of which is rendered in the Book of Genesis, belongs on the trash-heap of outdated folk theory. All the same, she concedes that there will be those who "may tend to see the revision of folk theory and the rise of neurobiological-psychological theory as the irreparable loss of our humanity." But not to worry, because

> it may be a loss, not of something necessary for our humanity, but of something . . . that, though second nature, blinkers our understanding and tethers our insight . . . . The loss, moreover, may include certain folk presumptions and myths, that, from the point of view of fairness and decency, we come to see as inhumane.

With the words "fairness," "decency," and "inhumane," Churchland makes her first and only reference to the ethical dimension of the mind-body problem, after having exclusively considered its scientific dimension. In her earlier discussion of Descartes's contributions she did not mention his motivation for holding the substance dualist view, namely the argument that the body, being a machine, could not be guided by moral principles; hence the mind, which obviously *is* guided by such principles, cannot be a physical part of the body-machine. More important, Churchland has provided an inadequate account of Kant's treatment of the mind-body problem, which is generally considered to have initiated the Copernican revolution in philosophy that Churchland thinks is about to be set off by neuroscientists. By pointing out that we live in two metaphysically distinct worlds, Kant had replaced Cartesian substance dualism with an "epistemic" dualism. One of these worlds is that constructed by the theoretical reason of science, whose natural objects (including the brain of *Homo sapiens*) are governed by laws of causal determination. The other world is that constructed by the practical reason of ethics, whose rational human subjects are governed by laws of freedom that individual free will imposes on each person's actions. Here practical reason justifies the concept of free will, not on the introspective basis, which, as Churchland justly points out, cannot claim evidential priority over all other empirical arguments advanced by theoretical reason, but as a logically necessary constituent of the intuitive theory of personhood that governs interpersonal human relations. As such, the practical concept of free will is not, in principle, subject to reduction by scientific theory, be it neurobiological or psychological.

Accordingly, from the perspective of epistemic dualism, the neurophilosophical brouhaha about the reducibility of psychology to neurobiology is, to use one of Churchland's phrases, "mere crinkum-crankum." It is, after all, immaterial for the resolution of the deep mind-body problem of praxis, posed by the paradoxical human condition of simultaneously facing the two incommensurate realities of science and of ethics, whether psychological theories are or are not reducible to neurobiological theories.

Neuroscientists and psychologists do not need much assistance from philosophers in their struggle with the mind-body problem, as it is posed within the context of theoretical reason. As Churchland herself points out, the controversy regarding neuroscientific reduction of psychological theories will be settled anyhow, in the wash of future experimental and theoretical developments. My own expectations are those of a member of the set styled "boggled skeptics" by Churchland. We boggled skeptics tend to view the human brain as belonging to a class of phenomena whose very complexity limits the extent to which theories designed to explain them can be successfully reduced by theories developed to explain less complex phenomena. As a neuroscientist, I believe that all mental phenomena are *in principle*, explainable by neurobiological theories, just as, as a physical chemist, I believe that all neurobiological theories are, *in principle*, explainable by physico-chemical theories. Moreover, I look forward to some progress still being made in the venerable enterprise of reductionist neuroscience. Yet, I doubt that a complete reduction is de facto possible. *My* cardinal hunch is that a significant residue of unreduced psychological, as well as neurobiological, theory will remain with us long into the future.

Where neuroscientists and psychologists do need philosophical help is in fathoming not the physical but the metaphysical infrastructure of folk presumptions and myths and the likely consequences for the human condition of their abandonment. Churchland is not one of the folks who can provide that help.

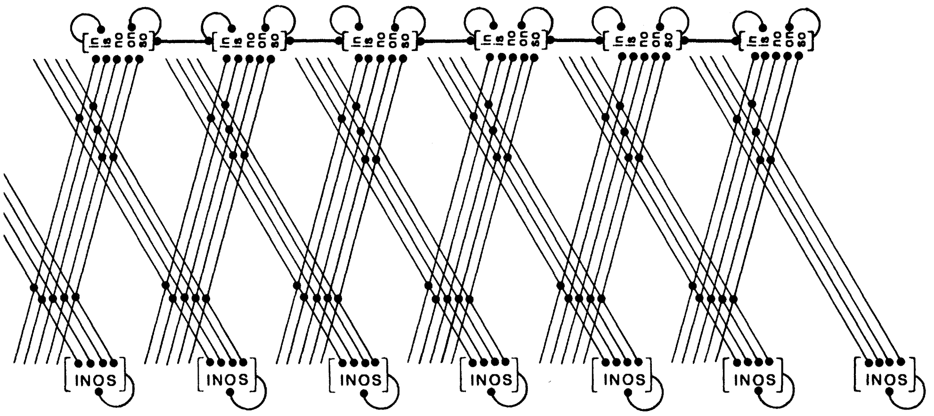
GUNTHER S. STENT
Department of Molecular Biology,
University of California,
Berkeley, CA 94720

A Connectionist View of Cognition

Parallel Distributed Processing. Explorations in the Microstructure of Cognition. DAVID E. RUMELHART, JAMES L. MCCLELLAND, and the PDP RESEARCH GROUP. MIT Press, Cambridge, MA, 1986. In two volumes. Vol. 1, Foundations. xx, 547 pp., illus. $27.50. Vol. 2, Psychological and Biological Models. xiv, 611 pp., illus. $27.50. Computational Models of Cognition and Perception. A Bradford Book.

Two types of devices are known that can support such functions as perception, memory, language, and problem solving. One is the modern digital computer, programmed to produce "artificial intelligence" (AI); the other is the human brain, which produces the natural variety. Given that the latter device seems more intimately connected to the human mind, it may seem surprising that the dominant metaphor for developing theories of mental processes has been based on the former. Modern cognitive psychology, which has come of age with the digital computer, has been more heavily influenced by computer science than by brain science. There are at least two reasons for this. First, we know vastly more about the functioning of computers than of brains. Second, the basic approach of cognitive psychology has been predicated on a philosophical position known as functionalism, which emphasizes that mental functions can be analyzed at an abstract level separate from their physical realization. Just as a computer program can be described without reference to the particular hardware on which it runs, a functional analysis of cognition need not directly refer to brain processes.

The avowed intent of the authors of *Parallel Distributed Processing* is "to replace the 'computer metaphor' as a model of mind with the 'brain metaphor' " (vol. 1, p. 75). The publication of this massive work is a landmark event in cognitive science for a mixture of scientific and sociological reasons. The principles of parallel distributed processing (PDP), a variety of "connectionism," challenge the functionalist attitudes of cognitive psychologists, offer a distinct alternative to conventional AI techniques, and suggest representations of linguistic knowledge very different from the rule systems typically used by linguists. The volumes appear against a backdrop of conferences, workshops, and seminars devoted to the PDP approach. Although connectionism in fact has a long heritage and current models have been actively developed over the past decade, the approach has recently acquired the vigor of a movement in the first bloom of youth. The movement has a proselytizing bent, and talk of a Kuhnian "paradigm shift" is in the air, accompanied by a spirited mix of hype and hope on the part of adherents and by expressions of skepticism from various critics. *Parallel Distributed Processing*



"A hardwired processing structure for bottom-up processing of the words *IN, NO, ON,* and *SO* presented to any two adjacent letter slots. Note that each letter slot participates in two different word slots, except at the edges." [From *Parallel Distributed Processing*, chapter 16.]

provides a focus for this excitement. As one contributor puts it, "The present book offers an alternative paradigm for cognitive science, the *subsymbolic paradigm*, in which the most powerful level of description of cognitive systems is hypothesized to be lower than the level that is naturally described by symbol manipulation" (vol. 1, p. 195).

The basic tenet of parallel distributed processing is that information is represented solely by patterns of activation over neuronlike "units" linked by synapselike "connections," which can be either excitatory or inhibitory. Input units respond directly to features of the environment, and output units represent responses of the system. In between may lie "hidden" units that perform internal processing. Individual units update their activation level (a numerical value) as a function of the activation levels and connection weights of other units that feed into the system. Processing consists of a set of input units being activated (typically by an environmental stimulus), which causes activation patterns to propagate (across hidden units if the network is multilayered), eventually activating a set of response units. Learning is based on mathematical algorithms that adjust connection weights so that inputs produce responses that are more appropriate by some specified criterion.

The term "distributed" has two distinct meanings within connectionist models. All such models perform *distributed processing* in that each unit adjusts its activation levels and associated connection weights on the basis of local computations. Such systems can exhibit massive parallelism and do not require oversight by a central processor. Some PDP models also use *distributed representations*, in which meaningful elements are represented by patterns of activation over a number of units, rather than by single units. For example, a distributed representation of the word *cat* might consist of two active units, one representing the occurrence of *ca* in the first two letter positions and another representing the occurrence of *at* in the final two positions. Each of these units would also form part of the representations of many other words. Distributed representations contribute to some of the more interesting properties of PDP networks as well as to some of their most difficult representational hurdles.

These two volumes consist of 26 chapters by David Rumelhart, James McClelland, and 14 other members of the "PDP Group" that became active at the University of California at San Diego in 1982, with more recent offshoots at Carnegie-Mellon University and elsewhere. Although several notable figures in the connectionist movement are not represented, the contributors constitute a major subset. They also represent a diverse set of scientific disciplines, including cognitive psychology, computer science, physics, mathematics, neuroscience, and molecular biology. This collaboration in itself constitutes a major contribution to the interdisciplinary field of cognitive science. Although the work has implications for several fields, the central focus is human cognition. This emphasis reflects the fact that both McClelland and Rumelhart are cognitive psychologists; the third major contributor, Geoffrey Hinton, is a psychologically oriented computer scientist.

The chapters fall into five major groups. Part 1, in volume 1, includes four overview chapters by McClelland, Rumelhart, and Hinton. (The final chapter in volume 2 is also an overview.) Chapters 1 through 3 in particular are essential reading, providing a general description of the motivation for parallel distributed processing, a survey of the core ideas and their variations, and an introduction to the notion of a "distributed representation."

The other sections of the books are much more technical. Part 2 includes four chapters on connectionist schemes for models of learning by adjusting connection weights, the area that constitutes the core of recent theoretical advances. Each chapter in this section combines mathematical analysis with computer simulation. Chapters 5 and 8, by Rumelhart and colleagues, present two general algorithms for learning in networks. *Competitive learning* (chapter 5) is an algorithm whereby each unit in a mutually inhibitory set learns to respond when patterns with a particular feature are presented. The basic idea is that whichever unit responds most actively to an input pattern adjusts its connection weights so as to respond slightly more strongly to future occurrences of that input. The competing units eventually learn, without guidance from an external "teacher," to partition the patterns in such a way that each unit responds to a unique feature combination. The *generalized delta rule* (chapter 8) is a procedure for adjusting the weights on incoming connections as a function of the difference between a unit's obtained activation value and the "correct" value. A "teacher" provides the correct activation values for the output units, and the algorithm then propagates weight adjustments back to hidden and input units, including units that do not connect directly to output units. By exploiting hidden units, the algorithm can learn to compute simple versions of functions such as "exclusive or" and "odd versus even" that cannot be computed with only input and output units. The latter limitation was highlighted by Minsky and Papert in 1969 in a critique of "perceptrons," which were networks that lacked hidden units. That critique effectively ended an earlier generation's interest in adaptive networks; the generalized delta rule is crucial in justifying the current revival.

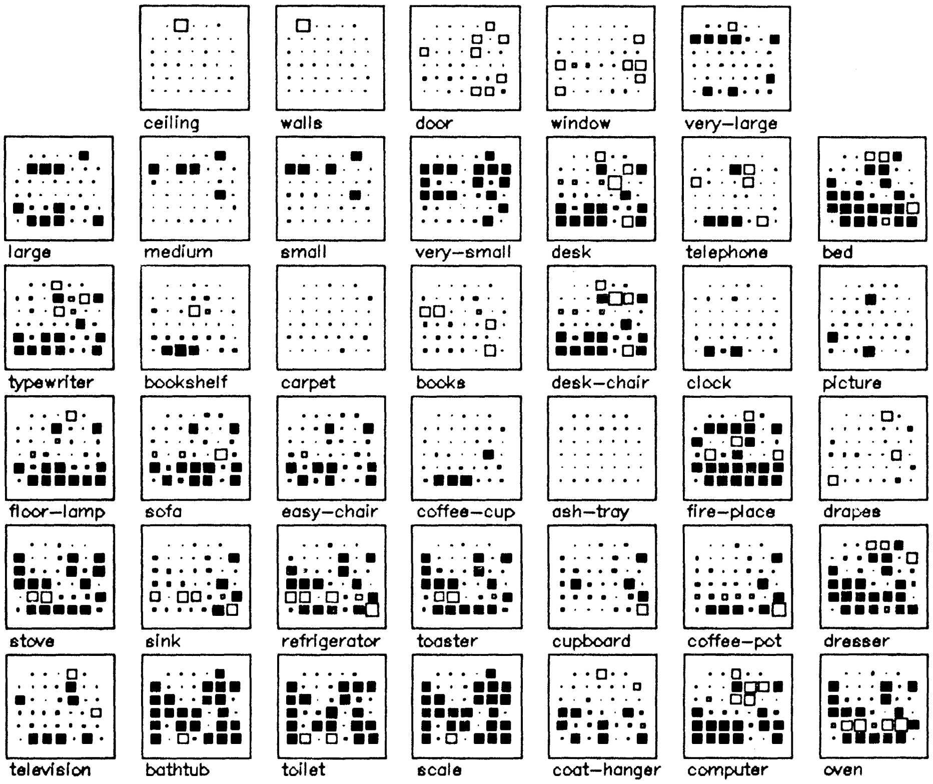
Chapters 6 and 7 respectively describe Smolensky's *harmony theory* and Hinton and Sejnowski's *Boltzmann machine*. These two models, although couched in different terminology, are essentially identical. Each is based on a formal isomorphism between information processing and statistical thermodynamics. Processing consists of maximizing an optimization function based on the consistency of the activation pattern over a network of units that have symmetrical connection weights and binary ("on" or "off") activation levels. For example, if two units have an excitatory connection, consistency is increased if they are both on or off rather than one on and one off. The updating process is probabilistic; the "noise" in the process is at first large and then is progressively reduced. This "simulated annealing" is based on an analogy with the behavior of particles that are heated and then allowed to cool. Annealing serves to alleviate the "local minima" problem that all gradient-descent optimization procedures must contend with (that is, the tendency for the search process to get trapped at some solution that appears optimal relative to similar solutions but is inferior to some other very different solution). Boltzmann machines also have a learning algorithm based on a comparison of the probability that any two connected units are on simultaneously when environmental stimuli are presented with the probability that both units are on when the network is running freely in the absence of environmental inputs. Like the generalized delta rule, the Boltzmann learning rule can modify weights throughout the entire network.

Part 3, in volume 1, includes five chapters dealing with technical aspects of PDP models. Part 4, in volume 2, consists of six chapters that describe applications of PDP models to a rich body of experimental data regarding such cognitive processes as word recognition, speech perception, categorization, the acquisition of verb morphology, and sentence processing. From the point of view of cognitive psychology, this section provides the central evidence that the "brain metaphor" is in fact theoretically fruitful. Chapter 14, the most speculative chapter in this set, addresses an especially controversial aspect of the PDP approach—its appropriateness for describing high-level reasoning and sequential thought processes. Finally, part 5 includes five chapters that deal with the biological relevance of PDP models (plus chapter 22, which more properly belongs in part 3). Chapter 20, by Crick and Asanuma, provides a survey of current knowledge of cerebral anatomy and physiology; Sejnowski's chapter 21 is an intriguing discussion of the kinds of computations that may be performed by the cerebral cortex; and chapters 23, 24, and 25 describe specific models of the coding of place information, neural plasticity, and human amnesia.

As the authors are careful to point out, the attractiveness of the "brain metaphor" underlying PDP models does not derive from the correspondence of the models to any recent breakthroughs in neuroscience. (Indeed, Crick and Asanuma's chapter ends with an interesting list of disanalogies between properties of real neurons and the "units" of the models described in other chapters.) The connectionist models presented here resemble the brain only in extremely general properties (there are many neurons, richly interconnected, that perform simple computations in parallel on a relatively slow time scale). As the authors admit, "The basic idea is that there is a mapping between elements of the model and the brain, but it is unknown and probably only approximate. A single unit may correspond to a neuron, a cluster of neurons, or a conceptual entity related in a complex way to actual neurons" (vol. 2, p. 329)—in short, to virtually anything (perhaps including meaningful "symbols," the connectionists' *bête noire*). Despite the admitted lack of constraint on what units can represent, the book adopts the rather misleading convention of terming their referents "microfeatures." This seems reasonable enough for a unit that represents a low-level construct such as "a vowel preceded by a stop consonant and followed by a nasal" (chapter 18); on the other hand, a microfeature corresponding to a "causal verb" (chapter 19) seems like a conventional semantic feature; and the microfeature "has television" (chapter 14) simply invites ridicule.

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. The connection between two units can be directly translated into a simple rule of the form, "If hypothesis 1 holds, then hypothesis 2 holds," with a strength measure attached to the rule. Summation of activation at each unit serves to integrate multiple sources of converging or contradictory evidence regarding a hypothesis. The various learning algorithms allow revision of the inferential relationships among hypotheses. As this description suggests, many 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.

It is also apparent that the PDP approach does not fundamentally contradict the functionalist approach to cognition. A connectionist model of mind, like any cognitive model, can be described at a level more abstract than any particular implementation. It also seems that the "computer metaphor" for mind is not really being abandoned by the connectionists, but simply modified: the classical von Neumann serial processor is replaced with a processor that performs some simple computation in parallel on many data elements (for example, the "Con-



A representation of the connection weights between units in a parallel network. Each labeled box stands for a unit. "Within each unit, the small black and white squares represent the weights from that unit to each of the other units in the system. The relative position of the small squares within each unit indicates the unit with which that unit is connected." White squares represent positive connections and black squares represent negative connections, with the size of the square representing the strength of the connection. The large, white squares in the "ceiling" and "walls" boxes show that there is a strong, positive weight between the two. [From *Parallel Distributed Processing*, chapter 14.]

nection Machine" recently marketed by Thinking Machines Corporation). The brain metaphor has influenced computers, yielding an updated computer metaphor for mind.

The research reported in the book provides a broad picture of the accomplishments of the PDP approach and of its future promise and problems. The accomplishments are substantial. The models provide simple mechanisms for allowing fragments of memory representations to complete themselves, for producing generalization to similar patterns, and for "graceful degradation" when the system is damaged. These are all fundamental characteristics of human cognition. At a more specific level, the individual psychological models, such as the word recognition model of McClelland and Rumelhart and the speech perception model of McClelland and Elman, provide detailed accounts of fine-grained empirical results. It is hard to imagine any adequate account of the impact of context effects on the interpretation of stimuli that will not embrace the principles of interactive activation embodied in these models. McClelland and Elman's TRACE theory provides an elegant account of the phenomenon of categorical speech perception in terms of the "canonicalizing" effect of feedback from phoneme to feature units.

Not surprisingly, the work is still a long way from providing a full account of human cognition. The book offers a great deal of intriguing and potentially fruitful speculation about how higher-level cognition might be modeled in PDP terms. The discussion of schemas and mental models in chapter 14 is of this nature, as is Smolensky's description in chapter 6 of how physics expertise might develop. The basic conception of mental processing as a form of parallel constraint satisfaction constitutes a rich theoretical framework. In a few cases, however, speculation crosses the border into flights of fancy, as when Hinton and Sejnowski identify the free-running stage of the Boltzmann machine with dreaming (chapter 7), or McClelland and Rumelhart postulate a mysterious "gamma" substance in the brain and endow it with properties required to reconcile PDP models with some of the evidence concerning amnesia (chapter 25).

The greatest challenges facing connectionists concern the adequacy of their learning mechanisms and of their knowledge representations. Each of the three major learning algorithms described in the book is capable of extracting statistical regularities present in patterned inputs, using combinations of hidden units to implicitly form novel feature detectors. These are significant accomplishments; however, it is an open question whether any of the algorithms provides a model for some form of human learning. The chapters in part 4 use only the ungeneralized version of the delta rule (restricted to networks without hidden units) in simulating detailed psychological data. In chapter 8 the generalized version is used to generate solutions to several small-scale relational problems that require hidden units. Its solutions are often nonobvious and constitute impressive demonstrations of feature extraction; however, the solutions generally do not seem "humanlike." Each solution to a relational problem, such as deciding whether a string of binary bits has an odd or even number of *1*'s, is obtained for an input string of some particular fixed length. For example, if strings of four bits are presented, each labeled "even" or "odd," the network will in effect learn to count the number of *1*'s and to associate the values *0*, *2*, and *4* with the response "even" and *1* and *3* with "odd." No investigations of transfer are reported, but clearly this representation provides no information about whether a novel input of *five* bits is even or odd. Unlike the case of a person who has abstracted the relational rule that "even" means "divisible by *2*," the delta rule's solution will never allow transfer to strings of unbounded length.

In addition, all the algorithms are extremely slow, requiring thousands of repetitions to reach solutions to even simple problems. The learning algorithms have so far been shown to support a degree of parallelism more aptly characterized as modest than massive, since learning within networks containing more than hundreds of units has yet to be demonstrated. The difficulty of applying the algorithms to very large networks may reflect inherent limitations of "bottom-up" approaches to learning based solely on gradient-descent methods that adjust the weights of preexisting connections. If a unit receives a large number of inputs, only a few of which actually correlate with the output the unit needs to produce in order to perform a new task, then most of the weight adjustments made on a given trial will fail to improve performance on the new task and may impair performance on tasks previously learned. The PDP learning algorithms fail to exploit potential "top-down" types of information that could restrict the size of the search space in which the weight-adjustment procedures operate (for example, prior knowledge of a specific domain, or general heuristics for identifying plausible causal relations). Organisms from rats to humans sometimes exhibit one-trial learning of relations consistent with their prior "theories" and may fail utterly to detect statistical regularities that violate them. The learning algorithms also have difficulty accounting for sequential behavior. In particular, none has yet demonstrated the capacity to learn tasks involving sequences of actions in which early components receive no direct feedback yet are crucial to ultimate success (for example, a rat learning to negotiate a maze to reach food, or a checkers player setting up a triple-jump). These aspects of biological learning are not well captured by the PDP learning algorithms.

The adequacy of the knowledge representations proposed in the book is also open to question. A general problem with PDP representations is that a great deal of redundancy is required. To process a word, for example, the TRACE model of speech perception assumes that the networks of feature, phoneme, and word units are reduplicated many times over the time segments into which the input signal is divided. In McClelland and Rumelhart's model of visual word recognition, each letter position corresponds to a completely different set of feature detectors. (The implemented model can only recognize words up to four letters long.) This representation, if taken seriously, implies that learning to recognize a particular letter in the second position of words will have no impact on recognizing it in the third position, a prediction that seems dubious. Also, it is quite unclear how such architectures could develop with experience, be extended to inputs of arbitrary length, or (in the case of the TRACE model) be tuned to such crucial idiosyncrasies as speech rate.

The use of distributed representations to represent sentences leads to additional architectural complexity. In a distributed representation, the concept "boy" will be mapped onto many of the same units as is "girl." It is therefore tricky to represent a sentence such as "The boy kissed the girl" without muddling the agent and patient roles. In chapter 19 McClelland and Kawamoto, extending a proposal by Hinton, implement a possible solution to this type of problem that involves defining sets of units representing different semantic roles, such as agent or patient, and including units that respond to conjunctions of features of the concepts filling two different roles. This technique exemplifies one approach to the general problem of representing "what goes with what" in PDP networks, which has yet to be definitively solved.

The representational complexities that the PDP approach must deal with severely exacerbate the shortcomings of the available learning procedures. The learning algorithms are not inductively sufficient; that is, simply modifying connection weights within a large homogeneous initial network on the basis of environmental inputs would not suffice to construct the specialized representations required for different tasks. The performance of a weight-adjustment algorithm generally depends on the prespecified architecture of the network within which it operates (for example, the number of layers, the number of units within each layer, and the overall pattern of connectivity). In all the simulations reported in the book, the network is crafted by the researcher to perform some specific task. As yet there seem to be no principles that constrain network architecture and no proposed learning techniques that might allow induction of modular subnets or other specialized architectures.

Although major hurdles loom ahead, it is clear that the PDP approach so forcefully articulated in this book will have a major impact on cognitive science. Researchers in cognitive psychology, artificial intelligence, parallel computation, neuroscience, linguistics, and other fields as well will find the work immensely stimulating. The greatest value of the book is that it clearly lays out a paradigm and applies it to concrete and interesting examples. The book opens a door that cognitive scientists can enter: some will stay and join the movement, others will steal a few ideas and leave, and others yet will learn why they prefer to stay outside. All will want to take a peek.

KEITH J. HOLYOAK
Department of Psychology, University of California, Los Angeles, CA 90024

Shapes in the Plane

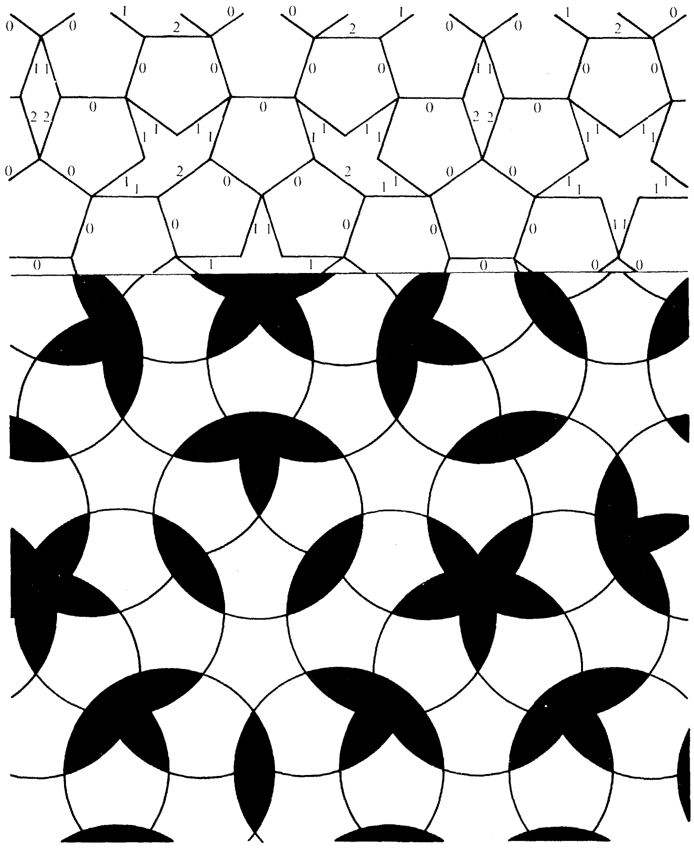
Tilings and Patterns. BRANKO GRÜNBAUM and G. C. SHEPHARD. Freeman, New York, 1986. xii, 700 pp., illus. $59.95.

Throughout history people have filled floors, stained-glass windows, and fabrics with shapes that occupied the plane without holes or overlaps. These shapes, or tiles, often endow plane surfaces with patterns of remarkable symmetry and beauty. Although tilings of the plane have been found in varied contexts and cultures, the systematic study of their types and properties is surprisingly new. Except for a modest beginning in the 17th century by Johannes Kepler, there were few studies of tilings until the 20th century, and not until the research of this book's authors did the study of tilings become a full-fledged subbranch of geometry.

To develop a theory of tilings, one must adhere to the rules concerning which shapes are allowed for tiles and the rules about how shapes can be placed next to one another. For example, if one starts with a supply of identical rectangular tiles, one may place the tiles so that the edges of the rectangles match, producing the familiar tiling of floorings, where four tiles meet at a point. But other tilings by rectangles can occur if we allow the tiles to adjoin in additional ways. Infinitely many types of tiling are possible when the tiles are not laid down edge to edge. Patterns, the other word sharing the book's title with tilings, involve symmetry considerations that arise when motifs of various kinds are systematically located in the plane. Although beautiful patterns have been created and examined by mathematicians in addition to craftsmen and artists for thousands of years, a coherent theory of patterns was not developed until that presented in this book.

What Grünbaum and Shephard have done, in a dazzling display of scholarship, erudition, and research, is collect in one volume a compendium of the accumulated knowledge about tilings and patterns developed by a wide range of individuals including artisans and craftsmen, mathematicians, crystallographers, and physicists. In doing so they were forced to take a fresh look at what was known, what was thought to be known, and what had yet to be investigated, and to provide a framework in which all of this information could be succinctly put down. The project, which was started about 10 years ago and has only now been brought to (partial) completion, is well worth the wait.

To present this material Grünbaum and Shephard needed to develop a standardized terminology. They have chosen terminology that is clear yet flexible and thus well suited to the rich range of phenomena encountered. Since this book begins with the simplest of concepts, it can be started and read with enjoyment by a high-school student. However, the reader needs staying power. As the book progresses, relatively elementary concepts and definitions are developed, but in the building-block style typical of a mathematical theory. Furthermore, the authors have felt free to use elementary ideas from topology, group theory, and number theory, defining the necessary terminology as they go along. Thus, although the treatment is elementary in the sense that all concepts and ideas that are not well known by persons with modest mathematics backgrounds are fully, clearly, and carefully explained, one cannot hope to understand the statements in the middle of the book with-



"A modification by Penrose of his set P1 of aperiodic prototiles. Each edge is replaced by a circular arc whose center is the 'point' of a pentacle or half-pentacle. A small portion of the original P1 tiling is reproduced at the top of the diagram to show the relationship between the tilings. Three of the prototiles have been colored black and three are white. It is conjectured that these 'curvilinear' tiles are also aperiodic." [From *Tilings and Patterns*]