tion. I suppose we could highlight problems that are not solved for people to look at if they are shopping around for interesting research problems. Maybe we need more articles in the *American Psychologist* to bring to the attention of the academic world some of the clinical issues for which we really could use more understanding.

Cf Weimer & Palermo #1; notes from Penn State Conference

4

The Cognitive Revolution: The Rise of a Theoretical Psychology

Science is built of facts the way a house is built of bricks, but an accumulation of facts is no more science than a pile of bricks is a house.

—HENRI POINCARÉ (1913, p. 12)

Between 1955 and 1965 a quiet revolution in thought took place in scientific psychology. Even before 1955 there were many hints of this development, and much remained to be done after 1965, but for most opinion leaders in the field, this decade was crucial. The cognitive revolution was not a spectacular one—no one stormed the Winter Palace, not even metaphorically. Public fireworks were rare. Furthermore, unlike Watson's revolution, the cognitive shift was not self-conscious. No one announced its existence until long after the fact. Experimental psychologists did not set out to make a revolution. Many of them were surprised to find themselves breaking the prohibitions of the behavioristic metatheory, and did so reluctantly. The understanding of the cognitive revolution followed the event; it only emerged slowly, and even now there are some who will not agree that any "revolution" took place. Only now, in the 1980s, are the historical retrospectives being written (*e.g.*, Bruner, 1984; Kendler, 1979).

Most of the changes occurred silently. There was one major public controversy following Noam Chomsky's critical review of B. F. Skinner's work *Verbal Behavior* (Skinner, 1957; Chomsky, 1959); and some years later, public debates took place between these two men. But in 1957, Chomsky was still a little-known linguist, and Skinner, a respected but atypical behaviorist. Most of the milestone experiments in the cognitive revolution were published without fanfare, and changes were occurring even as their profundity was being denied. Many such changes seemed to proceed by a process of euphemism. Psychologists did not speak of "mental representation" at first, but of "memory"; not of "consciousness," but of "selective attention"; not of "the organization of meaning," but of "semantic features." In each case, the modest euphemistic term was defined operationally, by precise and reliable experiments, and the results were interpreted within narrow theoretical limits. But as the new ideas gained momentum, theoretical terms such as "mental representation," "meaning organization," and recently, even "consciousness" burst the boundaries of the experimental situations in which they were first defined. Soon, all these terms were extended far beyond any single experimental model or technique.

"Protocognitive" Psychologies

We have noted that the psychology of common sense requires a cognitive metatheory; in this respect the cognitive approach is hardly new. Most formal schools of psychology are metatheoretically cognitive, and their existence may have helped to bring about the cognitive revolution. This includes psychoanalysis and the various schools that derive from it; Gestalt psychology and its intellectual descendants in the study of perception; and the whole field of social psychology, started in America by the refugee Gestalt psychologist Kurt Lewin. The applied field of psychometrics began much earlier, and produced numerous tests of intelligence, aptitude, and personality characteristics while coexisting uneasily with academic behaviorism. To most behaviorists, such measures had no scientific standing, because the very constructs "intelligence" and "personality" were undefined. Among the cognitive viewpoints in Europe we must mention Jean Piaget's epochal studies of child development, the work of Selz and de Groot on problem solving and chess playing, and Bartlett's studies of memory. In the Soviet Union, Pavlov's followers, such as Y. Sokolov, began increasingly to talk in cognitive terms, and the earlier work of Y. Vygotsky and A. R. Luria has a distinctly cognitive flavor. In the 1940s Soviet cybernetics began to keep pace with Western developments, and psychologists such as N. Bernstein created a cognitive movement using cybernetic concepts. But there was astonishingly little communication between American and European psychologists, who were separated not only by language barriers, but by formidable differences in outlook. Even when European schools of psychology were transplanted to America, there was remarkably little visible communication.

In this connection we may also mention *humanistic psychology*, an American school of thought that has affected clinical practice, self-help techniques, and what may be called "recreational psychology." In all these aspects of popular psychology humanistic thought has been very influential. Humanism was originally conceived as a response to the pessimism of Freudian thought, and to the mechanistic quality of academic behaviorism (*e.g.*, Maslow, 1962; Rogers, 1970). Its orientation has been both optimistic and antimechanistic. On the other hand, though a native American product, its influence on American scientific psychology has been nearly zero. Humanism has flowered outside of academic psychology, but scientific psychologists in their official roles tend to see it as unscientific, and therefore outside of their sphere. What influence humanism has exerted has been very indirect (see, for example, the interview with Ulric Neisser in Chapter 6).

The Triumph of Behaviorism in the Clinic

Human history seems to abound with irony and paradox, and the history of psychology is no exception. One of the ironies is that between 1955 and 1965, just as cognitive psychology was emerging from the remnants of behaviorism in experimental psychology, exactly the *opposite* was happening in the applied area of clinical psychology! Psychopathology was largely neglected during the first three decades of behaviorism (though John B. Watson and others performed pioneering studies on phobias and habit-control). Instead, clinical work was dominated by psychodynamic theories, which presuppose a cognitive metatheory, postulating invisible entities such as wishes, fears, conflicts, and the like. As discussed in Chapter 2, during the 1950s behavioristic psychologists began to make a decisive move toward clinical psychology (Kazdin, 1978). These behavioristic clinicians were deeply critical of psychoanalysis. In arguing against psychodynamics, they made use of the classic behavioristic arguments against mentalism, concluding that psychodynamics was inherently unscientific. Like their forebears in experimental psychology, clinical behaviorists were determined to eliminate unobservable constructs (*e.g.*, Ullmann & Krasnèr, 1969).

The spread of behaviorism into the clinic had little effect on the "pure science" issues discussed in experimental psychology. Experimentalists continued to shift to a cognitive perspective, undeterred by the fact that their clinical colleagues next door were shifting the opposite way. For better or worse, experimental psychology is largely insulated from applied fields—the direction of influence tends to go from pure to applied psychology, rather than the reverse.

Today, the cognitive shift in experimental psychology seems to be reaching the clinical world, but in greatly modified form (*e.g.*, Mahoney & Arnkoff, 1978). We cannot go into these very interesting and important historical developments, other than to note their existence. Thus we return to the story of the revolution in mainstream experimental psychology.

The Role of the Theoretical Imagination

We have defined the cognitive metatheory as a belief that psychology studies behavior in order to infer unobservable explanatory constructs, such as "memory," "attention," and "meaning." A psychological theory is a network of such constructs, serving to summarize empirical observations, predict new results, and explain them in an economical way. Like behaviorism, cognitive psychology is primarily a metatheory for psychology, one that simply *encourages psychologists to do theory, relatively free from prior philosophical constraints.* No longer is it thought necessary for theoretical constructs to resemble visible stimuli and responses, or to adhere to rigid conceptions of theoretical parsimony. As simple as this change may seem, its impact on the practice of scientific psychology is pervasive. Cognitive psychologists do not assume that learning is necessarily a matter of associating observable stimuli and responses, that conditioning is a fundamental form of learning, or that inherited factors are of minor importance in human action and experience. Nor do they assume that consciousness is merely a byproduct of the ordinary, physical functioning of the nervous system. One of our interview participants calls this "the liberation"—permission to use one's theoretical imagination free from prior constraints.

It is not unscientific to speak of the "theoretical imagination." In one way or another, scientists have always employed imagination to account for their observations: In Dalton's time, the atom was almost entirely an imaginative construction; Darwin imagined the continuity of descent between the species; and Newton had to imagine gravitational forces emanating from the sun to keep the planets in orbit. A witty remark attributed to the prominent physicist Richard Feynman makes this point about the centripetal forces of Newton's theory: "Before Newton," Feynman has reportedly said, "people thought that angels were pushing the planets in orbit around the sun. But now we know that the angels are pushing the planets *into* the sun!" In their way, the centripetal force vectors of Newtonian physics are as imaginary as angels; but they are scientifically useful and perhaps indispensable (Jammer, 1957).

Indeed, behaviorism itself was fundamentally an act of theoretical imagination, one that had psychologists for 50 years pretending in their professional lives that goals, memories, and images did not exist, even while they freely used commonsense psychology in their private lives. Behaviorism was a *metatheoretical* act of imagination, but one that prohibited the use of the theoretical imagination in the actual practice of scientific psychology. By the same token cognitive psychology is an act of imagination that permits wider latitude in imagining explanations for behavior. Whereas behaviorism taught psychologists to respect public empirical evidence, the cognitive metatheory may make it possible to do good theory. This involves a learning process, of course. The tools needed to do theory do not spring up instantly, any more than the tools for experimentation did after Watson's revolution. A new metatheory can clear the way, but the new opening must be exploited by hard, thoughtful work.

Consider what would happen if scientists were not allowed to use their imagination freely—if they could not imagine that some undiscovered fact were true, or that some inconvenient observation were false. Galileo's work is an interesting case in point, given that he has often been considered a model of the hardheaded empiricist, who presented plain facts that his theological opponents literally refused to see. Some modern historians of science view Galileo's contribution somewhat differently. Apparently, Galileo's facts were not quite as solid as subsequent generations have thought, and his opponents themselves could cite numerous facts in their favor (Feyerabend, 1975, and others). Many of Galileo's arguments for the Copernican solar system were actually quite false, but his overall conclusions were correct. In this light it becomes clear that Galileo's claims really involved a daring act of imagination. Conclusive empirical evidence in his favor only became available more than a century after his death.

Now imagine what might have happened if tough-minded experimentalists had dominated the physics of Galileo's time in the same way that behaviorists dominated psychology in this century. Consider Galileo's famous (though probably apocryphal) experiment of simultaneously dropping a musket ball and a cannon ball from the Leaning Tower of Pisa, and noting that both objects hit the ground at the same moment. This presumably showed that the mass of an object does not affect its rate of acceleration in gravity. But a good experimentalist would quickly refute Galileo's demonstration by dropping from the Tower a feather, a hot-air balloon, and a musket ball. Now, which would fall fastest? Obviously, the feather would encounter more resistance from the air than the musket ball, while the balloon would actually float upward! Is this an unfair experiment, because it neglects the effects of air resistance and buoyancy? Not really; even to conceive of air, and of its possible absence, required in Galileo's time a profound act of imagination, a leap that scientists would not be able to justify until several centuries later. The whole point of Galileo's experiment depends on our imaginative ability to pretend that air does not exist, and that the effect of gravity can be considered separately from the effects of air.

Without this imaginative leap, an entire experimental profession might have developed, collecting a multitude of facts about falling objects. Good experimentalists would not stop with testing feathers and hot-air balloons. As conscientious believers in empirical fact, they would perforce investigate other factors: the wind-speed and ambient temperature, the weather and the time of day. Soon they would have collected enormous numbers of facts, but without gaining any *insight* into the problem of gravity. Facts alone do not make a science; indeed, thoughtless fact-gathering can interfere with the work of science. It has now become commonplace in the philosophy of science to say that "the facts" cannot even be perceived *as* facts without some theoretical framework, explicit or not (*e.g.*, Kuhn, 1962, 1970; Lakatoš & Musgrave, 1970).

The Computational Metaphor and the Role of Experiments

But why were psychologists granted more theoretical freedom in the 1960s and '70s, when a few decades before, a psychologist such as Tolman was considered unscientific for talking about "purpose" and "cognitive maps"? Answers to this question must ultimately come from our interviews, but we may suggest two general reasons for the liberalization of theory: First, developments associated with the theory of computation led some psychologists and neurophysiologists to view the nervous system as a kind of information processor, a theoretical metaphor that made it legitimate to think in terms of goals and representations. Second, the experimental methodology developed by behavioristic psychologists provided proof-procedures whereby nascent cognitive thinkers could make compelling arguments for ideas such as "attention," "imagery," "mental representation," "unconscious inference," "goals," and the like. Computational theory provided a guarantee that the theoretical imagination did not exceed the bounds of physical possibility, and experimental demonstrations made a compelling case that computational ideas could apply to human beings.

The cognitive revolution took place in many places at the same time, and involved a number of areas, including memory, language, imagery, and attention. In each case, careful and reliable experiments forced psychologists to adopt a more theoretical stance than was permitted before. This chapter sketches these transformations in several areas, including mental representation, attention and consciousness, the problem of serial order, and the problem of inferring underlying structure and process. Space permits no more than a sketch of these events, however. More details may be found in any modern textbook of perception, human learning, or cognitive psychology, such as Norman (1976), Bransford (1979), Anderson (1980), and especially Neisser (1967). Other important perspectives on this historic metamorphosis are presented by Posner and Shulman (1979), Newell and Simon (1972), and Bruner (1984).

Behaviorists in general did not contest these developments publically, perhaps because they emerged clothed in the guise of simple, unadorned facts, using the accepted proof-procedures that the behaviorists themselves had pioneered. This chapter discusses only a few instances of public controversy. Finally, the chapter briefly traces the careers of five psychologists who figured prominently in the revolution: Charles E. Osgood, a neobehaviorist who tried to extend Hullian theory to encompass the concept of linguistic "meaning"; James J. Jenkins, who at first tried to pursue a behavioristic approach to verbal learning, failed, and then became a pioneer in exploring cognitive alternatives; George A. Miller, the most visible leader of the cognitive revolution, who exerted great influence by always staying one step ahead of the rest of the field; Jerome S. Bruner, less concerned with communicating with other experimental psychologists, and whose work was consistently far ahead of the rest; and Herbert A. Simon, whose work of the 1950s is only now entering the mainstream of cognitive psychology. In terms of their adherence to the behavioristic perspective, these men represent a spectrum ranging from the most conservative (Osgood) to the least (Simon), with the others ranged in between. Of course, "conservative" does not mean either "good" or "bad" in science; it is the interaction between adventurous souls and their hard-nosed colleagues that provides scientific work with the necessary mixture of daring and certainty, and a successful science will encourage the whole spectrum of scientists to thrive.

The Computational Rationale for the Cognitive Revolution

The persuasive mechanism of the cognitive shift was empirical rather than theoretical. Psychologists used experimental evidence to persuade themselves and each other of the need for change. It may seem peculiar, therefore, to start the story of this historic transformation with issues usually associated with the foundations of computer science. Nevertheless, it is important to be clear about these ideas. First, developments in computational theory led to the invention of the computer, which became the dominant "machine metaphor" of the cognitive shift. No one seriously maintains that humans resemble digital computers, but the nervous system needs to solve many of the same problems that must be solved by computers in performing similar tasks. Second, it is important to understand computational theory because it applies not merely to contemporary computational hardware; rather, it specifies mathematical principles that apply to an infinite class of symbolic devices. If nervous systems are specially adapted to represent and symbolically transform the world of the organism, then the abstract principles of symbol-manipulation must also apply to it. As a full-fledged realization this idea is rather recent in cognitive psychology; during most of the cognitive revolution it certainly was not accepted. Nevertheless, it is the most cogent scientific rationale for the revolution available today, and we may speculate that some unstated intuition like this may have played a part in the historical developments. Of course, the public proof-procedures used by psychologists were always empirical; but underlying these experiments were some strong commonsense intuitions about human beings. When these intuitions are rigorously analyzed, one arrives at something very much like a computational conception of human psychology (*e.g.*, Fodor, 1968). We will discuss the computational rationale of the cognitive shift before describing the actual story of the shift.

Information, Automata, and the Foundations of Mathematics

I am using the term "computational theory" very broadly to refer to the cluster of mathematical ideas that led up to the modern computer, its programs, and its higher-level languages. Some of these ideas go clear back to Aristotle's logic and to the set theory developed by George Boole in the 19th century. Russell and Whitehead's symbolic logic (1910) played a role, as did the 19th-century movement toward specifying the foundations of mathematics. About the beginning of this century, it became clear that any mathematical system can be expressed in symbolic logic, and that logic itself was largely reducible to Boolean set theory. When mathematicians and engineers discovered that electrical circuits could be modeled with Boolean set theory, it was only a matter of time before they realized that electrical circuits could also then *represent and compute* Boolean equations. But if electrical circuits could compute Boolean equations, they could also calculate symbolic logic expressions, and thence any mathematical formulae whatsoever. Furthermore, developments in mathematics made it clear that mathematics was not limited to numerical expressions—any symbolic pattern could be an object of mathematical analysis. In summary, these developments in the foundations of mathematics implied that electrical circuits could "act out" any precise set of instructions, whether they involved logic, arithmetic, or even natural language.

The Concept of Representation

"Representation" seems to be a commonsense idea. Ordinary words such as "knowledge," "thought," "feeling," "plan," and "description" express the idea that we can somehow model aspects of the world in our minds and in our speech. Like other commonsense ideas, the concept of "representation" is a good deal more subtle than we may suppose. It raises considerable philosophical difficulties. In philosophy there is a large technical literature on these difficulties, which I can only mention here. (For instance, some philosophers maintain that the mind–body problem emerges whenever one entity represents another.)

Certainly in behaviorism and logical positivism there was great resistance to the idea that people can represent things about the world "in their heads." The concept of an internal representation was often ridiculed with a *reductio ad absurdum* argument, the so-called "homunculus theory." This argument, which dates back to Aristotle, starts by supposing that people represent an object (say, a chair) internally. How does the nervous system deal with such a representation? Surely, the argument goes, there must be something that inspects the representation of the chair, or acts upon it. This "something" is just like a little inner man, a homunculus. But how does the homunculus know what to do with this information? Surely it needs to represent the representation somehow. But *that* representation in turn needs some little homunculus to act upon it, *etc.*, *ad infinitum*. Thus the idea of mental representation seemed inevitably to lead to an infinite regress.

Today no one can maintain this argument anymore. Computers have become so commonplace, and the business of representing some part of the world in a computer code has in many instances become so trivial that the homunculus theory has itself become absurd. The existence of the computer provides a concrete proof that the commonsense notion 7 of representation is indeed viable, even with all its attendant philosophical difficulties.

Mathematical Machines

Starting in the 1930s the English mathematician Alan Turing studied a remarkably simple, imaginary machine, consisting merely of a tape, and a head that could make marks on the tape, read them, and erase them (Turing, 1950). Suppose that such a simple machine had an unlimited amount of time and tape to work with—are there any limits to the mathematical functions it could compute? Turing proved that even this simple machine can compute any mathematical function whatsoever. This means that any device which is formally equivalent to Turing's machine will also be able to model any imaginable relationship that can be stated explicitly. An electronic digital computer is really just a large, fast Turing Machine, able in principle to compute *any conceivable symbolic relationship*. But of course, unlike Turing's imaginary machine, a physical computer has only finite time and memory.

In the 1940s, spurred on by the World War and its demands for automatically controlled radar-tracking devices, mathematicians and engineers developed various kinds of computers, as well as formal theories for describing information and cybernetic feedback. *Information* was at first defined mainly for the purpose of measuring the carrying capacity of telephone lines and the like (Shannon & Weaver, 1949), but this precise definition is directly relevant to fundamental questions in physics and biology. The mathematical definition of information is simple but profound. Basically, it states that information exists when a signal reaches a receiver, enabling the receiver to make a *choice* within a set of alternatives. One might recall in this respect the game of Twenty Questions, in which an answer to a question provides real information only when it permits a player to reduce the number of possible alternatives. A player may get many answers that provide no real information at all. In physics, the very basic notion of entropy—the tendency of energy to dissipate—is defined simply as the inverse of information, and in biology, the genetic rôle of DNA can be expressed in terms of information transmission. Thus, information is a foundation concept in the fundamental sciences.

The smallest conceivable unit of information is one that enables the receiver to discriminate between two alternatives, and this amount of information is called a "bit." Any larger amount of information can be expressed in bits, because any number can be restated to a numerical

base of two. Thus a message may be represented as either + or −, where + could mean, for example, "go to war," and −, "make peace." If there are four choices, say, "go to war," "make peace," "exchange ambassadors," and "protest to the U.N.," the alternatives can still be denoted with only two bits of information (− −, + +, − +, + −); and so on. All the axioms of Boolean algebra can be represented by a small number of bits, and only a few more are needed to symbolize the fundamental axioms of symbolic logic. Once this is done, any other mathematical symbol-system can be easily represented in terms of bits of information, and any physical system able to represent strings of pluses and minuses can then be used as a medium, able to "act out" the operations of any mathematics whatsoever. Thus, a computer can act out arithmetic operations, but also logical deductions, analyses of syntax, representations of objects in space, or control of the movements of a robot, as long as these events can be represented in terms of some explicit set of symbols and operations. Current computers use electrical switching circuits to do such things, with "switch open" versus, "switch closed" to represent plus or minus. But the same information may be equally well represented by magnetized pieces of metal, the position of the beads of an abacus, the position of gears in a mechanical calculator, or any other convenient physical system, including neurons. From the very beginnings of information theory, it was clearly understood that information could be defined *free of any particular physical medium of representation*.

In this way, computers are at once physical and nonphysical. They are physical machines. But the use and significance of computers is not in the switching circuits as such, but in the fact that these circuits can flexibly encode information. A certain piece of information may start as a mark on paper, be transferred mechanically to a hole in a paper tape, to a magnetized bubble, to an electrical pulse running down a wire, or to a radio signal echoing off the earth's ionosphere; it may be stored indefinitely as a magnetized region on a metal disk, and finally end up as a beam of electrons activating the phosphor on the screen of a computer terminal. All these media are different from one another, but the user is not interested in the physical medium, only in the outcome: whether the message is "make peace" or "go to war." In this sense, information is not precisely physical; *as information*, it exists only with reference to the set of alternatives held by the receiver (Bakan, 1980; Szilard, 1929/1964). And it did not take long before engineers and mathematicians began to think that human beings might also be viewed as bearers and transmitters of information.

Human Beings as Information Processors

Human beings in general show a similar disregard for the medium of information, and a decided preference for information itself. Most of the time, when we hear someone speak a sentence, we will forget the exact words within a matter of seconds (try to recall the first sentence of this paragraph). But we will usually be able to *paraphrase* the forgotten sentence. A paraphrase is of course another sentence that preserves the meaning but not the form of the original. The "same" sentence can be spoken in a man's bass voice or a child's falsetto; it can be sung, printed, handwritten, or tapped out in Morse Code; between long-time friends, meaning can be conveyed by a look, a gesture, or even by an unexpected silence. In all these cases it is *not* the physical event that is important, but only the information it conveys.

Further, people tend to confuse two physically different events that bear the same information. Physically different sounds that signal the same phoneme of a language may be perceptually indistinguishable. And in memory, we tend to confuse physically different sentences if they carry similar meaning (*e.g.*, Bransford & Franks, 1971). But adults rarely confuse physically similar sentences if they *differ* in meaning.

The nervous system itself is responsive to information rather than to physical energy as such. Even a single neuron, if it is stimulated by a certain electrical pulse train, will adapt to that *particular* stimulating frequency, but not to any other. That is, it will soon stop responding to the incoming stimulus, not because it is "fatigued," but only because the repeated input conveys no new information. We can show that the adapted neuron will still respond to a *new* stimulus by stimulating it with a different electrical frequency, which will make it respond again. Thus it seems to "recognize" the old stimulus and distinguish it from any new stimulus. In informational terms, a repetitive stimulus becomes redundant to the neuron—it carries no more information.

At a far more complex level of neural integration, every sensory tract in the nervous system adapts to redundant stimulation, so that incoming energy patterns that carry no new information are ignored. And at the highest level of integration, people lose awareness of predictable sensations, of well-learned automatic actions, and even of the meaning of repeated words. All these neural systems begin to activate again when new information becomes available. In general, mere physical input does not usually provide the nervous system with information, and without information the nervous system does not respond. A stimulus must convey some news, some signal that is not redundant (Bateson, 1979; Sokolov, 1963).

In the 1950s, it seemed natural for psychologists and neurophysiologists to investigate how information theory could be used to understand human beings (*e.g.*, Miller, 1964; Garner, 1962). Less than a decade previously, cybernetic theorists had provided a formal proof that normal physical systems could behave in a goal-directed fashion, finally resolving the controversy over "teleology" that had been debated since Aristotle (Rosenblueth, Wiener, & Bigelow, 1943). The "problem" of purpose, which had shaped the theories of Hull, Tolman, and Skinner, was shown to be no problem at all. Meanwhile, Turing's imaginary machine had led other mathematicians to work out a theory of such mathematical machines, and this "theory of automata" soon led Noam Chomsky to a formal proof of the inadequacy of existing behavioristic theories of language (Chomsky, 1957, 1959).

Already by the late 1940s, the conceptual apparatus existed to provide a formal justification for cognitive psychology. An engineer who wanted to discover what a cybernetic machine was doing, or what program was being run on a computer, would not look only at the inputs and outputs of the machine (the stimuli and responses), but would attempt to describe its *functional* innards—does it have the goal of maintaining a certain constant temperature? Does it compensate for changes in a target that it is tracking? Does it compute a theorem in formal logic? These questions are best described not by reference to the physical states of the machine, but to the way in which it functions—a higher-level of description. In fact, the engineer's description of an information machine begins to sound just like our commonsense way of describing a human being: It is stated in terms of goals, knowledge of the world, decisions, predicted gains and losses, and sometimes even intelligent strategies.

Levels of Reality in the Computer

There are a number of legitimate and useful ways of analyzing the functioning of any computer (Newell, 1981): At the physical level one can study the machine as such (the *device* level), and somewhat more abstractly, one can view its elements as electrical circuits (the *circuit* level); as memories with transfers between them (the *register–transfer* level); in terms of the program (the *symbolic* level, the most familiar one); and even further, in terms of the system architecture (the *configuration* level). According to Newell (1981), Each level is defined in two ways. First, it can be defined autonomously, without reference to any other level. To an amazing degree, programmers need not know logic circuits, logic designers need not know electrical circuits, managers can operate at the configuration level with no knowledge of programming, and so forth. Second, each level can be reduced to the level below. . . .

Thus each level can be treated as functionally autonomous *and also* as reducible to the next lower level.

If this is generally true of biological information processors as well, then one can see how a similar relationship might hold between psychology and neurophysiology—psychologists would seem to operate at the level of programs and system architecture, while neurophysiologists work with neurons, neural pathways, centers, and the like. Although this is a bit simplistic, it suggests one way to view the relationships between psychological concerns and the underlying physical system. In principle, psychological functions might also be treated autonomously, with emphasis on the unique properties of psychological regularities, without in any way denying the possibility of reducing the psychological level to neurophysiology. That reduction may be possible, but it is not necessarily useful (Fodor, 1968). To oversimplify a bit, psychologists may be primarily concerned with the "programming" of the human nervous system, rather than with the lower-level hardware.

"Machine metaphors" are not new to psychologists. Even in the 17th century, French thinkers were fascinated by the lifelike robots that were built for the amusement of the aristocracy, and Descartes himself suggested that the human body was just such an automaton, interacting with the divine soul by way of the pineal gland. To Descartes, the soul was the seat of reason, and reason was *not* a mechanical affair. But several centuries later, behaviorists such as Watson and Thorndike thought of learning in terms of a telephone switchboard, as a mechanical connecting of stimuli and responses. Sigmund Freud, probably the most creative theoretical psychologist to date, employed numerous mechanical metaphors to describe the "vicissitudes of the libido" in its various transformations, though Freud acknowledged these as merely metaphors, to be exchanged for more precise neurological descriptions when they became available (Erdelyi, 1985). Machine metaphors are common elsewhere in science as well. For example, the Newtonian solar system looks much like the elaborate clockwork mechanisms that were popular in Newton's time, and indeed, mechanical models inspired by Newton's theory were soon made. For psychologists, always looking for reassurance that their theories refer to some physically possible reality, the lure of machine metaphors is well-nigh irresistible.

Psychological Resistance to the Computational Metaphor

Nevertheless, the new computational ideas were not welcomed with open arms. In the early 1950s, George A. Miller inspired a number of psychologists to investigate information theory as a way of looking at human beings (Miller, 1956), but the results were never completely satisfactory. Behavioristic metatheory still required that the elements of the information system (the sender, receiver, and information-channel) be observable. Thus the stimulus could be treated as a sender, the response as the receiver, and the human subject as the channel. With some exceptions, this application of information theory was unproductive. In retrospect, it seems that information theory posed a challenge that psychologists were unable to meet at the time: If a human being processes incoming stimuli as information, the question suggested by information theory is, "What is the context of alternatives that the information helps to reduce?" This is equivalent to asking what representations people maintain of the world around them, and this question could not be tackled squarely until some years later. The mathematical theory of information presupposed that the context of alternatives was already known; but this is precisely what psychologists needed to discover.

Between 1946 and 1953, the Macy Foundation sponsored a number of meetings among mathematicians, psychologists and other social scientists, and neurobiologists, to explore the emerging parallels between machines and living organisms (Heims, 1975). The regular participants included anthropologists such as Gregory Bateson and Margaret Mead; mathematicians such as Norbert Wiener (the major founder of cybernetics), John von Neumann, Warren McCulloch, Walter Pitts, Julian Bigelow, and Arturo Rosenblueth; Gestalt psychologists such as Kurt Lewin, Wolfgang Köhler, and Heinrich Klüver; sociologists and social psychologists such as Paul Lazarsfeld and Alex Bavelas; neurophysiologists Hans Lukas Teuber and Ralph Gerard; and psychoanalysts such as Lawrence Kubie, and briefly, Erik Erickson. Few, if any, behaviorists attended the Macy conferences. The results were apparently mixed. A number of the leading thinkers in each field attended the conferences repeatedly, and we may suppose that they found the meetings worth their while, but the impact of the new computational theories did not seem to affect the scientific work of most participants (but see McCulloch & Pitts, 1943; Bateson, 1979).

But in 1956, only three years after the last Macy Conference, another conference was held in which all of the core ideas of the cognitive revolution were already contained. The participants included George A.

Miller, soon to become for several decades the most identifiable leader in the cognitive movement; Noam Chomsky, who was about to initiate a cognitive shift in linguistics that would profoundly influence many psychologists; David Green and John Swets, who were applying an engineering theory to human psychophysics, destined to have major impact; and perhaps most fatefully, Allan Newell and Herbert A. Simon, who were then helping to develop the first higher-level computer languages: They would soon be using these new languages to model human performance. Within the next few years, Newell and Simon showed that computers could be programmed to play chess, prove sophisticated logic theorems, and even simulate some important properties of human short-term memory (Newell & Simon, 1972). Over the next several decades, the number of simulations of human intelligence increased dramatically, and the resulting field of Artificial Intelligence is today having increasing impact in experimental psychology. But this is getting ahead of the story. The seeds were there in 1956, but any widespread impact on psychology was very much in the future.

The rapid advances in computational theory had a pervasive *but indirect* impact on the course of scientific psychology. Even today, most psychologists know little about the mathematical arguments originating from information theory, cybernetics, or the theory of automata. Contemporary psychologists seem to take for granted such ideas as "information," "cybernetics," and "computation," without making them explicit. But when these ideas were first proposed, psychologists were committed exclusively to an empirical proof procedure, and felt deeply suspicious of any theoretical argument. Very few psychologists became directly involved with computational theory, and those who did were often not considered to be doing psychology. The suspicion of "arm-chair theory" kept most experimentalists from becoming directly acquainted with the new ideas, especially because these ideas were not easy to test empirically. Some of this suspicion of theory remains in force today, even among cognitive psychologists. To some extent this is normal. Mature sciences often show a similar tension between experimental and theoretical scientists. But in psychology, the claim to be scientific seemed to rest *entirely* on the existence of experimental proof-procedures. Experimentation was treated with far more respect than theory, so that there was a profound imbalance between the influence of facts and the influence of ideas.

Some Indirect Influences of Computational Theory

A number of cognitive psychologists reject the analogy of computational processes for human functioning (*e.g.*, Neisser, 1976; Gibson, 1966). But

even those who reject a computational metaphor still profit from the concrete "existence proof" provided by the computer. The brute fact of computers helps to create the possibility of doing theory, because we can talk about the functioning of the computer abstractly, without being limited to physical stimuli and responses. Thus the mere existence of the computer has a liberating influence on the theoretical imagination, even for psychologists who do not believe in a computational metaphor.

Ideas that may seem obvious today were much less so a decade or two ago. Computational ideas have spread only gradually, and a formal connection between computational theory and psychological theory was not made until the founding of "cognitive science" in the late 70s. At the beginning of the cognitive revolution, experimentation was still the *sine qua non* of scientific psychology. The computer showed perhaps that representations and information processing were possible, but this possibility rarely convinced psychologists that human beings did *in fact* have representations and goals, or that they did something very much like the processing of information. The real burden of proof for the cognitive revolution was carried by a remarkable series of experiments, which very gradually brought an intensely skeptical community around to a more theoretical point of view.

The Cognitive Revolution: Persuasion by "the Respectable Experiment"

The major mechanism for change in experimental psychology is the persuasive experiment. This is not necessarily true for more highly developed sciences, in which purely theoretical arguments can be more persuasive than empirical observations. Einstein's famous paper on relativity triggered a revolution in the Newtonian physics of the previous two centuries, but it was only 11 pages long, and contained no empirical evidence whatsoever (Einstein, 1905). It merely posed a "thought experiment," asking, "How would the world look to an observer in a vehicle traveling near the speed of light?" A theoretical argument such as this can be persuasive only in a scientific community that views theory with the greatest respect. The behavioristic psychology of the 1950s, however, was deeply suspicious of theory and greatly enamored of experiments. As a result, the theoretical implications of the critical experiments in the cognitive revolution were realized silently, because they could not be stated openly within the accepted framework.

Therefore we must be cautious in discussing the critical experiments.

It is important to place them in their historical and theoretical context—that is the purpose of this book—and yet, we must keep in mind the fact that most psychologists at this time were ahistorical and antitheoretical. Wundt and James had anticipated many of the issues raised by the new cognitive experiments, but in the '50s they were considered unscientific, and their books simply were not read. Many ideas that seem so obvious today were unacceptable, even inconceivable, to serious experimental psychologists at the time. Each new "cognitive" experiment was seen as just another isolated contribution, not creating the coherent whole we see today.

I have noted that the word "cognitive" is ambiguous: For our purposes it refers primarily to a metatheory that encourages one to infer unobservable theoretical constructs from empirical observations. It is rather confusing that the shift toward the cognitive metatheory originated in a field of psychology that is also called "cognitive"—the field that is most concerned with functions such as language, thought, and attention. In a way, this is an historical accident: In principle, at least, the cognitive metatheory could have started in any other part of empirical psychology (and in fact, one could argue that it started first in social psychology). But the cognitive *metatheory*, like the behavioristic one, applies to all of psychology, and indeed today it is rapidly spreading to the rest of scientific psychology.

Yet the cognitive metatheory and the study of human cognition cannot be entirely separated. The study of human cognition provided the empirical domain in which the success or failure of the cognitive metatheory was to be demonstrated, just as the study of conditioning was to be the proof of the behavioristic pudding. A metatheory cannot stand on its own; it must promise not only a plausible approach to the science, but also point to a set of succesful applications of its approach. In just this way, "cognitive psychology" (in the narrow sense) provided a test-case for the cognitive metatheory.

There is a third, rather old-fashioned sense of "cognitive," meaning "conscious, or potentially conscious intellectual activity" such as thinking, problem solving, or memory. This is emphatically *not* the meaning used by modern cognitive psychologists; to them, cognitive functions are roughly synonymous with human informational processes, which can be either conscious or unconscious. Indeed, it has only recently become clear (again) that the overwhelming bulk of effective information processing in the nervous system is unconscious. Nevertheless, this old-fashioned meaning of "cognitive" as "conscious processes" continues to create confusion (*e.g.*, Zajonc, 1980; Baars, 1981).

The Theoreticians' Dilemma: Adequacy versus Testability

The persuasive force of the cognitive revolution was carried almost entirely by "the respectable experiment": By dint of sheer accumulation of novel experiments, the realization slowly dawned that psychological theory would have to become far more abstract and powerful than before. This is true even though explicit theoretical statements in these experiments were always strictly limited. The metatheory still held that science proceeds by small and continuous accumulations of facts, rather than by large, rapid insights. In fact, the theoretical tools needed to accommodate the new facts were developing in tandem with the experimental work—not in psychology as such, but in the new field of Artificial Intelligence. Until recently these new theoretical tools were not considered to be part of scientific psychology. Some Artificial Intelligence researchers thought of their work as a kind of theoretical psychology, but they gained only grudging acceptance among psychologists, often several decades after their work was published (Newell & Simon, 1972, historical addendum). Until very recently, those psychologists who worked in Artificial Intelligence were simply not considered to be doing scientific psychology. For one thing, the computer programs that were able to simulate intelligent functions such as language comprehension or visual recognition were far too complex, interactive, and fast to test empirically.

Unfortunately, the theories that emerged from the experimental laboratories had the opposite problem: They were extremely simple and economical, able to explain certain experimental results, but usually not much more. As Chomsky showed so devastatingly, behavioristic theories of language could not even explain some elementary properties of language (Chomsky, 1957; Bever, Fodor, & Garrett, 1968). Until very recently, *all* experimentally-based theories in psychology failed the test of adequacy—they simply lacked the power to explain even the basic properties of the domain they were supposed to explain. Thus, theories of language could not handle the relationships between paraphrases (such as active and passive sentences with similar meanings). And theories of perception could not explain perceptual constancies—those aspects of perception that allow us to consider an object to be "the same thing" no matter what its orientation and distance. This lack of explanatory adequacy did not make these models useless, because they could always handle the direct results of experiments. But it made them largely irrelevant to the job of explaining perception, memory, or language in the real world.

Over several decades of work, Artificial Intelligence researchers developed computational systems that came much closer to adequacy. Thus, Artificial Intelligence systems were able to do the sorts of things that human beings do with language: answer questions, resolve ambiguities, generate paraphrases, and the like. But these more adequate theories were so enormously complex that they could not be tested with the proof-procedures used by experimental psychologists. A fundamental dilemma arose between theoretical adequacy and empirical testability: One could do experimental research in language, and feel secure as a scientist, at the sacrifice of theoretical adequacy; or alternatively, one could build very large and complex interactive computational systems that could simulate many aspects of human linguistic functioning, without being able to test these ideas experimentally. This fundamental dilemma in cognitive research has not yet been resolved. Although the gap between theory and experiment has narrowed somewhat over the past few years, even a very recent textbook of cognitive psychology reflects the doubts of experimentalists concerning the value of complex computational theories. As Lachman, Lachman, and Butterfield (1979) write:

> Their colleagues have mixed feelings about the global modelers. On the one hand, most information-processing psychologists concede that they are asking the right questions. On the other hand, there is some objection to the movement away from traditional experimentation on the part of some modelers. The global modelers are similar to linguists in their willingness to use rational argument, and they are similar to artificial intelligence specialists in their willingness to be pragmatic for the purpose of implementing parts of their models on computers. To at least some information-processing psychologists, these are character flaws. (p. 437)

A long time lag between theory and empirical testing is not unknown in other scientific fields. We noted before that Galileo's theory was not really tested for 150 years after his death, and in this century in geology the theory of continental drift was not taken seriously for 50 years after it was first suggested. Many similar examples could be cited in other fields. But a great time lag between theory and experimental confirmation is always uncomfortable.

The Cognitive Revolution Proceeded Atheoretically

Thus the revolutionary experiments had to appeal to experimentalists on largely nontheoretical grounds. This may seem strange. To most people in the natural sciences, the idea that experiments can be evaluated without reference to theory may seem peculiar or even incoherent. But experimental psychologists had learned through hard experience that models built to account for one set of experimental findings usually failed to generalize to new results, and that any attempt to generalize to the world outside of the laboratory was likewise doomed to failure. The more ambitious the theory, it was believed, the less likely it was to succeed; and Artificial Intelligence theories seemed ambitious indeed. Suspicion of theory seemed justified by experience.

As a result, experimentalists developed their own criteria for evaluating data: Were the results precise and quantifiable? Were they replicable? Did they show simple functional relationships between the variables? Did they open up new areas of empirical inquiry? The most influential experiments of the cognitive revolution had these empiricist virtues. Quite a few of them appealed to experimentalists because they demonstrated linear relationships between the variables measured. There is no particular reason why psychological functions should be linear, but these kinds of results seemed to guarantee a kind of simplicity that inspired confidence. Straight-line graphs were very persuasive pieces of evidence (*e.g.*, Sternberg, 1966; Cooper & Shepard, 1973; Bransford & Franks, 1971).

Almost every mature science shows a tension between experimentalists and theoreticans. Experimental scientists often complain that the theoreticians simply ignore inconvenient evidence, whereas theoreticians tend to believe that experimentalists cannot see the larger picture. This tension can be healthy and creative, providing that there is also good communication.

A remarkable article in the journal *Science* illustrates this tension vividly in the case of biochemistry:

> At the 1958 Conference on Biophysics, at Boulder, there was a dramatic confrontation between the two points of view. Leo Szilard (Nobel Prize-winning physicist and biochemist) said: "The problem of how enzymes are induced, of how proteins are synthesized, of how antibodies are formed, are closer to solution than is generally believed. If you do stupid experiments, and finish one a year, it can take 50 years. But if you stop doing experiments for a little while and *think* how proteins can possibly be synthesized, there are only about 5 ways, not 50! And it will take only a few experiments to distinguish these."
>
> One of the young men added: "It is essentially the old question: How *small* and *elegant* an experiment can you perform?"
>
> These comments upset a number of those present. An electron microscopist said, "Gentlemen, this is off the track. This is philosophy of science."
>
> Szilard retorted: "I was not quarreling with third-rate scientists. I was quarreling with first-rate scientists."
>
> A physical chemist hurriedly asked: "Are we going to take the official photograph before lunch or after lunch?"
>
> But this did not deflect the dispute. A distinguished cell biologist rose and

said: "No two cells give the same properties. Biology is the science of heterogeneous systems." And he added privately, "You know there are *scientists* and there are people in science who are just working with these oversimplified model systems—DNA chains and *in vitro* systems—who are not doing science at all. We need their auxiliary work: They build apparatus, they make minor studies, but they are not scientists."

To which Cy Leventhal replied: "Well, there are two kinds of biologists, those who are looking to see if there is one thing that can be understood, and those who keep saying it is very complicated and that nothing can be understood . . . You must study the *simplest* system you think has the properties you are interested in."

As they were leaving the meeting, one man could be heard muttering, "What does Szilard expect me to do—kill myself?" (Platt, 1964, p. 350)

And so it goes. The fact is, of course, that theory and experiment are both indispensable, and that it is often useful if they are to some extent pursued independently of each other.

But in psychology, only one-half of this dialectic existed. There simply was no credible theory. In the absence of accepted theory, it was never clear which experiments had important implications. When such implications did become clear after a decade or two, many previous experimental findings suddenly seemed obvious, trivial, or irrelevant. Atheoretical empiricism is a costly way to do science.

We will discuss just a few of the many hundreds of experiments that created the cognitive revolution: experiments bearing on mental representation; on linguistic meaning, attention, and consciousness; on the problem of inferring structure and process; and on the problem of serial order in behavior. For additional detail the reader is referred to any standard text in cognitive psychology, human memory, or psychology of language (*e.g.*, Neisser, 1967; Norman, 1976; Anderson, 1980; Bransford, 1979; Newell & Simon, 1972; Clark & Clark, 1977; Blumenthal, 1977b).

The Issue of Mental Representation

Many cognitive psychologists study memory, and memories are clearly representations of the world. In some reasonable sense my memory of yesterday's breakfast is a representation of that event. With a little effort I could generate an image of yesterday's toast and orange juice, which an outside observer could verify by comparing my verbal description of the image to a photograph record of yesterday's breakfast. My mental image and description can serve as plans for tomorrow's breakfast as well, indicating that representations can also describe future states. Plausible commonsense? Yes, but not acceptable psychology, even to memory psychologists until well into the cognitive revolution. Psychologists were not comfortable with the idea of internal representation, even though they willingly spoke of all kinds of memories. Radical behaviorists considered the idea of "memory" to involve a scientifically unwarranted inference (see Chapter 3, Skinner and Rachlin). Behaviorists preferred to speak of "learning" rather than memory, because a memory is not observable, whereas learning was thought to be definable behaviorally.

On the other hand, computers have memories, and of particular significance, they have buffer memories. Buffers are small, local memories that are used whenever two events that arrive at different times must be related to each other. Buffer memories are always needed if the speed of processing inside the computer is different from the speed of external information. Take the word "they" for example, in the first sentence of this paragraph, "they have buffer memories"; in order to comprehend the sentence, the reader must understand that "they" refers to "computers" in the previous clause. Because the word "computers" was read a half second *before* the word "they," it is natural to suppose that the first word was held in some temporary store, so that when the pronoun "they" arrived, it was matched with its referent. A buffer memory is that kind of temporary holding bin. A significant part of the cognitive revolution began when evidence began to accumulate for the existence of such temporary holding memories in the human nervous system.

Wilhelm Wundt had already performed experiments using brief exposures of a grid of letters, and he reported that trained observers could report only three to six briefly glimpsed letters (1912). Imagine that the following grid is exposed for a few tenths of a second:

t	f	v	x
o	a	z	l
q	c	e	g

People invariably claimed that they could see more letters in a glance than they could report! It seemed as if the very act of recalling the letters or of pronouncing them, or perhaps the sheer effort of holding them long enough to say them, interferes with the memory for what was actually seen. In one of the landmark experiments of the cognitive shift, Sperling (1960) developed an ingenious way to circumvent this difficulty—to measure how many letters people can actually see, no matter how many they report at any one time. He presented an array of 12 letters in a 3×4 arrangement for about one-tenth of a second. Immediately afterwards he sounded a tone, signalling the observer which row of letters to report.

The results showed that people can report the letters from any arbitrarily designated row in this experiment, but not much more than that. We can be sure that the observers consistently perceived for a moment *more* than they could report, because they could name all the letters in any arbitrary row quite accurately. Because the reporting signal was given after the display was turned off, there was no way for the observers to know beforehand which row would be designated. Given that any arbitrary row could be reported, we must infer that *all* of the information in the display was briefly available to the subjects, in spite of their inability to report all the letters in any one exposure.

The Sperling experiment is extremely precise and reliable, and it compels the inference that we have just made: People can remember momentarily more than they can report. To explain this, we are forced to imagine a kind of memory store in which more can be held than the observer can talk about. But if that is so, then the experimental operation of reporting the letters from memory *cannot be a direct measure* of the memory that is being inferred. Operationist philosophy would hold that the measure is the same as the construct, and many psychologists apparently believe this (*e.g.*, Stevens, 1966). But Sperling's results suggested to some psychologists that we may sometimes be obliged to infer a construct that goes beyond the measuring operation—in this case, a visual buffer memory, whose contents cannot be fully reported at any one time.

Another construct, called "short-term memory," has much the same methodological status. Wundt and James had noted that people could report no more than about six separate stimuli in any sensory modality, but psychologists had largely ignored this phenomenon for more than 50 years. Starting in the mid-1950s, people such as George Miller and Herbert A. Simon began to draw attention to the massive evidence for this fact. In a charming paper entitled "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," Miller (1956) wrote,

> My problem is that I have been persecuted by an integer. For seven years this number has followed me around, has intruded in my most private data, and has assaulted me from the pages of our most public journals. This number assumes a variety of disguises, being sometimes a little larger and sometimes a little smaller than usual, but never changing so much as to be unrecognizable.

People can report only seven (plus or minus two) unrelated words or arbitrary numbers in immediate memory; this is why telephone numbers are limited to seven digits. In judging anything from "size" to "degree of happiness," we can reliably use only about seven judgment categories on a scale. We can accurately estimate in a single glance the number

of no more than seven objects. Thus, "short-term memory" seems to be drastically limited.

But there is a peculiar amendment to this observation: As Wundt had already noted, it does not matter what *size* each item in the short-term memory is. We can remember some seven letters, syllables, or multisyllabic words, and somewhat less than seven short, idiomatic sentences. This short-term memory, therefore, has the same capacity for seven separate, unrelated items, no matter how lengthy the items themselves are. If we think of short-term memory as a container, say a shoe box, we can immediately see the problem: What kind of a shoe box will take seven shoes of *any* size—either seven baby shoes or seven giant boots?

Miller gave a partial explanation in terms of "chunking"; the capacity of the short-term memory buffer appeared to be limited to roughly seven chunks of any size. By recoding ("chunking") letters into words and words into sentences, more efficient use could be made of this capacity. But a real explanation of this phenomenon had to await a deeper understanding, not of short-term memory, but of long-term memory.

The buffer memories discussed so far are only temporary holding bins. Obviously, people can remember some things for years, and under optimal conditions perhaps any experience can be retrieved long afterwards (*e.g.*, Bransford, 1979; Williams & Hollan, 1981). Evidence was soon discovered that items held in short-term memory tend to move into long-term memory, especially if people actively try to integrate the newer items into existing knowledge. The great difference between short-term and long-term memory seemed to be the fact that long-term memory was highly organized. The evidence for this is as follows:

It had long been known that if people are asked to say any words that come to mind, their responses will seem obviously connected to each other. The word "book" may be followed by "volume," "space," "container," "Coke bottle," "bottle-fed babies," "breast-fed babies," "baby carriage," "carriage-house," and so on. Each response seems related to the next. This free-association technique has been used extensively in psychoanalysis since the beginning of the 20th century, to encourage trains of association that people would normally avoid. Indeed, the early experimental work on free association was done by Carl Jung.

But evidence from free association did not fit the experimental design criteria of experimental psychology, because the connectedness between the words could only be known *post hoc*; *post hoc* designs did not fit the assumptions of the statistical tests needed to support experimental claims. Only if all conditions were specified in advance could one draw statistically valid conclusions, or so it was thought. (It is noteworthy that this requirement entirely rules out new observations, that is, results not anticipated in the design.)

How could a respectable experimental psychologist test the possibility that people would spontaneously organize words in memory? Psychologists such as Bousfield (1953) taught their subjects lists of words belonging to different categories (such as animals, furniture, and musical instruments) and randomly mixed together, then after a day or so, tested the subjects on these words. Because all the conditions were established ahead of time, this was an acceptable experimental design. The results showed what Jung had shown in the 1930s: that people retrieved like with like. This seemed to imply that the words were being categorized together in memory.

Although this idea may not seem earth-shaking, it was quite revolutionary to psychologists who believed that memories were arbitrarily connected to one another, as taught by the English Associationists and the behavioristic learning psychologists. That memories should actually be *classified* in some organized scheme profoundly challenged their views. Unlike Sperling's visual buffer memory and Miller's short-term memory, long-term memory seemed to be profoundly influenced by organizational considerations.

By 1965, it appeared that there were three kinds of memory: sensory memories, such as the visual buffer, which could hold about 12 items; short-term memory, which could hold seven, plus or minus two, items (if people were allowed to rehearse the items silently to themselves); and long-term memory, a mysterious entity that held the bulk of what we ordinarily call memory, and which seemed to be sensitive to organizational effects. Often these three types of memory were encoded in a linear model, as it were, end-to-end: First, sensory information entered a sensory buffer; then some of the items in the buffer were selected to go into short-term memory, where seven of them could be rehearsed, until finally, they could enter long-term memory. This linear model was perhaps naïve, but it was clearly cognitive—nobody expected actually to *see* these memories in the flesh—and it did account for a great many observations.

Because the linear model was so clearly cognitive, perhaps we should ignore what was wrong with it. But the cognitive revolution did not really stop in 1965; it has continued in one way or another and is still continuing today. It is true that by 1965, the *metatheory* of scientific psychology had already changed to include unobservable constructs as a routine matter. But while we will focus on the years 1955 to 1965, it would be a pity to neglect more recent developments.

What was wrong with the linear model, with the idea of information

flowing from sensory to short-term to long-term memory? For one thing, it was self-contradictory. There is a problem with the very idea of an "item" in memory. What would constitute an "item"? In Sperling's 3 × 4 visual display there are some 12 "items," but a backwards "t" would not show the same effects on a subject as a normal "t", because of its unfamiliarity. Similarly, in short-term memory, the sequence "1776, 1492, 1914, 1942" would be remembered easily, whereas the same numbers presented backwards would be very difficult to remember. Obviously, what constitutes an "item" depends on familiarity—but familiarity implies that something was experienced before, perhaps many times, and over a long time; thus, familiarity itself depends on long-term memory. The way in which information is handled in a visual buffer or in short-term memory depends, in itself, on long-term memory. Therefore it cannot be true that information flows only one way, from sensory buffers to short-term memory to long-term memory, because long-term memory is needed *in the first place* to define the items in the buffer memories.

Thus the linear model did not last very long. For the first time in the history of the new cognitive psychology, some purely theoretical influences were helping to change the thinking of some psychologists (*e.g.*, *cf.* the first and second editions of Norman, 1967 and 1976).

Artificial Intelligence researchers had at first attempted to set up computer systems able to recognize auditory, visual, and linguistic inputs in ways rather similar to the linear model; but they quickly realized that there were far too many ambiguities in the input to permit a deterministic flow of information. Thus, the sentence, "Time flies like an arrow," was interpreted by a computer as meaning "A certain kind of a fly, called a 'time fly,' has a liking for arrows." This may seem like a unusually silly interpretation, but the same interpretation is perfectly all right in a sentence such as, "Fruit flies like a banana," because fruit flies *are* a certain category of fly that have a liking for bananas. So the difficulty was not that the computer constructed a totally implausible interpretation of the sentence, but, rather, that it did not have enough contextual knowledge to resolve the genuine ambiguities in the sentence. Ambiguities abound in ordinary language. When human beings have the right contextual knowledge, we resolve such ambiguities very quickly. But without context, we encounter the same difficulties that hobbled the early computer models of language comprehension.

Artificial Intelligence workers soon recognized that any sentence requires abstract, prior knowledge in order to be understood. Comprehending "Time flies like an arrow" requires prior knowledge about metaphors, about the fleetingness of time, about flying and its similarities to other speedy forms of motion, and it requires information about the speaker and his or her purposes, the social and semantic context, and so on. Any linguistic input requires "top-down" information, which is derived from many kinds of abstract knowledge, as well as "bottom-up" information, which is derived from the input itself (Winograd, 1972). When this need for contextual information became widely understood, a great deal of work in Artificial Intelligence began to focus on specifying contextual knowledge in more and more detail.

Psychologists who were in touch with these developments now realized that they had made the same mistake in interpreting their experimental results. Information available to the senses also is frequently ambiguous, and it, too, requires prior knowledge to be decoded. "Long-term memory" was needed to deal with the material in "short-term memory" (*e.g.*, Norman, 1976). The upshot was that it made much more sense to interpret short-term memory as a special aspect of long-term memory, a buffer that not only introduces materials into long-term memory, but also reflects long-term memory. We might say that short-term memory is what happens to long-term memory when it is unable to organize information in the most compact way. When chunks of memory are separated from the context that normally gives them meaning, the relationships between the items is lost. But if the relationships between the seven items in short-term memory are restored, the items stop behaving as separate units, and start acting as a single item. Thus it seems that short-term memory is just a special case of long-term memory. The buffer memories are just a "front end" for the normal, very large, and highly organized repository of our knowledge.

This realization signalled a new era in the methodology of scientific psychology. For the first time there was a fruitful confluence between empirical psychology and theoretical work.

Other Aspects of Mental Representation

During the 1950s, '60s, and '70s, a great many other experiments were being performed on the general topic of mental representations. The issue of *imagery* was rediscovered, and shown to be testable after all, after some 50 years of rejection. Work by Paivio (1971) and Cooper and Shepard (1973) showed that precise, reliable experiments could be performed with mental images, and that the resulting data were clean and simple. There are several curious historical addenda to this. Consider, for example, the commonsense idea that an image is *conscious*, in fact, that an image

is a conscious representation of something. During the 1960s and '70s, in the midst of much research on imagery, this rather obvious idea could not be stated by experimental psychologists, because the word "conscious" was still under the ban. This curious gap has continued to lead to considerable confusion to this day (Baars, in press).

But is the expression "conscious representation" not redundant? Aren't all representations conscious? The answer is clearly no — my memory of yesterday's breakfast presumably exists even before I retrieve it, and it becomes conscious only when it is retrieved. Indeed, the Russian physiologist E. N. Sokolov (1963) had already made some elegant arguments for the existence of unconscious representations for those repetitive stimuli to which we have become habituated (the things we have gotten used to). Certain findings about attention also suggest that unconscious representation and information processing could take place (below). The fact is, however, that these realizations did not become generally accepted until the late 1970s.

Although Artificial Intelligence workers had been dealing with highly abstract, complex, and active representational systems since the 1950s, psychologists in general were not ready to accept the idea of abstract mental representation. Some remarkable work by John Bransford and Jeffrey Franks, two graduate students at the University of Minnesota, helped to convince many skeptics of the need to postulate more powerful kinds of knowledge representation.

Bransford and Franks conducted experiments in which they presented several simple stories in a series of sentences, without asking their subjects to memorize the sentences. They simply asked a short question about each sentence, to make sure that its meaning was understood. Afterward, people were presented with completely new sentences and asked which ones they had heard before. Invariably, if the meaning of a new sentence was consistent with the meaning of an old sentence, the subject reported having heard it before. Yet subjects were quite able to detect the novelty of even very similar-sounding sentences, *if* those sentences had a slightly different meaning. Recognition ratings for the new sentences were a linear function of the number of semantic propositions contained in the sentences (Bransford & Franks, 1971; Bransford, 1979).

These experiments are quite natural, because they simply ask the subjects to do what they normally would — that is, try to understand the sentences. But the results have had far-reaching consequences. They have compelled a deeper psychological theory than was hitherto permitted. The memory for "gist" — some sort of meaning representation — is obviously quite a bit better than the memory for actual sentences. Follow-up work showed that subjects also confused with the original sentences those new sentences which expressed immediate inferences from the original meanings (Bransford, 1979). The implication is that we construct semantic representations from the direct meaning conveyed by the sentences.

In the mid-1950s, it was still common to ridicule the notion of mental representation. Twenty years later, the issue became the core of the new psychology. Over several decades, psychologists were forced, by clear empirical results, to adopt a highly abstract theoretical language. Instead of chains of stimuli and responses, they found it necessary to speak of abstract semantic representations that encode not just the meaning of individual words, but an inferential, deeply interpreted representation of the world. In its theoretical sophistication the new perspective was light-years beyond stimulis-response theory.

It now appeared that human beings in general represent their world much as scientists do—in abstract theoretical terms. The abstract theories are highly active: People perform numerous symbolic inferences with them. Mental representations seem to be economical and effective, and they require fast-acting, automatic symbolic processes to make them work. The human nervous system, far from being a passive bundle of connections between stimuli and responses, begins more and more to resemble an enormous, highly sophisticated, active, intelligent, and flexible mechanism.

The Issues of Attention and Consciousness

Consciousness was of course the great bone of contention between 19th-century psychology and the behaviorists. Wundt and James believed that psychology *was* the study of conscious mental contents, and Watson and his followers often denied the very existence of consciousness. The analysis of "attention" was said to be the great accomplishment of 19th-century psychology, and surprisingly, the concept of attention did not entirely disappear during the behavioristic era. At least in the 1950s, American behaviorists began to speak again of stimulus-selection mechanisms, such as Pavlov's "orientation reaction" and the "observing response," whereby animals turn their sensory receptors toward a source of stimulation. But even this was an utterly impoverished version of what had been known about consciousness and attention in the 19th century.

Even though 19th-century psychology defined itself as the study of the contents of consciousness, it was handicapped in dealing with consciousness itself because it *presupposed* consciousness. On the very first page of his book, *An Introduction to Psychology* (1912/1973), Wundt remarks on this fact:

> If psychologists are asked, what the business of psychology is, they generally make some such answer as follows, if they belong to the empirical school: that this science has to investigate the facts of consciousness. . . .
>
> Now although this definition seems quite perfect, it is really *to some extent a vicious circle. For if we ask further, what is this consciousness which psychology investigates? the answer will be "It consists of the sum total of facts of which we are conscious."* (p. 1, italics added)

In this sense, psychologists such as Wundt and James were prevented from really seeing the issue of consciousness as a *problem* for scientific investigation. In part, this difficulty was due to their unwillingness to recognize the existence of unconscious processes; one cannot discuss an entity without considering the absence of that entity. But even Freud, who was certainly willing to infer the properties of unconscious processes, had little to say about consciousness *as such*. Thus, in *An Outline of Psychoanalysis* (1940) he writes:

> The starting point for this investigation is provided by a fact without parallel, which defies all explanation or description—the fact of consciousness. Nevertheless, if anyone speaks of consciousness, we know immediately and from our most personal experience what is meant by it. (p. 34)

Like many 19th-century thinkers, Freud was greatly impressed by the existence of unconscious factors, but had little to say about conscious processes. Perhaps they seemed too obvious.

Behaviorism broke the vicious circle regarding consciousness, only to deny its existence altogether. Watson viewed consciousness as nothing but "the soul of theology," and given that theology was clearly unscientific, so was consciousness.

All the same, behaviorists did break the conceptual logjam imposed by the assumptions of 19th-century psychology. Tolman wrote a speculative paper on consciousness in 1927. But it was Clark Hull who put into words something that was unclear to the psychologists of the 19th century—namely, that consciousness might actually be *a problem* demanding a psychological explanation. As he wrote in 1937:

> . . . to recognize the existence of a phenomenon (i.e. consciousness) is not the same thing as insisting upon its basic, i.e. logical, priority. Instead of furnishing a means for the solution of problems, consciousness appears to be itself a problem needing solution. (p. 29)

This remark represents a high point in behavioristic thought. Had he been allowed to think freely about this problem, Hull might have proposed an interesting theory of consciousness. But of couse in his day, the narrow limits on psychological theorizing kept him from fully using his theoretical imagination. Thus Wundt and James could not discuss consciousness because it was presupposed in everything they did, and behaviorists could not talk about consciousness because to do so required metatheoretical permission to postulate unobservable constructs. Only today is it possible to avoid both of these presuppositional traps.

Consciousness is a pervasive psychological issue. Commonsense psychology takes it for granted that people are conscious of anything they sense in the world around them, that dreams, images and hallucinations are conscious, and that we ordinarily exercise conscious control over our actions. On the other hand, everyday psychology also assumes that sleep and coma imply the absence of consciousness, that we lose consciousness of repetitive events as we get used to them, and that when we become very skilled at some routine action we tend to lose consciousness of it as well. Common sense, then, views consciousness as constrained in a number of ways.

Even during the behavioristic era it was possible in some cases to find behavioral correlates for the things commonsense psychology describes in terms of conscious experience. For example, the fact that people lose awareness of constant or predictable stimulation was studied behaviorally as a decrement in the "orientation reaction" first analyzed by Pavlov (1927/1960). A dog will prick up its ears and look at a source of stimulation, and a number of physiological changes will take place in its body as it does so (*e.g.*, Berlyne, 1960). But as the stimulus is repeated, the dog will no longer point its receptors to the source of stimulation, and the psychological concomitants of the orienting reaction will cease as well. This is a behavioral way to describe habituation of awareness, without recourse to concepts such as attention and consciousness.

But some things cannot easily be described in such behavioral terms, and the prohibition against the idea of consciousness can create an awkward theoretical handicap. A notable case was Jerome Bruner's attempt, in the 1950s, to test the psychodynamic notion of repression by means of a "perceptual defense" concept (Bruner & Postman, 1947; Erdelyi, 1974). Repression is defined as a defense against a potentially conscious thought, but of course this vocabulary was not acceptable within experimental psychology in the 1950s. Rather, Bruner was forced to speak in terms of *perceptual* defense. Instead of consciousness, one was compelled to speak of perception (some behaviorists even considered this term to be mentalistic and, hence, unscientific). Bruner and Postman (1947) were able to show that briefly exposed taboo words were reported less often than neutral words, and interpreted this as evidence for the idea that people defended against the perception of these words. Bruner was a cognitive psychologist from the very beginning, and his work on perceptual defense

was really an attempt to trigger an early cognitive revolution, one tha would also do justice to psychodynamic concepts. As it turned ou however, although many hundreds of experiments were published o perceptual defense, the idea was criticized and lost credibility (see Erdelyi 1974; Bruner, 1984).

There were several reasons for Bruner's failure to start a cognitiv revolution. Perhaps the most telling was the inability of psychologist of the time to give a coherent account of a process that can *recognize* taboo words, and yet *prevent* the taboo words from being perceived. It made no sense to believe in this "Judas eye"—to think that a word could be detected and yet not perceived; but this is of course an idea that gives modern theoreticians no difficulty in principle. Words may be represented and processed in the nervous system long before they become conscious. But at the time, it made no sense to speak of "unconscious representation" of the taboo words, which would prevent them from becoming conscious. The distinction between conscious and unconscious representation could not be maintained, and as a result, the notion of "perceptual defense" was rejected as incoherent. Indeed, the same objection was often made to Freud's notion of repression. This rather sad episode in the history of psychology illustrates again that the effect of behaviorism was often to make perfectly good theoretical proposals incoherent and incomprehensible. Behaviorists did not even feel the need to argue against such ideas, because they seemed to make no sense, and a nonsensical idea cannot be tested empirically. In this way, some of the most interesting hypotheses in psychology simply went untested.

Psychologists began to rediscover consciousness and attention in the 1950s, when a group in England led by E. Colin Cherry (1953) and Donald E. Broadbent (1958, 1971) began to investigate the effects of listening to two spoken messages at once. In the 19th century it was common to speak of "the unity of consciousness," a phrase by which one meant that people can only experience one interpretation of a stimulus at a time. One way to show this unity is by giving people two streams of information, especially two streams of continuous speech. To be sure that the subjects are actually attending to the speech, they are asked to repeat one stream of speech even while they hear it. With practice, people become very good at this "shadowing" task, and are able to follow the input with a delay of a few tenths of a second; but under these circumstances people can only follow one stream of speech at a time. With this technique, it became possible to investigate consciousness and attention experimentally, though the *word* "consciousness," of course, was not used for several more decades. Much like the work on short-term memory, the findings were interesting, but ultimately led to a difficult paradox.

Broadbent initially proposed that the selectivity of attention occurred because the subject was filtering out unwanted information (1958). Thus, one could shadow only one stream of speech, because any other source of stimulation was being filtered out. Electronic filters were well understood at that time, and it seemed plausible that one simply "turned down the volume" in the ear that was receiving the unwanted speech. By 1958, however, this filter hypothesis had been elegantly disproved by two Cambridge undergraduates, who simply alternated two streams of speech between their subjects' two ears several times a second. Even under such circumstances, people continued to monitor only one message, and they still were unable to report the "nonattended" speech (Gray & Wedderburn, 1960). Thus it could not be true that one filters out the speech in one ear. Furthermore, in order to recognize the rapidly switching stream of speech as *the same one* even though it was coming into different ears, the subjects had to be sensitive to the *content* of the speech—the words' syntax and meaning. Otherwise the attended stream of speech would be confused with the nonattended speech. But if such decisions were being made on the basis of content, then how could one be filtering out the content?

Some time later it was shown that even the nonattended stream of speech had observable effects. For example, if the attended stream of speech contained an ambiguous sentence ("I am going to sit near the *bank*"), and simultaneous with the word *bank* the nonattended channel contained the word "water," the subject would tend to interpret "bank" as "river bank"; but if the nonattended channel instead contained the word "money," the subject's interpretation would tend to shift to "financial bank" (MacKay, 1973). Thus Broadbent's early proposal that the nonattended stream of speech was being filtered out was contradicted: The meaning of "water" or "money" had to be available unconsciously, so that the conscious word "bank" could be interpreted. But if unconscious information is processed just as conscious information is, what is the real difference between conscious and unconscious events? Broadbent had proposed that filtering out unwanted material helps to save processing capacity for more important information; but if all input is in fact analyzed, no work is saved. Again, a promising research direction seemed to lead to paradox.

The same kind of paradox was encountered in the study of short-term memory, where it seemed at first that sensory input is analyzed bottom-up. That is, we were presumed to analyze speech by examining first the sound, then the phonemes and morphemes, then words and syntax, and finally the meaning of the input. But work on short-term memory, as well as Artificial Intelligence theory, showed that this could not be true; that, in fact, one needed to use prior high-level knowledge in order to process even the lowest level of input. Top-down processes were needed as much as bottom-up ones. Thus, the filter theory of attention, which assumed that one could shut out the *sound* of nonattended messages, was bound to fail. Even the lowest level of input analysis may require high-level top-down processes. This realization explained, at least in part, why unconscious words seem to affect the semantic interpretation of conscious input, and also why it is possible to switch two streams of information rapidly between the two ears, without the subject's losing track of the attended message.

Clearly a great deal about consciousness and attention remained to be understood. But it was not until the late 1970s that psychologists were even willing to resurrect the word "consciousness" from its premature grave, and to address the issues it raised in an experimentally precise and theroetically creative way (*e.g.*, Mandler, 1975; Shallice, 1972, 1978; Baars, 1983a, in press). This is currently perhaps the most important theoretical issue in cognitive psychology, one whose outcome is by no means clear.

The Problem of Serial Order

Behavior takes place over time, and the components of behavior must be ordered. During the behavioristic period, only one theoretical mechanism was invoked to handle this serial aspect of behavior—namely the *chain* of stimuli and responses. An external stimulus was understood to evoke a response, which itself had stimulus properties, which were connected to the next response with its own stimulus properties, and so on. This, in essence, is Hull's theory, which was the most sophisticated attempt made to deal with this problem (1937). But in 1951, Karl Lashley pointed out some fundamental inadequacies of this conception. A chain of stimuli and responses cannot explain, for example, the cases where elements in the chain appear to "leapfrog" over each other. To make his point, Lashley cites the case of spoonerisms—"our dear old Queen" being transposed into "our queer old Dean."

There are numerous situations where parts of a sequential action appear to be leapfrogging. Take the case of active and passive sentences. "The boy chased the cat" is a near-paraphrase of "The cat was chased by the boy." In order to recognize the similarity of these sentences, one must be able to transpose the order of "boy" and "cat," and this capacity would require more than just a chain of word-responses. Wilhelm Wundt had made this very point when he noted that Latin grammar allows the order of words to change quite freely without changing the meaning of a sentence (Blumenthal, 1970).

The fact that chaining theories cannot explain leapfrogging elements can be made clear to a child: Picture a train going into a tunnel, with a locomotive, a passenger train, and a caboose. But at the other end of the tunnel, the train comes out in reverse order: The caboose comes first, followed by the passenger car, and finally the locomotive. What could have happened inside the tunnel? Well, maybe there is a parallel switching track or maybe a giant crane was used to exchange the caboose and the locomotive. But whatever mechanism we imagine, it must involve *more than a single* railroad track. In the same way, one must imagine more than a serial chain of elements in order to explain the exchange of phonemes in a spoonerism, or the ability people have to transform active to passive sentences. This realization forces one to abandon the sole theoretical mechanism that behavioristic theoreticians, such as Hull, were able to use.

Noam Chomsky first achieved fame with his monograph, *Syntactic Structures* (1957), in which he used the mathematical theory of automata to prove formally that the class of chaining theories (technically called Markov processes) cannot in principle represent even the simplest kind of grammar that is needed to represent the linguistic competence of human beings. (The simplest grammar that can describe some important aspects of sentence structure is called a "finite-state" or "tree-structure" grammar. It is effectively the same formalism that most people learn in elementary school by "diagramming sentences.") Chomsky went on to prove further that any grammar able to represent such things as the similarity between active and passive sentences would have to be even more powerful than a tree-structure grammar. This powerful but minimally necessary grammar he called a Transformational Grammar.

In many ways Chomsky's argument was a formal equivalent of the informal argument given by Lashley (1951) and by Wundt in the previous century (Blumenthal, 1970). There is a class of symbolic strings called "mirror-image strings," in which the first half of the string is a mirror-image of the second half. Thus ab, abba, baab, abbbba, and aabbaa are mirror-image strings, whereas aabb and baa are not. Anybody can learn to recognize mirror-image strings very quickly, and almost anyone can suggest a formal procedure to test whether a string is a mirror-image string. But it can be proved mathematically that such a string cannot be represented in a general form by a chain grammar (Markov Process). Thus, even so trivial a learning task as recognizing a mirror-image string

cannot be represented by a Markov process. Hull's chaining theory was proved unable to account for linguistic competence, or for any other ability of similar complexity. This proof was later generalized to cover all chain theories of serial-order behavior (Bever, Fodor, & Garrett, 1968).

Chomsky's work caught on slowly, but once it became widely understood it had very great impact. Its effect must be understood in terms of the enormous resistance to complex theory shown by behavioristic psychologists. Behaviorists were intent on producing extremely parsimonious theories, theories that did not postulate any more elements than were strictly necessary. Chomsky's formal proofs made it clear that these parsimonious theories *could never, in principle, explain what they were intended to explain*. The theories lacked adequacy. Thus psychologists were forced to postulate more complex theories than before.

This development had a curious outcome, because it was not long before it was proven mathematically that Chomsky's proposed Transformational Grammar was formally equivalent to a Turing Machine (Turing, 1950). But a Turing Machine can compute *any* function whatsoever. Thus a Transformational Grammar, which is the *minimum* necessary grammar to account for language, is formally equivalent to the *maximally* powerful kind of theory. Very soon after Chomsky's work, computer scientists working in Artificial Intelligence began to develop grammars other than Transformational Grammar, but these too, were invariably equivalent to Turing Machines. Indeed, there seems to be no viable system that can model some significant part of human behavior that is less powerful than the most powerful mathematical system. What all this means is that *the mathematical theory of automata can suggest no kind of parsimony for any system able to account for some significant part of human behavior*. All theories must be mathematically maximum theories.

One can suggest other aspects of theoretical parsimony, such as computational efficiency, fast-processing speed, and limited memory load, but these properties cannot be handled by the theory of automata. In order to place constraints on the set of possible psychological theories, psychologists would have to look elsewhere than pure mathematics.

The Problem of Inferring Process and Structure

Increasingly the question was raised: Under what conditions can one infer the existence of invisible, but theoretically necessary, constructs? Chomsky had suggested that language cannot be understood except by reference to *rules*. Indeed, according to Chomsky, the sentences of a language are generated by rules in much the way algebraic formulas are generated by the basic elements and operations of algebra. But rules are abstract entities—they are symbolic representations that can account for very large sets of observable events. And for the speakers of a language, of course, the rules are not conscious. Sets of rules make up a structure, a relatively stable knowledge representation. But structure is not enough to make knowledge work. Rules must actively operate on incoming information, on information in memory, and on plans for controlling actions. These requirements raise the issue of *process* as well as *structure*. An adequate cognitive psychology must be able to infer both process and structure.

Structure was usually inferred by what might be called *equivalence operations*. People will treat two superficially different events as part of one unified structure: Two different orientations of the same physical object (for example, a chair) are treated perceptually as just two aspects of the same "thing"; two paraphrases of a sentence are often confused with each other in memory, as are two similar-sounding syllables, or two acoustically different versions of the same phoneme. These equivalence classes can be represented theoretically as a single structure, so that at an abstract level, the two sentences that are confused in memory may be represented in one structure.

This view of mental structure corresponds directly to what people for ages have called "knowledge." The Indo-European roots of the word "knowledge" go back to the very earliest known languages, with roughly the same meaning; so the idea of knowledge is at least 4000 years old. Very often we can infer that people know something by presenting material in one form and testing their knowledge in a different form. College examinations may present paraphrases of the original material, to test whether students can remember the abstract content of the sentences they hear as well as the sentences themselves. If this equivalence class exists, one can infer knowledge. In a way, then, cognitive psychologists assess knowledge in the same way everyday psychology does.

Finding evidence for mental *processes* is a good deal trickier. The main source of evidence for inferring process has come from measurements of reaction time in fairly simple tasks. In the 19th century the Dutch psychologist Donders (1868) suggested a way of measuring the duration of mental processes by interposing " . . . into the process of the physiological time some new components of mental action" (p. 418). Donders knew, for example, that the time needed to respond to a single predictable stimulus (called *simple* reaction time) was always shorter than the time needed to respond differentially to one of two possible stimuli (*choice* reaction time). For example, he showed that the time needed to react to an electric shock delivered to *either* the right or left leg was longer than

the time needed to react to a shock that would invariably be given to only one leg. The time difference between these conditions was 66 milliseconds, and Donders interpreted this time difference as the time required to choose between the two possible responses in the choice reaction time task. But notice that some of this time might be needed to choose between the two *stimuli* (left or right shock) rather than between the two *responses* (saying "left" or "right"). To distinguish between these two interpretations, Donders added a *third* experimental condition, in which several stimuli were presented but only *one* had to be responded to (the so-called "c-reaction"). He proposed that:

(1) Stimulus Discrimination = c-Reaction Time – Simple Reaction Time Time

and

(2) Response Choice Time = Choice Reaction Time – c-Reaction Time

Thus he could separate, by inference, the time needed to select the right stimulus from the time needed to select the right response. These values he found to be 36 and 47 milliseconds, respectively, in a verbal reaction time task.

Donders's clever solution to the problem of inferring processing time was based on some debatable assumptions. Even in the 19th century some psychologists argued that other explanations for Donders's results were possible. For instance, in simple reaction time (where only one response is possible), subjects might be better prepared to make the response than in a choice reaction time task. The difference between simple and choice reaction time is not just that a "stage" has been added—there is a qualitative difference as well. Moreover, Donders assumed that all mental processes must be discrete and serial. This may sometimes be so, but there is much reason to think that some processes are gradual, interactive, and parallel.

Interest in the use of reaction time tasks for inferring the duration of mental processes received a great boost as a result of the cognitive shift (*e.g.*, Sternberg, 1966; Posner, 1978). In 1969, Sternberg proposed a new approach to the inference problem, the "additive factors method," based on a statistical technique called analysis of variance. Using analysis of variance, one can discern the effect of several independent variables on some dependent measure such as reaction time. It also provides a way of deciding whether two or more variables interact in their effect, or whether the individual effects are only additive. Sternberg proposed applying the same logic to the study of reaction time. His method is not based on adding or deleting stages of processing, but on the possibility of *selectively influencing* the duration of different stages (Ashby & Townsend, 1980).

In his first demonstration of the technique, Sternberg manipulated three variables in a choice reaction time task: First, the quality of the stimulus was either degraded or normal. Second, the number of stimulus–response alternatives was varied. And finally, the compatibility between stimuli and responses was varied between high and low compatibility. The results indicated that stimulus quality and stimulus–response compatibility both had additive effects on reaction time, but that these two variables interacted in a nonadditive way with the third, the number of stimulus–response alternatives. These results were interpreted to mean that there were, indeed, two separate processing stages. The first, or *stimulus encoding* stage, was affected by the quality of the stimulus and by the number of alternative stimuli that had to be evaluated. The second stage, consisting of *translation* and *response organization*, was affected by stimulus–response compatibility, and also by the number of alternative responses available.

Sternberg's additive-factors logic has been widely applied, and other methods have also been proposed. A large and active research literature has sprung up. But is the goal of it all—a reliable method for inferring the details of human information processing—really within our reach? It may be too early to tell.

Cognitive Science: The Rise of a Theoretical Psychology

Perhaps the most exciting recent development growing out of the cognitive revolution is a new trend toward integrating all the major disciplines concerned with studying the nature of knowledge. This new interdisciplinary field has been called "cognitive science," and it includes cognitive psychology, linguistics, philosophy, Artificial Intelligence, and the neurosciences. These fields are indeed developing a set of common concerns, and even, to some extent, a common language. This new, integrative trend is most encouraging—no longer is it necessary for psychologists to prove their scientific standing by rejecting philosophy or neurophysiology as irrelevant to their concerns. Rather, there is a wide appreciation of commonalities in the problems encountered by philosophers, psychologists, linguists, Artificial Intelligence workers, and neuroscientists. Indeed, at a theoretical level, these problems often seem to be identical.

Artifical Intelligence provides the theoretical core of cognitive science, because it gives us a theoretical language that permits us to be precise about cognitive issues. I have briefly mentioned the study of Artificial Intelligence, but only in passing. Given the widespread suspicion of theory in experimental psychology, for many years only a few scientific psychologists took Artificial Intelligence work seriously. It appeared to be entirely theoretical and nonempirical. It was not until the 1970s that a significant group of psychologists began to look seriously at this field (Norman, 1976; Anderson & Bower, 1973; Anderson, 1980). The rise of cognitive science signals a change in the rejection of theory among psychologists, and the reasons for this new receptivity are worth discussing.

Psychologists have been involved with the development of computer languages from quite an early point (see the interview with Herbert Simon, Chapter 7). Allan Newell and Herbert Simon, who considered themselves to be psychological scientists, helped develop some of the first higher-level computer language, and were apparently the first to articulate the very important idea of recursiveness in high-level computer languages (Newell & Simon, 1972). They and other pioneers in the 1950s developed the first simulations of intelligent human functions, including short-term memory, problem solving, chess playing, and logical theorem proving (see Newell & Simon, 1972; Feigenbaum & Feldman, 1963; Schank & Colby, 1973). Such simulations performed several functions: They showed that computers can do some things that are considered signs of intelligence when performed by human beings; and further, they suggested a possible theory of the way in which human beings do the same task. Thus, a simulation of chess playing can serve as a theory of how humans play chess—in this case, a largely incorrect theory. Finally, any other theory of chess playing, language comprehension, or emotional conflict could be encoded in a computer program if it were made sufficiently explicit. Thus, computer programs provide a natural theoretical language for stating psychological hypotheses in a very explicit way. And any theory that cannot be made explicit enough to run on a computer is *ipso facto* a faulty theory.

This last point is of enormous importance for psychology. Historically, scientific development is severely hampered without a theoretical language that can naturally model the subject matter of the science. If Newton had done nothing more than invent the infinitesimal calculus, which is a natural language for modeling the movement of objects in space, his place in the history of physics would have been assured. Psychology has never had a language able to express *in a natural way* the facts we observe. That language now seems to exist, and if this were the only contribution made by Artificial Intelligence, its place in the history of psychology would be assured.

Psychologists have often tried to apply to their subject the kinds of mathematics that proved so successful in the physical sciences, with generally poor results. Early in the 19th century, the physicist J. F. Herbart tried to apply algebra and even calculus to model the way in which thoughts rise to consciousness, but this attempt was essentially metaphorical (Miller, 1964). There was no way to quantify the strength of a thought with any precision. Similarly, in this century, Hull attempted to use algebraic formulas in an essentially metaphoric fashion; and Kurt Lewin tried to apply ideas from mathematical topology, using vectors to indicate the strength of a goal, a bounded region to represent the "life space," and so on (see Miller, 1964). By the 1950s, a more sophisticated and modest group of mathematical psychologists appeared, who applied more appropriate kinds of mathematics to certain limited psychological problems: probability theory, decision theory, and signal detection theory. Some of this work has been quite valuable, and yet, all of these mathematical applications still assume that the kinds of mathematics that work in physics and chemistry will fit the requirements of psychology.

In particular, it is assumed that *quantification* is important for representing psychological phenomena. It is quite possible, however, to get mathematical precision without quantification—for example, in symbolic logic, Boolean algebra, topology, and, perhaps most important for psychology, in the theory of automata and a related field called recursive function theory. There are many reasons to believe that some kinds of nonquantitative mathematics are more important for psychology than standard algebra. For example, Chomsky's theory of linguistics is based entirely on automaton theory. In linguistics, one deals most naturally with a string of symbols, and quantitative questions about these symbols are irrelevant. What is important about a sentence is not how loudly it is spoken, nor even how long it is; what is important is its meaning, its syntactic structure, the words it contains, and the like.

There are some remarkably simple ideas that make it possible to model language, semantics, and other kinds of qualitative representations of the world in a natural way. I have mentioned Chomsky's work on grammar, which used tree-structures and Transformational Grammar. In effect, a Transformational Grammar is a mathematical system that allows tree-structures to be transformed into other tree-structures. With this property, the Transformational Grammar becomes equivalent to a Turing Machine (see p. 177)—that is, it can then be shown formally to be capable of computing any mathematical function. Transformational

Grammar is one language that is capable of expressing an important set of facts about natural language that cannot be expressed by a less powerful system. But there are other languages of equal computational power, many of which are more convenient than Chomsky's grammar for performing certain symbolic operations. These languages were developed in the 1950s and '60s, based on a mathematical theory about tree-structures called "recursive function theory." The most popular computer language based on this theory is called LISP (for List Processing language).

The basic idea of LISP is very simple. A tree can have any number of branches, and the labeled intersection of each branch is called a node. LISP provides a set of commands able to trace along the branches of a tree to find the nodes, and the labels at these intersections can direct the program either to another node or to another tree. Thus, the syntax of a sentence can be modeled in LISP, just as it is in the "sentence diagrams" that most people learn in elementary school. But trees can be related to other trees, and this capacity raises the formal power of LISP to that of a Transformational Grammar (equivalent to a Turing Machine).

Suppose now, that we want somehow to represent the meaning of a sentence in such a way that we can model the effects found by memory psychologists. If we wish to represent the meaning of the word "buy," it is clear that we must relate it to a number of similar meanings: For example, there is reciprocity between the words "buy" and "sell," which can be shown in a LISP format. Similarly, we would want to relate the word "buy" to the words "buyer," "seller," and "medium of exchange," and to a subjective assessment of value by the buyer and seller, and so on. In a natural way, a very dense network of relationships begins to emerge between all these components, which can be symbolized by labels on a complex set of tree-structures. Such semantic networks are quite effective in representing meaning structures so that a computer containing the semantic network can answer questions about the subject of buying and selling, generate semantically equivalent paraphrases, relate one part of a discourse to another, retrieve relevant facts, make inferences from the facts given, obey commands stated in ordinary language, solve certain problems, and even generate some analogies.

This is really the proof of the pudding, of course. One can make a plausible case that a language such as LISP is a natural way to encode human knowledge, but this argument will not convince many people unless one can show its effectiveness in particular cases. That has been done abundantly. LISP has been used to model parts of human knowledge in a variety of domains, to parse and generate sentences, recognize speech, play chess and other games, simulate the behavior of a paranoid patient in a mental hospital, generate pictures, and control the movement of a robot (Boden, 1977). All this does not mean that LISP is the ultimate language for the representation of psychological facts, but it does mean that it is far more effective than any previous theoretical medium. And of course, LISP is not at all a *quantitative* language. LISP makes it easy to represent sentences, but hard to represent algebraic equations. Yet it is entirely precise from a mathematical point of view.

Psychologists have found it difficult to deal with a mathematically precise but nonquantifiable theoretical language. All experimental proof-procedures available to psychologists were aimed at testing quantitative hypotheses. Further, any system that simulates an intelligent human ability must be enormously complex and interactive. In fact, it is a behaviorist's nightmare: Artificial Intelligence theories contain thousands of theoretical entities, all interacting with each other at enormous speed. There simply were no experimental methods for assessing the empirical implications of such a complex system, but it seemed that nothing less complex and symbolic would do the job of simulating interesting human functions. Thus, empirical psychologists were caught in a bind: The minimum system needed to model the process of understanding natural language is enormously complex, far beyond the proof-procedures available; and yet something at this level of complexity was needed to do the job. (Neurophysiologists are probably not so distressed by this, given that the nervous system has a comparable awesome level of complexity.) In this situation, it is perhaps surprising that psychologists were receptive to Artificial Intelligence work at all.

Not all cognitive psychologists even today are receptive to Artificial Intelligence (*e.g.*, Neisser, 1963b). Even those who are not, still profit from the "breathing space" created by the computer for psychological theorizing. The capabilities of computers makes it easier to propose all kinds of theories. Of all cognitive psychologists, perhaps the most uncomfortable with Artificial Intelligence are the strongest empiricists: those who still maintain the suspicion of theory inherited from behaviorism. They will continue to insist on direct empirical proof, and this is of course all to the good. There is widespread misunderstanding of this "computational metaphor" for psychology, as well as some intense disagreement from those who are familiar with it (Weizenbaum, 1976; Neisser, 1963b). Psychologists are often said to believe that people are like computers, but this is a crude way of stating the case. There are, in fact, a number of useful levels of the computational metaphor, some more conservative than others.

First, and most conservatively, there is the claim that any precise psychological theory can be modeled on a computer. This is almost certainly true. The modern digital computer has its limitations, of course: It is only a very fast, serial device for executing entirely explicit, discrete instructions. Thus, it would seem unable to represent situations that require parallel operations, or analog (as opposed to digital) operations, or those that are imprecise. But all of these limitations can be overcome. Digital computers can simulate analog processes to any desired degree of precision. The serial operating systems of these computers can also simulate parallel processes, and in recent years a great deal of very important work has been done with parallel "distributed" systems, which are networks of ordinary computers. And, finally, the representations maintained by computers do not even have to be entirely explicit: They may be probabilistic or make use of a logic called "fuzzy set theory" (Zadeh, 1975). In sum, there is no reason at this point to believe that computers have *principled* limitations in representing interesting psychological facts.

A stronger, less conservative, claim about the computational metaphor for psychology is that *computers and humans are both information processors*. (This claim is not inherent in the previous point that computers can represent interesting things about people—computers can also represent interesting things about houses, but no one claims that houses are information processors.) We have previously cited some reasons to think that the human nervous system works to process information—that it is not sensitive to physical energy as such, but only to information; that it tends to treat different physical events as identical if they are abstractly the same, and so on.

Perhaps the strongest version of the computational metaphor claims that when computers are confronted with tasks similar to those that humans must solve, they are driven to the same kinds of solutions. In analyzing speech, computers must use both top-down and bottom-up processes, just as people apparently are obliged to do. In input and output, both computers and people require buffer memories in order to relate events arriving at different times. A number of such similarities have been found, but no one is likely to take them for granted without investigating the human side of the analogy with great care.

Any of these reasons are adequate justification for the psychological interest in Artificial Intelligence. This is a very significant development in the history of psychology, but it raises some troubling questions as well. As a young science, psychology has not yet had to face the possibility that psychological knowledge might be misused. If psychological theory achieves some genuine level of adequacy at some point in the future, this possibility will have to be faced.

Behavioristic Responses to the Cognitive Revolution

There has been surprisingly little in the way of controversy in the cognitive revolution. No doubt many behaviorists felt opposed to the new approach, as the interviews in Chapter 3 indicate, but not much of this opposition appeared in print. As I mentioned, the single great public controversy took place as a result of Noam Chomsky's negative review of B. F. Skinner's book *Verbal Behavior* (1957). In his autobiography (1976, 1979), Skinner tells us how this attempt to provide a detailed behavioristic account of language was to be the ultimate achievement of his life's work, a project on which he had worked for almost two decades before it was ready to appear in print. In view of his original training as a serious writer of "psychological" short stories, we may guess that Skinner's attempt to explain language in terms of operant conditioning had a great personal as well as scientific significance.

Skinner considered language to be essentially a chain of responses emitted by a speaker when the environment called them forth. His theory holds that children learn language by a process of trial-and-error conditioning, facilitated by the community of speakers surrounding the child. To some extent, verbal responses would be differentially reinforced by adults, so that "correct" utterances would gain the child attention, hugs, and other reinforcing stimuli. As these utterances were reinforced, they would tend to grow in frequency.

I have already indicated that by the time Skinner's book was published in 1957, the world had already changed. Noam Chomsky, then a young and still obscure linguist at MIT, was in the process of gaining a reputation for a new and disturbing approach to language, which claimed to prove that all contemporary theories of language were grossly inadequate to explain even the simplest and most obvious facts. Thus, when Skinner's book was published, it became grist for Chomsky's mill. Asked to review *Verbal Behavior* for the influential journal *Language*, Chomsky attacked Skinner's theory with the kind of vigor and theoretical perspicacity that has seemed to be a specialty in linguistics (Chomsky, 1959). In effect, he argued, Skinner's work had nothing to do with language, because Skinner knew nothing about language as such. Skinner treated

language simply as another instance of the kind of behavior that could be observed and manipulated in animal experiments. But language was demonstrably *rule-governed* and *generative* (in that an infinite number of new sentences could be generated from the grammar of a language), and to understand language one had to postulate the existence of a powerful theoretical construct, a Transformational Grammar, which was sensitive to the regularities *underlying* the immediate observable "surface" facts of language. Any sentence in a language and, indeed, any natural language as a whole, was only a phenotypic instance of an underlying genotype, just as any organism is only a particular realization of an underlying genetic plan. Indeed, Chomsky argued, there is much reason to think that our capacity for language is not learned in any simple sense, but that it is part of our biological inheritance. Needless to say, according to Chomsky, Skinner's theory failed in all respects to account for these facts. Even further, no imaginable behavioristic theory could handle such facts, for much the same reasons that Skinner's theory failed: Behavioristic theory was averse to postulating rules, because rules are abstractions, and behaviorism was committed to the analysis of concrete, observable things. Generativity of responses was not acceptable to behaviorists, because responses were thought to be paired with stimuli through a process of conditioning. The deep rule-structure underlying surface sentences could not be observed directly and, hence, was considered to be unscientific. In almost every respect, Chomsky's arguments contradicted the behaviorism of Watson, Skinner, and Hull. His facts were not collected in the laboratory using precise experimental designs and operational definitions, but came simply from everyone's knowledge of their native language. Moreover, the idea that language capacity could be largely inherited ran counter to the behavioristic emphasis on environmentalism.

Chomsky's review of Skinner's *Verbal Behavior* had great impact among linguists and psychologists, far more so than the work it attacked. At just about this time, a new field called "psycholinguistics" was taking shape—a field created to study language by combining the efforts of psychologists and linguists in one coherent domain. Although psycholinguistics began behavioristically, it was open to new ideas, and soon cognitive thinkers were taking over. Psycholinguistics provided the first ground for collaboration between Chomsky and perhaps the single most influential figure in the cognitive revolution, George A. Miller. Miller, whose interview appears in the next chapter, propagated Chomsky's views among scientific psychologists. In a joint paper, Chomsky and Miller (1958) showed that it was impossible for people to acquire by conditioning all the sentences they could speak and understand—indeed, people appeared able to understand more sentences than there are seconds in a lifetime! Clearly these sentences could not be learned one by one, whether by conditioning or by any other means. People must have a relatively small set of rules that enables them to put together many different sentences.

Skinner never replied to Chomsky in print, although the two did meet in public debate on several occasions. In any case, they had little to say to each other (see their interviews, Chapters 3 and 7, respectively). Their theories were so different, the kinds of evidence they considered had so few common elements, and their methods were so alien to each other, that genuine communication appeared to be impossible. In fact, their attempts to communicate illustrates graphically one of Thomas Kuhn's claims about scientific revolutions: that even with the best intentions, adherents of different paradigms are not able to reason out their differences—they simply talk past each other.

Some of Skinner's followers answered Chomsky in print (*e.g.*, McCorquodale, 1970; Salzinger, 1970). Some years later, Skinner himself wrote an interesting article called "Why I Am Not a Cognitive Psychologist" (1977) which I discussed earlier. These replies came too late to be effective. By this time—the late '50s and early '60s—the cognitive point of view had already emerged in many different places, for many different reasons, not all of them inspired by Chomsky. Indeed, one of the striking aspects of the cognitive revolution was this sudden emergence of common insights among people who were socially and geographically far removed from one another. When water turns to ice, crystal formation begins in many places at once, and the transition to the solid state is quite sudden. We cannot point to one place in the liquid as *the* starting point of the phase-change from liquid to solid—the phenomenon is quite literally a function of the whole. There is an appealing analogy between the cognitive shift and such a change of phase in a physical liquid.

There have been occasional attempts to bridge the gap between behavioristic and cognitive points of view, but these attempts at integration have not been very persuasive. Some psychologists tried to develop an intermediate position, but were ultimately driven to believe either one or the other, and it soon appeared that most of the active researchers in human psychology were moving to a stronger and stronger cognitive position. That trend has continued for several decades, and at this writing it has not yet stopped.

The Liberal-Conservative Continuum: Five Cognitive Psychologists

Five individuals may be said to typify the new psychology. All five can be called cognitive psychologists, but they differ greatly in the degree to which they accept cognitive premises. From the most to the least conservative, they include the following: *Charles E. Osgood* attempted to adapt Hullian behavior theory to the question of linguistic meaning in the 1950s, and has held firmly to this position ever since. *James J. Jenkins* began with a major research program designed to test an intermediate theory of language learning, appropriately called "mediational theory". Finding his empirical results disappointing, he moved toward a more Chomskyan point of view. *George A. Miller* called himself a "good behaviorist" in a book published in 1951, but soon afterwards took the lead in guiding other psychologists toward an increasingly cognitive position. *Jerome S. Bruner* was clearly too impatient to wait for others to catch up, and as a result probably touched on more interesting issues in his career than any number of other psychologists, at the cost of some loss of influence. And finally, *Herbert A. Simon* started outside of psychology proper, and began doing work in Artificial Intelligence 20 years before it became popular; he has since seen the field move more and more in his direction.

Osgood was deeply concerned from a very young age with the issue of meaning in language (1975). As a good experimentalist, he saw the question of linguistic meaning as essentially a problem of measurement. If scientific psychology was to address the issue of meaning in a respectable way, it would have to be operationalized, brought into the laboratory, and best of all, quantified. This is, of course, the familiar emphasis on quantification derived from the physical sciences, and it does not appear that Osgood considered that meaning might not be *naturally* quantifiable in the usual sense of the word, or that one could be mathematically explicit about meaning without using numbers. But in the 1940s and '50s, few people considered this possibility seriously (but see Brunswik, 1946). Within the limits so defined, Osgood's work was remarkable: He did, indeed, manage to develop a well-specified measure of something much like meaning, but not the kind of meaning we speak of when we say "the meaning of this word or sentence"—rather, Osgood's work seems to deal with *connotative meaning.*

Before Osgood's work, behavioristic psychologists interested in the question of meaning were much concerned with a phenomenon called "semantic generalization" (Razran, 1961). Generalization refers to the sentation in human psychology, a level that need not be directly observed, but helps to make sense out of a host of phenomena that cannot otherwise be understood (*e.g.*, Miller & Johnson-Laird, 1976).

As a conscientious experimentalist, Charles Osgood could not permit himself such theoretical freedom in the 1950s. People who exercised such freedom at the time were simply not considered to be scientific, and indeed, some had to pay the price of exclusion from the community of "serious" psychologists for this very reason.

James J. Jenkins is also very much an experimental psychologist, though his career differs from Osgood in interesting ways (see interview in Chapter 5). Jenkins attempted to perform a rather heroic series of experimental tests of the mediational theory of meaning and syntax. His interview describes these experiences well, and the reader is referred to it for details. Jenkins's attempt to test mediation theory failed in certain critical ways. In the face of this failure, Jenkins was able to stand back from the standard behavioristic framework, and reexamine the issues from the ground up. In his ability to keep changing with the field and discover a fruitful *modus vivendi* with the new point of view, Jenkins's work was remarkable. Yet his research remained essentially experimental, and, thus, in the rough conservative–liberal dimension, he belongs somewhere between Osgood and a somewhat more theoretical psychologist, George A. Miller.

Miller's work has been mentioned already in connection with the linguist Chomsky, who triggered a cognitive revolution in linguistics at roughly the same time that the cognitive revolution took place in psychology. Miller began as a psychophysicist at Harvard with a strong mathematical background. On three different occasions he imported into experimental psychology a quasimathematical framework for the consideration of other psychologists, and three times psychologists accepted his point of view. The first imported framework was *information theory*, a mathematical theory developed during the Second World War to describe the carrying capacity of communication channels such as telephone lines (Miller, 1964). Although this theory was and is of fundamental importance to computer scientists and psychologists, it seemed to have little to do with the kinds of problems faced by experimental psychologists. In particular, the mathematical definition of "information" seemed to circumvent the whole issue of representation—information was defined as a choice made between alternatives in some existing representation in a message-receiver, as a consequence of the same choice signalled by the sender of the message. Insofar as the message caused the receiver to make the same choice, information could be said to have been transmitted.

But how was one to represent the alternatives to be chosen in the first place, especially when those alternatives approached the kind of complexity needed to represent real human concerns? This question was simply not dealt with. Thus, information theory seemed to beg the question of knowledge representation. Nevertheless, in the theoretically impoverished environment of the 1950s, it was a step forward for psychologists to learn something about information theory.

Next, Miller discovered Noam Chomsky. Although he was a linguist, Chomsky's criticisms of the kind of theoretical mechanisms used in behavioristic linguistics also applied to behavioristic psychology. Further, psychologists found some experimental evidence in favor of Chomskyan ideas, though the experiments were only suggestive. As a result of Miller's espousal of Chomsky's ideas, language psychologists began to study whole *sentences*—prior to this, they had focused only on single words, and on the connections between words. But if language was to be understood in terms of a set of rules able to generate sentences, then sentences were clearly the proper units of study. It was not long before Chomsky's theory passed from the scene as a viable psychological theory of language, but here again, the very act of thinking about a genuine theory which did not apologize for itself, had a favorable impact: In the process of considering and rejecting such theories, psychologists were becoming much more theoretically sophisticated.

Miller had led the field toward information theory and Chomsky's Transformational Grammar, but he was a bit late in arriving at the next major step—the use of full-fledged computational theories. Others were there before he moved "Toward a Third Metaphor for Psycholinguistics" (1974), in which the perception, production, and acquisition of language was viewed in terms of symbolic information processing. This is still the dominant metaphor in the psychology of language, and it seems likely to last for some time. The computational metaphor is very theoretical, and even in this third metaphor, Miller was able to make an original contribution (*e.g.*, Miller & Johnson-Laird, 1976). But the kind of leadership that he supplied to psychologists before—essentially the courage to consider abstract mathematical systems that might have some relevance to psychology—was no longer needed. From being the major theoretical leader in the field, Miller had by 1974 become one of many leaders.

Miller's career contrasts with the career of his friend and colleague, Jerome S. Bruner, who is well known outside of psychology proper, especially in the field of education. Bruner is a man of great charm and brilliance, as are many of the others disucssed here, and has made major contributions in many different parts of psychology. Yet his influence has been surprisingly attenuated, especially in the most rigorous circles in experimental psychology. The impression one gets is that Bruner simply did not have the patience to wait for the rest of the field to catch up. Always 10 years ahead of the community, he was not recognized and relied on in the way Miller was. Conversely, one suspects that George Miller may have sometimes adopted a point of view in which he did not really believe, simply in order to stay in touch with the rest of the experimental community. This is admittedly speculative, but some of this book's interview participants hint at something along these lines.

The most "radical" cognitive psychologist discussed here is Herbert A. Simon. By present standards, his radicalism is well within the bounds of modern cognitive thinking, but for most of his career, although the psychological community was moving in his direction, it moved very slowly, and he was considered an outsider until rather recently. Herbert Simon is currently the only cognitive scientist who has won the Nobel Prize—or rather, he is the only Nobel Laureate who considers himself a cognitive scientist. (There is no Nobel Prize in psychology, of course. Simon received the prize for his work on the economics of the firm, which he considers to be cognitive science, although the Nobel committee probably conceived of it as pure economics.)

Because he was not trained in the behavioristic point of view, Simon never felt theoretically constrained by it. He was familiar with formalisms in other sciences, and when the opportunity presented itself he felt free to participate, in collaboration with Allen Newell and others, in the development of the key ideas now used with great success in Artificial Intelligence. Thus he has helped to make major contributions in economics, computer science, and cognitive science. But for most of his career, Simon was clearly not considered to be a psychologist by most psychologists. (For this reason, I present the interview with Herbert A. Simon in my chapter on "Nucleators," people who created nuclei of psychological interest outside the self-defined boundaries of psychology at the time, and whose work forced those boundaries to expand.)

Thus we have a continuum from Osgood, who made a strong attempt to apply Hullian behaviorism to a cognitive problem; to Jenkins, who found that the mediational theory would not work, and was able to move to a new theoretical perspective; to Miller, who introduced psychologists to three new theoretical approaches, staying always close enough to the field to be understood, but far enough ahead to be followed; to Bruner, who made a number of remarkable contributions, but was too far ahead to be fully trusted by many psychologists; and to Simon, clearly an outsider who made his major contributions in the guise of economics and computer science, when in fact he saw his work very much as theoretical cognitive science.

Organization of the Cognitive Interviews: Adapters, Persuaders, and Nucleators

All of our cognitive interview participants can be placed on this continuum of theoretically conservative versus liberal thinkers. In the following chapters, however, their interviews are grouped in a slightly different way. Chapter 5 contains interviews with *Adapters*, those psychologists who were able to adapt to the revolutionary change in psychology—from a strong professional commitment to behaviorism to an equally strong commitment to cognitive psychology: George A. Miller, Marvin Levine, George Mandler, and James J. Jenkins.

Chapter 6 is about *Persuaders*, cognitivists who were never behaviorists, and who view their role as a matter of persuading the rest of the field to give up an inadequate approach. Whereas the Adapters may have had to face a crisis of personal change, the Persuaders confronted an indifferent or perhaps hostile scientific community, and they had to maintain their independent perspective regardless of the resulting intellectual and social pressures. This group of interviewees consists of Ernest R. Hilgard, Ulric Neisser, Walter B. Weimer, and Michael Wapner. With the cognitive shift in psychology, the Persuaders have not been satisfied to accept membership in the cognitive community. Rather, they have continued to evolve new perspectives, and now present the cognitive community with a set of challenging new ideas. They have not ceased in their attempt to persuade the psychological community of the need for change—they have only shifted their ground.

The third and last group of cognitivists I call the *Nucleators* (Chapter 7). The word "nucleator" is borrowed from physics, where it signifies a particle that serves as a nucleus for a cluster of new particles, much as small dust particles in the atmosphere can trigger the formation of ice-crystals to produce snowflakes. By analogy, a scientific Nucleator serves as a nucleus for scientific activity outside the boundaries of the conventional scientifc community. These individuals came of age intellectually before the cognitive shift, outside of what was then considered to be psychology. They came from engineering (Donald A. Norman), from mathematical linguistics (Noam Chomsky), philosophy (Jerrold A. Fodor), or, in the case of Herbert A. Simon, from a combination of economics, political science, and computer science. According to Kuhn (1962), it is common for outsiders to have a disproportionate impact at a time of scientific revolution, and it is not surprising that these individuals have been extraordinarily influential. Some psychologists may not consider Fodor, Chomsky, and Simon to be "true" psychologists, but it is interesting to note that in their own thinking, all these individuals considered themselves to be "doing psychology" from very early on. Sometimes it is unclear whether it is Mohammed who comes to the mountain of psychology, or *vice versa*. But there is no doubt that these Nucleators have enormously enriched scientific psychology by the gifts they have brought from their respective fields.

The Adapters: Psychologists Who Changed with the Revolution

Our first group of cognitive psychologists started their professional live as behaviorists and changed their perspective along with the mainstrean of the research community. George A. Miller, Marvin Levine, Jame J. Jenkins, and George Mandler all began by considering themselves as behaviorists, although we can always detect some nonbehavioristic influences in their graduate training. Each performed state-of-the-art work for some time during the cognitive shift, and each represents an important theme in the shift.

George A. Miller indeed represents not one but several significant themes: the early role of mathematical psychology and information theory; the rediscovery of short-term memory; the influence of Noam Chomsky in demonstrating that behavioristic theories of language were inadequate *in principle* (see Chapter 7); early influences from computer simulation studies along the line of Simon and Newell; and more recent work on advanced topics in cognitive psychology, including a computational theory of language and perception (with Philip Johnson-Laird), studies in metaphor, and studies of word meaning. Miller has led significantly in all of these areas, making an extraordinary range of contributions. In spite of his continuing leadership, Miller's work has been relatively free from controversy; indeed, his prestige has been so well established that his involvement in a new problem has often served to signal other psychologists that some previously taboo topic was now safe for respectable researchers.

Marvin Levine is probably the most behavioristic of the cognitive psychologists represented here. In fact, his work has been consistently acceptable to both behaviorists and cognitive [illegible] work focused on mathem[illegible]