there is Azoulay's translation of Cajal's textbook the *Histologie du Système Nerveux* and the translations of Cajal's original reports on the cerebral cortex recently published by DeFelipe and Jones. For the new students interested in the origins of neuroscience and Cajal's contribution to it, there is Cajal's autobiography, *Recuerdos de mi vida* (Recollections of My Life) which has now been reprinted from an earlier English translation. The autobiography gives a popular and easy to read outline of Cajal's work in relation to the events of his life. But anything Cajal wrote is worth reading. The book is an introduction to the great breadth of Cajal's studies and it has some of his beautiful drawings of the structure of the brain and spinal cord, along with his clear and succinct descriptions of the cell types and their interconnections. The translation is most useful to our appreciation of Cajal's great contribution to modern neural science.

Reprint requests should be sent to Dr. M. Glickstein, Department of Anatomy and Developmental Biology, University College London, Gower Street, London WC1, England.

Unified Theories of Cognition

Allen Newell
Harvard University Press, 1991, 549 pp., $39.95

Review by Richard Granger
University of California, Irvine

The past three decades has witnessed the development of interdisciplinary research encompassing cognitive psychology and artificial intelligence (AI), with the goal of identifying formal computational hypotheses of mind. Allen Newell has been one of the leaders of this research from its inception, and publication of his new book *Unified Theories of Cognition*, based on the author's William James lectures at Harvard University, represents the amassed results of a major research effort in this field. It will be reviewed for its content and clarity in many forums; here instead a brief attempt will be made to address the book, and the line of research it represents, from the perspective of cognitive neuroscience. To this end, the main question that will be raised is how can we identify what functional psychological operations are primary or fundamental, i.e., those that combine to produce complex cognitive phenomena? This is a shared quest of cognitive psychology, artificial intelligence, and cognitive neuroscience; Newell's book offers an opportunity for evaluation of successes and prospects in these fields.

Newell dedicates his book to a specific task: "to urge on psychology unified theories of cognition . . . theories that gain their power by positing a single system of mechanisms that operate together to produce the full range of human cognition" (p. 1). Newell indeed assumes both the existence of a single set of mechanisms and the ability in principle to identify that set. Most of the book is devoted to the presentation of a proposed example of such a theory, consisting of a rich set of coherent hypotheses that purports to explain a wide range of sometimes disparate psychological phenomena, together with a computer system implementing these hypotheses. This computer system, called SOAR, is an artificial intelligence computer program, based on a "production system" architecture, i.e., a set of "If-Then" rules for behavior. This fact raises some immediate controversy: the very architecture of the system contends that all of memory is uniformly composed of these productions or rules. Moreover, it is proposed that this uniform architecture accounts for psychological phenomena ranging from the power law of speedup with practice, to solving "cryptarithmetic" problems: finding appropriate numbers to solve an algebraic addition. Newell is quick to admit that this proposal of unification via a production system is disconcerting in light of the many findings of neuropsychological dissociations and other evidence for multiple forms of memory, but he offers little in the way of amelioration, and does not consider modifying the unified theory to account for such dichotomies. Rather, Newell intends the very unification of his theories to amend what he sees as excessive dichotomizing in psychology: serial versus parallel processing, massed versus distributed practice, procedural versus declarative memory. For Newell, the demonstration of apparent modularities in cognition (being able to point to a thing but not to name it, being able to remember what happened four minutes ago but not ten, being able to learn to read in a mirror but be entirely unaware that any such learning has occurred) does not remove the necessity of "a theory that . . . explains the role of the [modules] and why they exist" (p. 18).

When the object of study is an artifact, understanding it can be aided by prior "top-down" information about its intended operation; thus the performance of David Marr's cash register (*Vision*, 1982) can be addressed with respect to its intended operation: adherence to the Peano axioms of arithmetic [adding something and then taking it away should result in zero (law of inverses), adding two things individually or as a group gives the same result (associativity), and so on]. Cognitive psychology attempts to apply such means to the behavior of natural systems; the complex meanders of Herbert Simon's hypothetical ant on the beach (*The Sciences of the Artificial*, 1969) can be greatly simplified by reference to the interactions of two factors: the ant's intended (internal) target and the (external) obstacles in its path. Attribution to the ant of a specified set of intentions enables a computational-level analysis of its activity. In the absence of knowing the "intentions" of the object, we can know only measured behaviors (as in the standard "behaviorist" argument), including, of course, measured neural responses. We nonetheless wish to know what mental, psychological, cognitive, operations are occurring: we wish to study thinking, not only behavior per se. In the absence of powerful neurobiological data to bring to bear on these points, psychology has proceeded to attempt the identification of mental operations predominantly from psychological experimentation. Newell suggests a number of constraints, with particular attention to timing and latency, that arise from the characteristics of neurons and networks (see especially Chapter 3), and incorporates them into the theoretical framework of SOAR. However, he also suggests that more theory, rather than more data, is what is most needed: "My bet is that enough constraining regularities already exist in the literature to identify the architecture to all intents and purposes. We could stop doing experiments entirely and, just from what now exists, discover what the architecture is" (p. 244). In this regard, psychology is unique among the sciences for arguing that the details of the physical object of study, in this case, the brain, may not be centrally needed in furthering the goals of the science.

It has in fact been argued that there are many questions that cannot in principle be answered using psychological and behavioral experiments alone, such as whether a process is serial or parallel, or what type of representation is used in any given cognitive task (Townsend, 1974; Anderson, 1978). John Anderson, professor of psychology at Newell's own institution, Carnegie-Mellon University, has been another primary purveyor of the type of unified theory advocated by Newell; Anderson is author of, among many other works, *The Architecture of Cognition* (1983), and the ACT system, which shares many goals with Newell's SOAR system. Anderson has stated that in light of the identifiability problem, all proposed models of cognitive mechanisms must be regarded merely as notations, without deeper explanatory power (Anderson, 1989). Newell dismisses these views as "misplaced," and proposes that "the way to solve the identifiability problem is by bringing to bear large, diverse collections of quasi-independent sources of knowledge—structural, behavioral, functional—that finally allow the system to be pinned down" (p. 22). Newell hypothesizes a single set of underlying mechanisms unifying the diverse domains, but let us for the moment assume an alternative that we might term the "dissociation" hypothesis: many distinct brain circuits exist, each operating relatively independently of the others, and each giving rise to distinct portions of the vast range of psychological and behavioral data that Newell hopes to unify. If brain organization turns out to have this nonunified character, it will be difficult to argue that mental organization does not share it. Newell argues that unified psychological theories are now possible, but that does not mean that they will be desirable.

This brings us back to the question raised at the outset: how can we identify fundamental psychological operations and their interactions? Cognitive neuroscience addresses this issue by studying specific anatomically defined brain areas in an attempt to identify their emergent functionality. Arguably, functions that arise directly from the normal operation of primary brain circuits will be fundamental: they are the "building block" operations from which other operations are constructed. Newell objects to arguments based on purely Popperian falsifiability of theories, yet biologically identified operations will be fundamental not just in terms of falsifiability, but also in terms of predictive power and, perhaps most importantly, explanatory power: by showing us what is primary and what composite, what is fundamental and what derived, such findings may help inform us why we do what we do; how it is that we are the way we are. It is by success at answering these central questions about our mental lives that cognitive psychology and cognitive neuroscience ultimately will be measured. Allen Newell has never shied away from attacking these deep questions, and *Unified Theories of Cognition* represents one of the largest coordinated efforts ever undertaken in cognitive psychology in pursuit of this goal; it is an important landmark, and an exceedingly rich and provocative one.

Reprint requests should be sent to Dr. Richard Granger, Center for the Neurobiology of Learning and Memory, Bonney Center, University of California, Irvine, CA 92717

REFERENCES

Anderson, J. R. (1978). Argument concerning representations for mental imagery. *Psychological Review, 85,* 249–277.

Anderson, J. R. (1983). The Architecture of Cognition. Cambridge, Mass: Harvard Univ. Press.

Anderson, J. R. (1990). The place of cognitive architectures in a rational analysis. In: Architectures for Intelligence (K. VanLehn, Ed.), Hillsdale, NJ: Erlbaum.

Marr, David (1982). Vision: a computational investigation into the human representation and processing of visual information. San Francisco: W. H. Freeman.

Simon, H. A. (1969). The Sciences of the Artificial. Cambridge, Mass: MIT Press.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In: Human Information Processing: Tutorials in Performance and Cognition (B. H. Kantowitz, Ed.), Hillsdale, NJ: Erlbaum.