer as the referent of "he." The constraint network that is constructed will then use parallel constraint satisfaction to identify a coherent subset of the units, deactivating possible interpretations that do not fit the discourse context.

In addition to accounting for psycholinguistic data on text comprehension, the construction-integration model has recently been extended to account for levels of expertise in planning routine computing tasks (Doane, Kintsch, & Polson, 1990; Mannes & Kintsch, 1989).

**Analogical thinking.** One of the central mechanisms for transfer of knowledge is reasoning by analogy. Two of the basic components of analogical thinking are retrieval of useful analogies from memory and mapping of the elements of a known situation (the source) and a new situation (the target) to identify useful correspondences. Both of these components can be challenging, especially when the analogues have few direct similarities between their elements. People often fail to retrieve potentially useful source analogues (Gick & Holyoak, 1980, 1983), but performance improves when multiple analogues allow induction of a more abstract schema (Bassok & Holyoak, 1989; Brown et al., 1986; Catrambone & Holyoak, 1989; Gick & Holyoak, 1983). Also, once people are directed to relate two analogues, they often succeed in using remote analogues effectively. Indeed, cross-domain analogies are routinely

used as devices to aid in teaching new concepts (e.g., Thagard, Cohen, & Holyoak, 1989).

Because analogical mapping sometimes requires finding relational correspondences in the absence of overt similarities, diffuse connectionist models lack the requisite representational tools. Purely symbolic models have difficulty avoiding combinatorial explosion in searching either for possible analogues in a large memory store or for optimal mappings between two analogues, without being forced to impose unduly limiting restrictions on the search process. Paul Thagard and I, together with our colleagues, have recently constructed symbolic connectionist models of both mapping and analogue retrieval (Holyoak & Thagard, 1989a; Thagard, Holyoak, Nelson, & Gochfeld, 1990). Both the mapping model, ACME, and the retrieval model, ARCS, operate by taking symbolic, predicate-calculus-style representations of situations as inputs, applying a small set of abstract constraints to build a network of units representing possible mappings between elements of two analogues, and then allowing parallel constraint satisfaction to settle the network into a stable state in which asymptotic activations of units reflect the degrees of confidence in possible mappings. The constraints on mapping lead to preference for sets of mapping hypotheses that yield isomorphic correspondences, link similar elements, and map elements of special importance. These same constraints (with differing relative impacts) operate in both the mapping and retrieval models.

As in Kintsch's model of discourse comprehension, ACME and ARCS first overgenerate a large pool of potential candidate hypotheses, and then use constraint satisfaction to select a coherent subset. As an example, the Appendix presents two mathematical word problems that were used by Novick and Holyoak (1991) to investigate analogical transfer in mathematical problem solving. College students first studied the "garden problem," plus a solution to it based on finding least common multiples. They then attempted to solve the target "band problem" using the garden problem as a source analogue. In addition, some subjects were explicitly asked to state the correct mapping for various key concepts and numbers in the band problem. For example, the band members should map onto plants, the number of members in a row of a column onto the number of plants of a kind, and the successful divisor in the band problem (5, which leaves a zero remainder) onto the successful divisor (6) in the garden problem. Note that the two problems have many surface dissimilarities (e.g., band members have no obvious resemblance to plants), contain some misleading similarities (e.g., the divisor 5 in the band problem should map onto 6, not 5, in the garden problem), and are far from isomorphic (e.g., the band problem involves two people who consider a single total number of band members, whereas the garden problem involves three people who consider two different possible total numbers of plants). There was therefore good reason to expect that mapping these two analogues would be challenging for either a person or a computational model.

The ACME model was applied to predicate-calculus representations of the

two problems. Each representation was quite detailed, requiring about 70 propositions to represent the band problem and over 80 to represent the garden problem and its solution. The program proceeded to construct a constraint network consisting of over 1,600 units representing mapping hypotheses (e.g., "band members = plants") interconnected by over 30,000 excitatory and inhibitory links. After about 200 cycles of settling, the network converged on a set of best mappings for the elements of the band problem that were consistent with the intuitively correct set of mappings. These included the optimal mappings between dissimilar concepts, such as band members and plants, and between specific numerical values, such as 5 and 6.

Similarly, the college subjects tested by Novick and Holyoak (1991) achieved over 80% accuracy in providing the correct mappings for the key concepts and numbers. Oral protocols collected from some subjects revealed few overt signs of the mapping process, consistent with the use of a parallel and relatively fast mapping mechanism. Interestingly, knowing the correct mapping did not guarantee successful transfer of the solution method, as about a third of the subjects who had appropriate mappings still failed to develop the analogous solution; furthermore, protocols consisted largely of laborious efforts to work out the *implications* of the correspondences found between the two analogues, after the initial mapping process was apparently completed. These results suggest that general skill in analogical mapping develops quite naturally, but is only one component of skill in analogical transfer. An implication is that adaptive expertise in humans is in part built on powerful constraint-satisfaction mechanisms for finding mappings between complex representations.

**Explanatory coherence.** The problem of evaluating competing explanations arises in an enormous range of domains, including medical diagnosis (Patel & Groen, chapter 4, this volume), legal reasoning, science (Anzai, chapter 3, this volume), and everyday language comprehension and reasoning. A longstanding problem in arriving at criteria for preferring one explanation to a rival is that observations and their relations to possible explanations often are intertwined in complex ways. As Quine put it, "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body" (1961, p. 41). This extreme interdependence has defeated attempts to formulate strict rules for assessing explanatory adequacy.

Thagard (1989) has shown that the problem of evaluating competing explanations can be addressed by a symbolic connectionist model of explanatory coherence: ECHO. The model takes as inputs symbolic representations of basic explanatory relations among propositions corresponding to data and explanatory hypotheses. The system then builds a constraint network linking units representing the propositions. As in ACME and ARCS, a few very general constraints are used in network construction. In ECHO, the constraints support explanations with greater explanatory breadth (more links to data), greater simplicity (fewer constituent assumptions), and greater correspon-

Table 12.1. *Input propositions for ECHO analysis of phlogiston and oxygen explanations*

*Evidence*

- (proposition 'E1 "In combustion, heat and light are given off.")
- (proposition 'E2 "Inflammability is transmittable from one body to another.")
- (proposition 'E3 "Combustion only occurs in the presence of pure air.")
- (proposition 'E4 "Increase in weight of a burned body is exactly equal to weight of air absorbed.")
- (proposition 'E5 "Metals undergo calcination.")
- (proposition 'E6 "In calcination, bodies increase weight.")
- (proposition 'E7 "In calcination, volume of air diminishes.")
- (proposition 'E8 "In reduction, effervescence appears.")

*Oxygen hypotheses*

- (proposition 'OH1 "Pure air contains oxygen principle.")
- (proposition 'OH2 "Pure air contains matter of fire and heat.")
- (proposition 'OH3 "In combustion, oxygen from the air combines with the burning body.")
- (proposition 'OH4 "Oxygen has weight.")
- (proposition 'OH5 "In calcination, metals add oxygen to become calxes.")
- (proposition 'OH6 "In reduction, oxygen is given off.")

*Phlogiston hypotheses*

- (proposition 'PH1 "Combustible bodies contain phlogiston.")
- (proposition 'PH2 "Combustible bodies contain matter of heat.")
- (proposition 'PH3 "In combustion, phlogiston is given off.")
- (proposition 'PH4 "Phlogiston can pass from one body to another.")
- (proposition 'PH5 "Metals contain phlogiston.")
- (proposition 'PH6 "In calcination, phlogiston is given off.")

Source: From Thagard (1989).

dence to analogous explanations of other phenomena. Relations of mutual coherence (modeled by symmetrical excitatory links) hold between hypotheses and the data they explain; relations of mutual incoherence (inhibitory links) hold between competing hypotheses. The resulting network thus typically contains multiple contradictory propositions. Parallel constraint satisfaction then settles the network into an asymptotic state in which units representing the most mutually coherent hypotheses and data are active, and units representing inconsistent rivals are deactivated.

As an example, Table 12.1 lists the sets of propositions relevant to Lavoisier's eighteenth-century arguments comparing the phlogiston and oxygen theories of such phenomena as combustion and respiration. These consist of sets of propositions representing observed evidence and sets representing the compo-

Table 12.2. *Input explanations and contradictions for phlogiston and oxygen explanations*

*Oxygen explanations*

(explain '(OH1 OH2 OH3) 'E1)  
 (explain '(OH1 OH3) 'E3)  
 (explain '(OH1 OH3 OH4) 'E4)  
 (explain '(OH1 OH5) 'E5)  
 (explain '(OH1 OH4 OH5) 'E6)  
 (explain '(OH1 OH5) 'E7)  
 (explain '(OH1 OH6) 'E8)

*Phlogiston explanations*

(explain '(PH1 PH2 PH3) 'E1)  
 (explain '(PH1 PH3 PH4) 'E2)  
 (explain '(PH5 PH6) 'E5)

*Contradictions*

(contradict 'PH3 'OH3)  
 (contradict 'PH6 'OH5)

*Data*

data '(E1 E2 E3 E4 E5 E6 E7 E8))

Source: From Thagard (1989).

nent hypotheses for the two theories. Table 12.2 lists the explanatory relations among the propositions that serve as the input to ECHO, and Figure 12.2 depicts a portion of the resulting constraint network. The solid lines represent excitatory links between cohering propositions, and the inhibitory link exemplifies a relation of contradiction, here between the rival hypotheses that oxygen or phlogiston is the product of combustion. After about 100 cycles, the network settles into an asymptotic state in which the oxygen hypotheses are active and the phlogiston hypotheses have been deactivated, reflecting Lavoisier's conclusion that the oxygen theory was globally superior. Thagard (1989) showed that ECHO is able to model a number of realistic cases of explanation evaluation in both scientific and legal contexts. The basic idea that explanations are evaluated on the basis of their internal coherence and completeness is quite consistent with findings concerning the reasoning of doctors performing medical diagnoses (Patel & Groen, chapter 4, this volume).

*Commonalities among the models*

Although the examples of symbolic connectionist models described earlier span quite different tasks and differ in many important ways, some major commonalities are evident, and they distinguish these models from

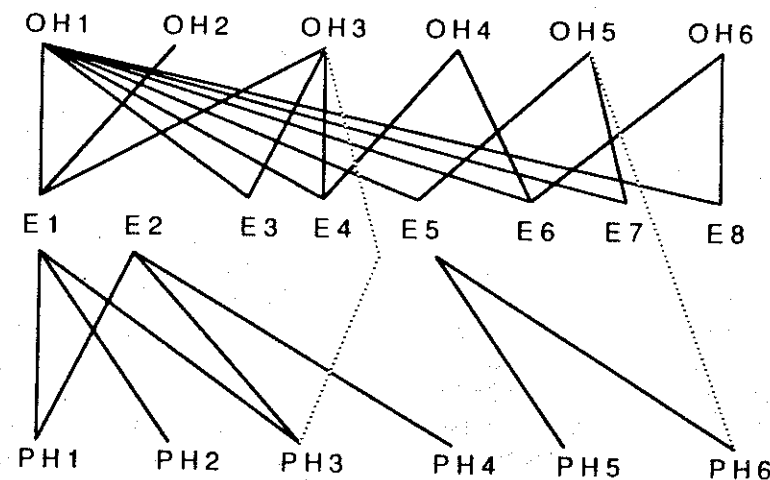


Figure 12.2. A portion of the constraint network formed by ECHO to represent Lavoisier's argument comparing the oxygen and phlogiston theories. From Thagard (1989).

those typical of strictly connectionist or symbolic approaches (Thagard et al., 1990). First, all use highly general constraints, potentially applicable to an unlimited range of examples, to transform symbolic structures into specific constraint networks. The constraints may take the form of parsing rules, inference rules, or other procedures for forming interconnected hypotheses. This on-line construction process is quite different from typical connectionist algorithms for building networks by learning algorithms operating incrementally on numerous examples. A second, related commonality is that the resulting constraint networks are essentially ephemeral. They are built to find a coherent interpretation and then are discarded (although the asymptotic states might readily be used for learning). In this respect these symbolic connectionist models resemble connectionist models of perception (in which ephemeral networks are formed in the process of interpreting sensory inputs) more than typical connectionist models of learning. Third, all the models allow the network-construction procedure to operate in a loose and uncritical manner, overgenerating a set of inconsistent hypotheses; they then rely on parallel constraint satisfaction to select a coherent subset of the hypotheses and discard the rest. This aspect of the models frees them from much of the brittleness that plagues purely symbolic treatments of similar cognitive tasks.

THE FUTURE OF THE THIRD GENERATION

I hope I am not mistaken in taking these examples of symbolic connectionist models as evidence that the third generation of theories of expertise, though still in its infancy, has in fact been born. Although none of these models

addresses the nature of expertise directly, they offer theoretical mechanisms for modeling aspects of higher-level cognition that must surely figure prominently in any model of expert performance. The most salient gap in the current models is that none addresses the crucial issue of learning. Nonetheless, it seems reasonable to expect that learning models can be developed within the symbolic connectionist paradigm. Connectionist schemes for associative learning could readily be incorporated, as could more knowledge-intensive mechanisms associated with symbolic models. The potential of learning mechanisms based on selective recombination of useful existing representations (Holland, 1986; Holland, Holyoak, Nisbett, & Thagard, 1986) deserves exploration. Such mechanisms might, for example, help account for the flexible repairing of conditions and actions found in studies of procedural transfer (e.g., Allard & Starkes, chapter 5, this volume).

The coordination of multiple approaches to learning may in fact yield synergistic benefits. For example, simple models of generalization operate by identifying similarities between elements in successive positive examples of a category. A major limitation of this approach is that in complex relational structures (as opposed to simple feature vectors of the sort typically provided as inputs to connectionist learning algorithms), it is difficult to identify which elements ought to be compared. However, a model of analogical mapping, such as ACME, can in effect force previously unrelated elements (such as band members and plants in the algebra-word problems discussed earlier) into correspondence, making salient the common relational roles that link them. This information could readily be used to guide the process of generalization in the aftermath of analogical transfer (Holyoak & Thagard, 1989b).

### *Inconsistencies and anomalies revisited*

Symbolic connectionism, in its current natal stage, certainly offers no panacea for our incomplete understanding of the inconsistencies and anomalies of expertise reviewed earlier. Still, it may be worthwhile to briefly reconsider these phenomena and offer some very preliminary speculations as to what each may imply about future theoretical developments related to expertise. In some cases the phenomena suggest potential applications of connectionist mechanisms, in other cases the need for symbolic components, and in some cases possible interactions between symbolic representations and connectionist processing.

**Experts sometimes achieve mediocrity.** The deficiencies of experts in clinical decision making, relative to simple statistical models that are closely related to some connectionist learning procedures, suggest some important ways in which these connectionist schemes may require modification or augmentation to account for human learning. The most salient characteristics of the learning situation confronting decision-makers in many areas, such as personnel selec-

tion, is that the pool of relevant features is poorly constrained, and feedback on accuracy of predictions is delayed, intermittent, and sometimes unreliable. Such learning situations are known to impair covariation detection (Holland et al., 1986, chapter 5). It is possible that people often assume that they are correct when feedback is lacking (thus potentially reinforcing flawed decisions). They may also switch attention to new cues when an established regularity is shown to lead to an error, fostering the learning of special-case exceptions, rather than the correction of erroneous weights relating old cues to predictions. In addition, top-down expectations generated by prior beliefs, to the extent they are invalid, may impair bottom-up contingency learning.

**Experts sometimes feel more pain.** The extended efforts at problem reformulation, planning, and revision of goals and methods that characterize experts working at their creative edge remain poorly understood.

**Means-ends analysis can impair learning.** The work of Sweller and his colleagues indicates that free forward search from the given information in the absence of an explicit goal, rather than means-ends analysis, can lead to superior subsequent problem-solving performance. Relatively free problem exploration would be expected to foster the acquisition of broad knowledge of problem constraints and regularities, perhaps using prediction-based learning procedures. Such learning would yield a rich constraint network, which in turn would facilitate the solution of relatively novel problems in the domain.

**Conditions and actions sometimes can be flexibly recoupled.** Such results may have implications for the acquisition of constraint networks. For example, consider the situation facing a macrosurgeon learning to perform microsurgery (Allard & Starkes, chapter 5, this volume). The visual cues that elicit a certain pattern of macromovements must now come to elicit a totally different set of micromovements; yet the old condition-action links must be preserved to allow continued skill in macrosurgery. A connectionist-style solution to this problem in skill acquisition might be to preserve the existing excitatory connections among related visual cues, add new excitatory connections from the visual cues to the required micromovements, and at the same time add inhibitory links between the two sets of visual-to-motor connections, so that the context would be able to "flip a switch" to choose which set of connections would be allowed to operate at a given time. The new skill thus would build on the old (by using preexisting connections among condition cues) while minimizing interference between the two. Similar mechanisms might allow useful but imperfect generalizations to be preserved by coupling them to more specific exception conditions that would override the default when they conflicted (Fahlman, 1979; Holland et al., 1986).

Most connectionist learning schemes have not built new knowledge on top of prior knowledge in this sense; rather, new learning using these algorithms

has involved explicit "unlearning" of existing connections, creating interference. However, recent proposals for learning mechanisms of a more incremental sort, in which established connections are "frozen" before new connections are added to the network, show promise of being better able to capture humanlike flexibility in building new knowledge (Fahlman & Lebiere, 1990).

**Expertise sometimes can be decoupled from memory performance.** From the symbolic connectionist perspective, routine expert performance is basically controlled by the activity of constraint networks. The same networks of constraint relations may help experts to form new memory representations for problems, producing the typical expert advantage in memory tests (Chase & Simon, 1973). However, such memory advantages may be fundamentally incidental in nature. If the expert's constraint network does not include nonessential problem-specific details (as may be the case in expert medical diagnosis; see Patel & Groen, chapter 4, this volume), or if other factors interfere with setting up new memory representations (such as the effects of aging; Charness, 1981), then expert problem solving may be decoupled to some extent from expert memory performance.

**Expertise sometimes can be decoupled from pattern perception.** Similarly, changes in pattern perception may accompany expertise only to the extent the new patterns are required by the constraint network developed to perform the task. In cases such as expertise in volleyball (Allard & Starkes, chapter 5, this volume), a game in which some patterns may actually be misleading, the constraint network of the expert player may not facilitate (or even may inhibit) their detection.

**Expert search strategies are extremely varied.** If high-level heuristics and strategies are represented by units in a constraint network, then contextual cues provided by problem situations can potentially drive the flexible selection of search strategies.

**Performance may not show continuous improvement with practice.** Although connectionist learning algorithms typically are slow and incremental in nature, constraint-satisfaction models of decision processes can undergo relatively radical reorganizations (e.g., Rumelhart et al., 1986c). A change in activation of a unit at one place in the constraint network can trigger a major revision in the overall state of the network. Such rapid changes can be observed in the ECHO model of explanatory coherence when new evidence overturns a crucial assumption of a previously dominant theory (Thagard, 1989). Such rapid changes in network states may underlie major conceptual shifts that sometimes accompany changes in levels of expertise.

**Learning need not require goals or feedback.** Some connectionist learning schemes can identify regularities in inputs without overt feedback (Rumelhart

& Zipser, 1986); see Bharucha (in press) for an application to the learning of musical chords.

**Knowledge can be transferred across domains.** Symbolic connectionist models of analogical reasoning provide potential mechanisms for cross-domain knowledge transfer (Holyoak & Thagard, 1989a).

**Teaching expert rules may not yield expertise.** The knowledge embodied in a constraint network typically will involve subtle interactions and contextual shading that "expert" rules often may miss.

**Rules elicited from experts may not predict their performance.** For the same reason, experts may be unable to articulate the complex interactions between small pieces of knowledge embodied in their constraint networks and thus will be unable to provide accurate descriptions of the basis for their superior task performances. A constraint network can represent the kind of difficult-to-verbalize knowledge associated with expert "intuition."

## CONCLUSION

Although still more speculative than substantive, the symbolic connectionist paradigm appears to have at least the potential to blossom into a new generation of models of cognition and expertise. Whether or not it will in fact do so, and what it will mean if it does, remain open questions. If more radical connectionists are right, elegant general solutions to the kinds of representational issues discussed earlier will quickly enable them to sweep aside the final vestiges of symbol systems, rendering hybrid systems obsolete. But if such solutions come slowly (or not at all), then symbolic connectionism offers at least pragmatic advantages for those trying to construct models of expertise in high-level cognitive tasks. Symbolic connectionists have the luxury of remaining officially agnostic regarding the "ultimate" resolution of debates about the status of symbolic representation. The pragmatic position is that by incorporating symbolic representations as needed, we can take advantage of connectionist processing mechanisms to build models with more of the flexibility in dealing with novelty and complexity that characterizes adaptive expertise. If the radical connectionists are fundamentally right, but overly optimistic about the time scale of their progress, then the symbolic part of symbolic connectionism will serve as a useful stopgap, slowly being replaced by "pure" connectionist mechanisms as these are discovered.

There are, of course, other possibilities. In particular, perhaps the radicals are wrong; perhaps human intelligence is based in part on mental representations that by any reasonable set of criteria are symbol systems. As I survey the current work directed at allowing connectionist networks to perform symbolic functions, my impression is that the more promising lines of attack are better characterized as connectionist-style implementations of symbolic representa-



tions, rather than connectionist eliminations of symbols. This is not in any way to slight the importance of such efforts; on the contrary, new implementations of symbolic functions may have important psychological (and perhaps biological) implications we cannot yet fully anticipate. However, these new symbolic connectionist models will have identifiable components that will perform such symbolic functions as representing variables and their bindings, objects of belief, and other abstract types of knowledge; they will postulate processes by which knowledge can become available for structured recombination and self-reflection. At the same time, these models will use connectionist principles to allow decisions about novel situations to emerge from the graceful integration of multiple constraints. In this possible future for models of high-level cognition, symbolic connectionism will flourish. Time will tell.

#### APPENDIX: ANALOGOUS MATHEMATICAL WORD PROBLEMS USED IN MAPPING EXPERIMENT<sup>1</sup>

##### *Garden problem (source)*

Mr. and Mrs. Renshaw were planning how to arrange vegetable plants in their new garden. They agreed on the total number of plants to buy, but not on how many of each kind to get. Mr. Renshaw wanted to have a few kinds of vegetables, and 10 of each kind. Mrs. Renshaw wanted more different kinds of vegetables, so she suggested having only four of each kind. Mr. Renshaw did not like that, because if some of the plants died, there would not be many left of each kind. So they agreed to have five of each vegetable. But then their daughter pointed out that there was room in the garden for two more plants, although then there would not be the same numbers of all kinds of vegetables. To remedy this, she suggested buying six of each vegetable. Everyone was satisfied with this plan. Given this information, what is the fewest number of vegetable plants the Renshaws can have in their garden?

**Solution.** Because at the beginning Mr. and Mrs. Renshaw agreed on the total number of plants to buy, 10, 4, and 5 must all go evenly into that number, whatever it is. Thus, the first thing to do is to find the smallest number that is evenly divisible by those three numbers, which is 20. So the original number of vegetable plants the Renshaws were thinking of buying could have been any multiple of 20, that is, 20 or 40 or 60 or 80, and so forth. But then they decided to buy two additional plants that they had not been planning to buy originally, so the total number of plants they actually end up buying must be 2 more than the multiples of 20 listed earlier, that is, 22 or 42 or 62 or 82, and so forth. This means that 10, 4, and 5 will now no longer go evenly into the total number of plants. Finally, the problem states that they agree to buy six of each vegetable, so the total number of plants must be evenly divisible by 6. The

<sup>1</sup> Adapted from Novick and Holyoak (1991).

smallest total number of plants that is evenly divisible by 6 is 42, and that is the answer.

##### *Band problem (target)*

Members of the West High School band were hard at work practicing for the annual homecoming parade. First they tried marching in rows of 12, but Andrew was left by himself to bring up the rear. The band director was annoyed because it did not look good to have one row with only a single person in it, and of course Andrew was not pleased either. To get rid of this problem, the director told the band members to march in columns of eight. But Andrew was still left to march alone. Even when the band marched in rows of three, Andrew was left out. Finally, in exasperation, Andrew told the band director that they should march in rows of five in order to have all the rows filled. He was right. This time, all the rows were filled, and Andrew was not alone any more. Given that there were at least 45 musicians on the field, but fewer than 200 musicians, how many students were there in the West High School band?

#### ACKNOWLEDGMENTS

Preparation of this chapter was supported by contract MDA903-89-K-0179 from the Army Research Institute. Miriam Bassok, Patricia Cheng, K. Anders Ericsson, Laura Novick, Paul Thagard, and Michael Waldmann provided helpful comments on earlier drafts.

#### REFERENCES

- Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 483-495.
- Ajjanagadde, V., & Shastri, L. (1989). Efficient inference with multi-place predicates and variables in a connectionist system. In *Proceedings of the 11th Annual Meeting of the Cognitive Science Society* (pp. 396-403). Hillsdale, NJ: Erlbaum.
- Amaiwa, S. (1987). Transfer of subtraction procedures from abacus to paper and pencil (in Japanese with English summary). *Japanese Journal of Educational Psychology*, 35, 41-48.
- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, 94, 192-210.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., Farrell, R., & Sauers, R. (1984). Learning to program in LISP. *Cognitive Science*, 8, 87-129.

- Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 153-166.
- Bharucha, J. J. (1987a). A connectionist model of musical harmony. In *Proceedings of the 9th Annual Conference of the Cognitive Science Society* (pp. 508-517). Hillsdale, NJ: Erlbaum.
- Bharucha, J. J. (1987b). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5, 1-30.
- Bharucha, J. J. (in press). Tonality, expectation and learnability. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication*. Washington, DC: American Psychological Association.
- Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development*, 1, 103-121.
- Carr, B., & Goldstein, I. P. (1977). Overlays: A theory of modeling for computer-aided instruction (AI Lab Memo 406, Logo Memo 45). Cambridge, MA: MIT, Artificial Intelligence Laboratory.
- Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1987). Written and oral mathematics. *Journal for Research in Mathematics Education*, 18, 83-97.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1147-1156.
- Charness, N. (1981). Visual short-term memory and aging in chess players. *Journal of Gerontology*, 36, 615-619.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review*, 92, 414-423.
- Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, 18, 293-328.
- Chi, M. T. H., Bassok, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182.
- Chi, M. T. H., Feltoovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Clancey, W. (1986). Qualitative student models. *Annual Review of Computational Science*, 1, 381-450.
- Clement, J. (1989). Learning via model construction and criticism: Protocol evidence on sources of creativity in science. In J. Glover, R. Ronning, & C. Reynolds (Eds.), *Handbook of creativity: Assessment, theory, and research* (pp. 341-381). New York: Plenum.
- Cottrell, G. W. (1985). Connectionist parsing. In *Proceedings of the 7th Annual Conference of the Cognitive Science Society* (pp. 201-211). Hillsdale, NJ: Erlbaum.
- de Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton.
- Doane, S. M., Kintsch, W., & Polson, P. G. (1990). Modeling UNIX command production: What experts must know (ICS Technical Report 90-1). Institute of Cognitive Science, University of Colorado.
- Dolan, C., & Dyer, M. G. (1988). Parallel retrieval and application of conceptual

- knowledge (Technical Report UCLA-AI-88-3). Department of Computer Science, UCLA.
- Dolan, C., & Smolensky, P. (1988). Implementing a connectionist production system using tensor products (Technical Report UCLA-AI-88-15). Department of Computer Science, UCLA.
- Dyer, M. G. (1991). Symbolic NeuroEngineering for natural language processing: A multilevel research approach. In J. Barnden & J. Pollack (Eds.), *Advances in connectionist and neural computation theory*. Vol. 1: *High level connectionist models* (pp. 32-86). Norwood, NJ: Ablex.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.
- Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235-267). Hillsdale, NJ: Erlbaum.
- Fahlman, S. (1979). *NETL: A system for representing and using real-world knowledge*. Cambridge, MA: MIT Press.
- Fahlman, S., & Lebiere, C. (1990). The cascade-correlation learning architecture (Technical Report CMU-CS-90-100). School of Computer Science, Carnegie-Mellon University.
- Feigenbaum, E. A. (1989). What hath Simon wrought? In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 165-182). Hillsdale, NJ: Erlbaum.
- Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, 6, 205-254.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. In S. Pinker & J. Mehler (Eds.), *Connections and symbols* (pp. 3-71). Cambridge, MA: MIT Press.
- Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, 18, 253-292.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
- Gentner, D. R. (1983). The acquisition of typing skill. *Acta Psychologica*, 54, 233-248.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306-355.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.
- Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of training: Contemporary research and applications* (pp. 9-46). New York: Academic Press.
- Goldstone, R. L., Gentner, D., & Medin, D. L. (1989). Relations relating relations. In *Proceedings of the 11th Annual Meeting of the Cognitive Science Society* (pp. 131-138). Hillsdale, NJ: Erlbaum.
- Hatano, G. (1988). Social and motivational bases for mathematical understanding. In G. B. Saxe & M. Gearhart (Eds.), *Children's mathematics* (pp. 55-70). San Francisco: Jossey-Bass.
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262-272). San Francisco: Freeman.

- Hatano, G., & Osawa, K. (1983). Digit memory of grand masters in abacus-derived mental calculation. *Cognition*, 15, 95-110.
- Hendler, J. A. (1989). Marker-passing over microfeatures: Towards a hybrid symbolic/connectionist model. *Cognitive Science*, 13, 79-106.
- Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 191-217). Hillsdale, NJ: Erlbaum.
- Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.
- Holding, D. H., & Reynolds, R. I. (1982). Recall or evaluation of chess positions as determinants of chess skill. *Memory & Cognition*, 10, 237-242.
- Holland, J. H. (1986). Escaping bitterness: The possibilities of general purpose machine learning algorithms applied to parallel rule-based systems. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 593-623). Los Altos, CA: Morgan Kaufmann.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.
- Holyoak, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 19, pp. 59-87). New York: Academic Press.
- Holyoak, K. J. (1987). A connectionist view of cognition [Review of *Parallel distributed processing* by D. E. Rumelhart, J. L. McClelland, & PDP Research Group]. *Science*, 236, 992-996.
- Holyoak, K. J., & Thagard, P. (1989a). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295-355.
- Holyoak, K. J., & Thagard, P. (1989b). A computational model of analogical problem solving. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 242-266). Cambridge University Press.
- Hunt, E. (1989). Connectionist and rule-based representations of expert knowledge. *Behavior Research Methods, Instruments, & Computers*, 21, 88-95.
- Jeffries, R., Turner, A. T., Polson, P. G., & Atwood, M. E. (1981). The processes involved in designing software. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 255-283). Hillsdale, NJ: Erlbaum.
- Kintsch, W. (1987). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
- Koedinger, K. R., & Anderson, J. R. (1989). Perceptual chunks in geometry problem solving: A challenge to theories of skill acquisition. In *Proceedings of the 11th Annual Meeting of the Cognitive Science Society* (pp. 442-449). Hillsdale, NJ: Erlbaum.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.
- Lange, T., & Dyer, M. G. (1989). High-level inferencing in a connectionist network. *Connection Science*, 1, 181-217.
- Larkin, J. H., McDermott, J., Simon, D., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.
- Lesgold, A. (1989). Context-specific requirements for models of expertise. In D. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 373-400). Cambridge, MA: MIT Press.
- Lundell, J. W. (1988). *Knowledge extraction and the modelling of expertise in a diagnostic task*. Unpublished Ph.D. dissertation, University of Washington.

- McClelland, J. L., & Rumelhart, D. E. (1986). Introduction to Part V. In J. L. McClelland, D. E. Rumelhart, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 327-331). Cambridge, MA: MIT Press.
- McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, 13, 307-325.
- Mannes, S. M., & Kintsch, W. (1989). *Planning routine computing tasks: Understanding what to do* (ICS Technical Report 89-9). Institute of Cognitive Science, University of Colorado.
- Miyata, M. (1989). A PDP model of sequence learning that exhibits the power law. In *Proceedings of the 11th Annual Meeting of the Cognitive Science Society* (pp. 9-16). Hillsdale, NJ: Erlbaum.
- Newell, A. (1973). Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing* (pp. 463-526). New York: Academic Press.
- Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, 65, 151-166.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science*, 238, 625-631.
- Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 531-546). Cambridge, MA: MIT Press.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 510-520.
- Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology*, 77, 272-284.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In S. Pinker & J. Mehler (Eds.), *Connections and symbols* (pp. 73-193). Cambridge, MA: MIT Press.
- Quine, W. V. O. (1961). *From a logical point of view* (2nd ed.). New York: Harper Torchbooks.
- Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, 21, 139-155.
- Reitman, J. S. (1976). Skilled perception in go: Deducing memory structures from inter-response times. *Cognitive Psychology*, 8, 336-356.
- Rosenbloom, P. S., & Newell, A. (1986). The chunking of goal hierarchies: A generalized model of practice. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 247-288). Los Altos, CA: Morgan Kaufmann.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986a). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318-362). Cambridge, MA: MIT Press.
- Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (Eds.). (1986b). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 327-331). Cambridge, MA: MIT Press.

- lel distributed processing: *Explorations in the microstructure of cognition* (2 vols.). Cambridge, MA: MIT Press.
- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986c). Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 7-57). Cambridge, MA: MIT Press.
- Rumelhart, D. E., & Zipser, D. (1986). Feature discovery by competitive learning. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 151-193). Cambridge, MA: MIT Press.
- Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, 24, 113-142.
- Saxe, G. B. (1988). Candy selling and math learning. *Educational Researcher*, 17, 14-21.
- Schank, R. C. (1982). *Dynamic memory*. Cambridge: Cambridge University Press.
- Shastri, L. (1988). A connectionist approach to knowledge representation and limited inference. *Cognitive Science*, 12, 331-392.
- Shastri, L., & Ajjanagadde, V. (1990). *From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings* (Technical Report MS-CIS-90-05). Computer and Information Science Department, University of Pennsylvania.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 194-281). Cambridge, MA: MIT Press.
- Smolensky, P. (1987). *On variable binding and the representation of symbolic structures in connectionist systems* (Technical Report CU-CS-355-87). Department of Computer Science, University of Colorado.
- Stratton, G. M. (1897). Vision without inversion of the retinal image. *Psychological Review*, 4, 341-360, 463-481.
- Sweller, J., Mawer, R. F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, 112, 639-661.
- Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, 12, 435-467.
- Thagard, P., Cohen, D. M., & Holyoak, K. J. (1989). Chemical analogies: Two kinds of explanation. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence* (pp. 818-824). San Mateo, CA: Morgan Kaufmann.
- Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence*.
- Touretsky, D., & Hinton, G. (1988). A distributed connectionist production system. *Cognitive Science*, 12, 423-466.
- Tweney, R. (1985). Faraday's discovery of induction: A cognitive approach. In D.

- Goodling & F. James (Eds.), *Faraday rediscovered: Essays on the life and work of Michael Faraday, 1791-1867* (pp. 189-209). New York: Stockton Press.
- Waldmann, M. R., & Weinert, F. E. (1990). *Intelligenz und Denken* [Intelligence and thinking]. Göttingen: Hogrefe.
- Wenger, E. (1987). *Artificial intelligence and tutoring systems: Computational and cognitive approaches to the communication of knowledge*. Los Altos, CA: Morgan Kaufmann.