1 History and core themes

Adele Abrahamsen and William Bechtel

1.1 Overview

Two characteristics of cognitive science are central and obvious.

First, it is *cognitive*, aiming toward empirical and theoretical understanding of cognition. Its founding disciplines addressed intelligence in humans (cognitive psychology) and computers (artificial intelligence), with special scrutiny given to language as a paradigmatic domain of human competence (linguistics). Over time, the understanding of what it is to be cognitive has expanded, diversified, and become more contentious.

Second, it is *interdisciplinary*: ideas and methods of inquiry propagate across traditional boundaries, and collaborations thrive among the founding fields and also philosophy, sociology, anthropology, developmental psychology, education, and neuroscience. Some of these collaborations have created or reinvigorated interdisciplinary fields such as psycholinguistics, language acquisition, linguistic anthropology, cognitive sociology, computational linguistics, and cognitive neuroscience; others have contributed research strategies, especially computer simulation of mental activity; and many more have contributed to particular strands of inquiry.

Beyond these, there are few if any core characteristics embraced by all cognitive scientists. Instead, there are themes that emerge as important in particular eras or approaches and also some dichotomies that unify advocates on each side but divide them from each other. To get more of the full story, see the Further Reading at chapter's end. Here we situate the mutable themes of cognitive science within a brief historical tour, organized as follows:

- Cognitive science has deep roots in several fields, but the most relevant advances were largely abandoned in the United States during the first half of the twentieth century as psychology became behaviorist, linguistics became structuralist, and what later became known as neuroscience was limited by the available methods and by anti-localizationist leanings.
- Key innovations in the 1940s – the idea of information and the advent of electronic computers – gave rise to new fields, such as information theory, artificial intelligence, and artificial neural networks, and set the stage for

interdisciplinary ferment and a "cognitive revolution" in psychology and linguistics.

- By the late 1950s pioneers in several fields were embarking on a vigorous pursuit of what is now called symbolic architecture, in which representations and the rules specifying operations on them consist of discrete symbols. This took the form of information-processing models in psychology, computer programs in artificial intelligence, generative grammar in linguistics, and the language of thought hypothesis in philosophy. In this volume, there is more extensive discussion of representation in Chapter 2 and of organized operations ("procedures") in Chapter 3.
- By 1975 "cognitive science" had a name, but its focus on symbolic rules and representations was challenged when some key cognitive scientists took another look at artificial neural networks and adapted them to obtain statistical, "subsymbolic" connectionist models in the early 1980s. Ever since, symbol manipulation and statistical approaches have offered quite different insights into perception, action, learning, memory, reasoning, decision making, concepts and language – aspects of cognition covered in Chapters 4–9. Moreover, certain alternatives to generative grammar – especially cognitive linguistics and optimality theory – had an impact beyond linguistics.
- Since the 1990s cognitive science has expanded in ways making it even more diverse and increasing the salience of previously peripheral fields, as discussed in Chapters 10–15.

Throughout this history, philosophy has been a player at arm's length from day-to-day empirical research. However, its concepts, theories, and tools often get adapted or applied by researchers in other cognitive science disciplines, and philosophers reciprocate by collaborating and by probing those disciplines. Moreover, philosophy of mind has provided an ongoing forum for interdisciplinary conversation and inspired certain lines of empirical research.

1.2 The roots of cognitive science

Theoretical inquiries into mental phenomena date back at least 2,500 years, but substantial lines of empirical research first developed in the nineteenth century – an era in which disciplinary boundaries were being established as universities grew dramatically. Contributions that are still relevant today were made by European researchers trained in biology or physics, notably Weber's and Fechner's psychophysical laws, Helmholtz's and Hering's accounts of color perception, Donders' techniques for inferring mental processes from reaction times, and (on a wider canvas) Darwin's insightful observations and theory of evolution. Certain philosophical frameworks were influential, including J. S. Mill's "mental chemistry" and Brentano's introspective analyses of mental acts and their "intentionality" (being about something). These strands converged in a new discipline when two young scholars established psychology laboratories and began publishing in the 1870s. At Harvard University, William James arrived at rich characterizations of consciousness, memory, habit, sensation, emotion, and other mental functions. In Leipzig, Wilhelm Wundt's great breadth, ranging from neural to cultural investigations, precluded a unitary legacy. Most students pursued either his program of experimentation or his use of systematic introspection for inferring mental content. By 1900 psychology was a thriving discipline, complete with factional disputes.

Research emphasizing the physical facets of the mind/brain followed a separate trajectory. The idea that different brain areas subserve different functions (localizationism) goes back at least to the early nineteenth century, when Gall's *phrenology* was built on apparent associations between size of particular brain areas (as inferred from the skull) and differences in ability to recall words, perform music, show kindness, and so forth. There was popular uptake but strong scientific opposition. By the 1860s, influential autopsy studies by Paul Broca had associated lesions with patients' loss of articulate speech (*Broca's aphasia*). In the early twentieth century, anatomist Korbinian Brodmann was able to use the layout and layering of different types of neurons to map areas of neocortex precisely enough that his numerical designations are still in use. Concurrently, neurophysiologists began using behavioral effects of lesions and electrical stimulation in animals to successfully identify brain areas for vision (striate cortex), motor control (precentral gyrus), and finally other sensory systems. Their attempts to localize memory or other higher functions in parts of the large "association cortex" that remained were less well received. Karl Lashley famously concluded that this brain tissue exhibited (1) *equipotentiality*, the ability to take on different functions (e.g., following damage); and (2) *mass action*, in which it is the total area available, not location, that matters. Today localizationism dominates neuroscience, ranging from single-cell recording to functional magnetic resonance imaging (fMRI), while anti-localizationism has a new home in dynamical systems modeling. Cognitive science encompasses both.

The prevalence of anti-localizationist views like Lashley's through the first half of the twentieth century minimized neurophysiology's influence on psychology. However, another part of physiology – Pavlov's studies of salivation in Russia – helped to shape an entirely new school of thought that came to dominate psychology in the United States during that period. *Behaviorism* originated in John Watson's growing conviction that both Jamesian functionalism and Wundtian introspectionism suffered from insufficient empirical grounding and objectivity, largely in consequence of their focus on mental activity or contents. Once Watson learned of Pavlov's classical conditioning (in which, e.g., a bell repeatedly paired with food can itself elicit salivation), he embraced it as a tool by which psychologists could obtain objective accounts of observable behavior. Soon he had narrowed psychology's focus to learning and promoted the use of animals as model organisms. In mid-century B. F. Skinner championed operant conditioning (in which reinforcing an act brings it under control, e.g., increasing the rate at which a rat presses a bar). Although Skinner was also known for his *radical behaviorism*, which repudiated appeals to inner states in explanations of behavior, during the same period Watson's approach was further developed by the less restrictive *neo-behaviorists*. In particular, Clark Hull allowed intervening variables for *drive* and other unobservables in his influential mathematical laws of learning, but he stopped short of overtly mentalistic constructs such as memory, attention, or representation. Edward Tolman went that extra step by positing cognitive maps in explaining the navigational behavior of rats (making him a maverick), and other less behaviorist research paths were pursued in psychophysics and parts of developmental, social, and clinical psychology.

Outside the United States, behaviorism had little impact on psychology. A variety of approaches that emerged in the 1920s and 1930s continued to advance in the UK (e.g., Sir Frederic Bartlett's appeal to schemata to explain memory distortions), Germany and Austria (Gestalt psychology's emphasis on organized wholes), Switzerland (Jean Piaget's genetic epistemology, in which schemata develop into stably organized systems), and the Soviet Union (Lev Vygotsky and Alexander Luria's studies of language and thought). These proceeded independently until all found a degree of uptake in cognitive psychology in the 1960s.

One other discipline with deep historic roots played a major role in early cognitive science: linguistics. As far back as the eighth century BCE, Panini systematically described the phonology (sound structure) and morphology (word structure) of Sanskrit, and touched on its syntax (sentence structure). His counterparts in the first half of the twentieth century were the *structural linguists*, many of whom showed an affinity with behaviorism. Leonard Bloomfield, for example, was influential in his insistence on cataloguing and analyzing only directly observed speech.

There were two periods of especially fertile interaction between psychology and linguistics. Around the turn of the twentieth century in Europe, certain linguists and psychologists shared an interest in mechanisms of change based on analogy or association. Other linguists were influenced by two of the many strands of Wundt's work: his emphasis on holistic creative mental processes and his use of tree notation to convey grammatical structure as hierarchical (vs. the flat but less constrained structures implied by associationism). The second period of interaction began in the early 1950s in the USA and relied on empiricism as a common ground between neo-behaviorist psychologists and structural linguists. Participants in a 1953 summer seminar reintroduced the term *psycholinguistics* and set an ambitious agenda for cooperative research. Although many of the specific theories from both linguistics and psychology were soon to be replaced in the cognitive revolution, the major research goals they set, such as establishing the psychological reality of *phoneme* and other linguistic constructs, would continue to be pursued.

1.3 Information, computation, and the cognitive revolution (1940–1956)

It is sometimes said that the cognitive revolution stemmed from seizing on a new technology – the digital computer – as a metaphor for the mind. This indeed was the dominant metaphor by the 1960s, but earlier technologies – the telegraph and telephone – provided a transitional metaphor in which certain human systems were likened to electronic communication channels. Formal engineering analyses of communication, especially in the 1920s to 1960s at Bell Laboratories, gave rise to a cluster of fields which changed the intellectual landscape. Those making the greatest impact on psychology included *information theory* (a quantitative treatment of information transmission across channels subject to capacity limits, rate limits, and noise) and *coding theory* (concerned with the form of the message, especially ways of recoding it into a compressed format – such as today's MP3 – that can be transmitted and stored more efficiently). Ignoring semantic content, a message was taken to be informative to the extent that it reduced uncertainty: the more possible messages, the greater the uncertainty. Claude Shannon (1948) introduced the *bit* (from *bi*nary digi*t*, 0/1) to quantify this; for example, if there are two equally likely messages, one bit is sufficient to distinguish them, but eight messages (2^3) require at minimum a three-bit encoding.

George Miller began a long, influential career in cognitive science by bringing Shannon's framework to bear on data he collected for his 1946 PhD dissertation in S. S. Stevens' psychophysics laboratory at Harvard. Specifically, he showed that spoken English messages that were most intelligible in noise were those that were more redundant – that is, those requiring fewer bits of information to narrow the interpretations to one. A small stream of research in what Miller called *statistical behavioristics* ensued. Within the next decade the computer metaphor began overtaking the communication metaphor, ultimately favoring a less statistical notion of information as mental content held in computer memory-like storage registers and manipulated by program-like processes. Both metaphors gave rise to *information processing* by offering engineering-based ways to open the "black box" between stimuli and responses and model mental activity. Two communication-based accounts in the 1950s helped shape the research paradigms and computer-based models of the 1960s. First, British psychologist Donald Broadbent posited multiple sensory channels, each with a memory buffer feeding into a central attentional filter that selects (and can switch) which channel's input gets sent through a limited-capacity information channel for further processing. His research paradigms and use of flow charts were as influential as his model. Second, Miller (1956) himself offered "the magical number seven plus or minus two" as the capacity limit of immediate memory (Broadbent's information channel, redubbed *short-term memory* in the 1960s) and proposed that humans often deal with this by "chunking" incoming information. As a simple example, 149217761860 exceeds capacity unless we recode the twelve digits as three familiar dates (1492, 1776, 1860). Miller presented this work to a Symposium on Information Theory at MIT on September 11, 1956.

Two other papers at the same symposium (by Noam Chomsky and by Allen Newell and Herbert Simon) advanced the computer metaphor by focusing on symbolic rules and representations. The roots of this approach lie in symbolic logic as formulated by Frege and further developed by Whitehead and Russell near the turn of the twentieth century; we use a simpler formulation to provide a glimpse. First, in *propositional logic*, symbolic expressions composed of propositions and connectives, such as $\vee$ ("or"), $\wedge$ ("and"), and $\neg$ ("not"), can be derived using rules of inference such as:

$$\mathrm{p} \rightarrow (\mathrm{q} \vee \mathrm{p})$$

where p and q are any two propositions, $(\mathrm{q} \vee \mathrm{p})$ indicates that q or p or both are true, and $\rightarrow$ ("implies") indicates that if the expression on the left is true, then the expression on the right must be true. Second, in *predicate logic*, propositions are replaced by predicates (F, G, . . .) taking one or more arguments, each of which is a constant or a variable (x, y, . . .) that can be bound by the quantifiers "for all" ($\forall$) and "there exists" ($\exists$). For example:

$$\forall x\ \forall y\ \mathrm{F}xy \rightarrow \forall x\ \forall y\ (\mathrm{G}xy \vee \mathrm{F}xy).$$

A more immediate influence is automata theory, a mathematically rigorous exploration of virtual machines for computation. A *finite state automaton* takes as input symbols from a finite set; each of its rules specifies transition to its next state based solely on its current state and current symbol (i.e., "if state A and symbol S at time t, then state B at time $t+1$"). In the 1930s Alan Turing proposed a more powerful type of automaton with an indefinitely extendable tape holding symbols. Each rule specifies, based on the current state and symbol, a state transition and also actions with respect to the tape (writing or deleting a symbol, moving left or right) – actions that amount to adding a memory. This abstract class of *Turing machines* influenced John von Neumann's design work on the overall architecture that has dominated computer design for decades and, in turn, cognitive scientists' conceptions of mind and language.

Chomsky had the revolutionary idea of construing the grammar of a natural language as equivalent to an automaton capable of generating the sentences of that language – a *generative grammar* – and asked what sort of grammar would be adequate. He used the 1956 symposium and the book *Syntactic Structures* (Chomsky 1957) to persuade key thinkers beyond linguistics to accept the question itself and his answer. As explained in Chapter 9, he concluded that a transformational grammar was required. By 1965 he described this as a grammar in which *phrase structure rules* such as S → NP VP and VP → V (NP) generated a sentence's *deep structure* (a tree suitable for computing meaning) and then *transformational rules* altered it to obtain the *surface structure* (a tree suitable for computing how the sentence should sound). When assessed as automata, transformational grammars were shown to have the power of a Turing machine. However, Chomsky's furthest-reaching impact beyond linguistics was that deep structure trees and transformations offered a specific vision of how mental representations and operations might look.

The symposium paper by Newell and Simon offered another such vision, as realized in the first functioning computer program in the new field of *artificial intelligence (AI)*. The historical background overlapped with that of generative grammar, and they shared Chomsky's basic commitment to rules specifying operations on symbols. But Newell and Simon anchored their work to digital computers – physical realizations of the kinds of devices abstractly explored in automata theory. Shannon had shown in the late 1930s that electric switches could be arranged to turn one another on and off so as to perform arithmetic operations, and World War II made this a priority. The first general-purpose digital computer, ENIAC, was delivered in 1946. The first with the serial von Neumann architecture was EDVAC in 1949–51: in the computer's memory are stored programs, data, and the results of each processing step, and these communicate (at the next step) with the central processing unit that carries out computations. Just ten years after ENIAC, Newell and Simon (with J. C. Shaw) wrote the first AI program (Logic Theorist) in the first list-processing language (IPL) and had it running on a digital computer. The influence of symbolic logic was obvious in its task: discovering proofs for theorems in propositional logic.

Looking back on his excitement at the nascent symbol-processing approaches in linguistics, AI, and his own corner of psychology, Miller (1979) identified September 11, 1956, as the birthday of cognitive science. It is important to note, though, that the information sciences were making connections not only with these fields but also with neuroscience during the 1940s and 1950s. A key example is the joint work by neurophysiologist Warren McCulloch and logician Walter Pitts on formal networks of simplified neuron-like units (*McCulloch–Pitts neurons*). Each unit could fire or not at each time-step, based on whether the sum of its individually weighted excitatory inputs across connections from other units exceeded a threshold. In 1943 they showed that any logical function could be computed by a network with this kind of parallel architecture, and by 1947 they were designing networks to simulate real-life tasks like sensory-motor mappings in the superior colliculus. McCulloch also helped organize an interdisciplinary conference on *cybernetics* that thrived from 1945 to 1953. For Norbert Wiener (1948), who coined the term and defined the field of cybernetics, the central concern was the role of feedback in controlling natural and artificial systems and guiding them toward goals. Cybernetics did not endure as a unified movement, but sent splinters of influence into a variety of fields. Most notably *artificial neural networks* were vigorously pursued in the 1940s through 1960s and revived in the 1980s. They represent a counterpoint to discrete computation (the von Neumann computer architecture and the symbolic models it inspired). We return to them in a later section.

This period saw the introduction or increased salience of a number of dichotomies that were inherited by cognitive science. Among them are content (the meaning of a symbol) vs. form (the "shape" of a symbol); digital/discrete vs. analog representation (some of the earliest computer designs were analog, as are mental images); serial vs. parallel processing; symbolic vs. statistical/quantitative models; and artificial vs. human intelligence.

1.4 Building symbolic models (1956–1975)

What are now called symbolic architectures or models continued to develop in generative grammar and artificial intelligence, and by the 1960s they were reshaping the information-processing approach in psychology as well. A *symbol* is a discrete form (e.g., the word "stop" or a stop sign) that stands for (represents) something else. Symbolic architectures share a commitment to (1) *representations* whose elements are symbols and (2) *operations* on those representations that typically involve moving, copying, deleting, comparing, or replacing symbols. A *rule* specifies one or more operations (e.g., S → NP VP). Typically the result is a different representation which then triggers a different rule, and so on until no further rules apply. An organized rule sequence such as this may be called a *process*, *procedure*, or (in linguistics) *derivation*. In many fields structured representations (rather than flat symbol sequences) are involved. For example, grammatical rules provide a combinatorial capacity that is constrained but productive, yielding sentences ("The car should stop here") along with trees indicating their structure, but not word salad ("stop the should here car"). Overall, the rules and representations approach is *formal* in that rules focus on the form of symbols, not what they represent, and *computational* in that it involves the manipulation of discrete forms.

As it became once again respectable to inquire into the inner workings of the mind, a major challenge was to develop tools for characterizing information processing. Taking inspiration from Weiner's cybernetics, George Miller, Eugene Galanter, and Karl Pribram (1960) proposed to model purposive human action using hierarchically organized goal structures that were flexible and recursive and that repeatedly assessed their own success. This book marked a turning point in North American psychology.

Miller next collaborated with social psychologist Jerome Bruner in creating the Center for Cognitive Studies at Harvard, an important influence on a new generation of cognitive psychologists. Bruner's own work in the previous decade – the "New Look" – had emphasized that a person's internal values and expectations affected their perception of external stimuli, and in the 1950s he pushed further into the mind by examining strategies of concept acquisition. A frequent visitor at the Center was Ulric Neisser, whose landmark *Cognitive Psychology* (Neisser 1967) emphasized the constructive nature of cognitive processes and brought European frameworks (e.g., Broadbent's attention filter, Bartlett's and Piaget's schemata, and gestalt psychology) to bear on the new information-processing approach in North America. The book served both to provide a name for the new subfield and to initiate the next generation of students into it; by 1970 there was a journal of the same name.

Harvard was not the only university at which strong faculty–student collaborations produced rapid advances in cognitive psychology in the late 1960s and early 1970s. Several psychology faculty at Stanford University whose original training was in mathematical and behavioral approaches to learning, most prominently William Estes, made a brilliant transition to innovative experiments and mathematical and computer models that were increasingly cognitive. Here are just three examples: Richard Atkinson and Richard Shiffrin developed an especially influential model of memory processes in which flexible control structures were involved in converting information from sensory to short-term to long-term memory stores. John Anderson and Gordon Bower proposed a pioneering semantic network model of human associative memory (HAM) that was the forerunner of Anderson's (1983) architecture ACT* (Adaptive Control of Thought). And Roger Shepard and Lynn Cooper asked people to mentally rotate geometric figures; finding a linear relation between the amount of rotation required and reaction times, they argued for analog mental operations.

Also influential was the new University of California at San Diego (UCSD). George Mandler, known for his work on active organization of memory, became first chair of its psychology department in 1965 and hired three young cognitive psychologists: Donald A. Norman, David E. Rumelhart, and Peter Lindsay. They and their graduate students (Norman, Rumelhart, and the LNR Research Group 1975) developed models of word recognition, analogy, memory, and semantic interpretation of verbs, sentences, and even brief stories. Underlying much of the work was a computer-implemented semantic network model of memory (ELINOR) that brought together influences from artificial intelligence, psychology, and linguistics.

During the same period research on memory continued within cognitive psychology (see Chapter 6), yielding more detailed characterizations of sensory, short- and long-term memory and of recognition and recall processes, plus discernment of procedural, episodic, and working memory. Research on reasoning gained traction, e.g., by positing mental models (see Chapter 7). But the greatest impact on cognitive science came from the rapid development of artificial intelligence within the new discipline of computer science. Embarking on a four-decade collaboration that made Carnegie-Mellon University a major incubator of cognitive science, Newell and Simon followed their landmark Logic Theorist program with General Problem Solver (GPS) in 1957. Both were written in their innovative Information Processing Language (IPL), which stored symbols in a *list structure* (i.e., one item was linked to another by specifying at the first site the address of the second item). In GPS they added their powerful idea of a *production system* architecture, in which conditional rules operate on representations in working memory (e.g., if expressions X and Y are in working memory, delete X and add Z). Newell and Simon (1972) asked people to think aloud while solving a problem and incorporated some of the strategies (e.g., reasoning backward from a goal state) in their models. Another AI pioneer, John McCarthy, created LISP (*LIS*t *P*rocessing language) in 1958; it incorporated some of IPL's features and became a standard tool. At MIT, Marvin Minsky (1968) introduced a wider readership to LISP programs adapting predicate logic toward simulating semantic activities such as solving analogies, proving theorems, and answering questions. The simplest example is in F. Black's chapter: "Where is my pencil?" was represented as a predicate with two arguments, **at (pencil, *y*)**, and answers were deduced from stored statements, e.g., **in (pencil, desk)**, **at (desk, home)**. In an outlier chapter M. Ross Quillian pioneered a different format – *semantic networks* – that found uptake in the 1970s (initially in HAM and ELINOR).

A related endeavor, *robotics*, began the transition from science fiction to engineering project in the 1960s. Minsky's group designed a Blocks Micro-World in which robots must see and move blocks (not simply cogitate like most AI programs). At Stanford, Charles Rosen's group endowed "Shakey" with wheels, a TV camera, and control by rules akin to a production system. However, what many regarded as the most impressive research with a blocks world did not involve a robot. Focusing on natural language processing (NLP) at MIT in 1972, Terry Winograd wrote a program, SHRDLU, that could follow commands and answer questions in English regarding a simulated blocks world displayed on a monitor. Its large number of specialist subprograms picked out aspects of the syntax and semantics of a command and combined them with constraints from the current situation to arrive at its response.

Significantly, Winograd did not find Chomsky's generative grammar a suitable tool in writing these subprograms, and even the one he chose (Halliday's functional grammar) required extensive adaptation. A distinction made by Chomsky (1968) – *competence* versus *performance* – suggested one way to think about this. Generative grammar was offered as a formalization of people's tacit knowledge of language – their linguistic competence. Chomsky did not regard it as the linguist's job to study individual acts of comprehending or producing particular sentences in real time – linguistic performance – or to ask how a competence theory might be applied in explaining performance.

It was psycholinguists who faced that question, beginning with Miller's 1962 finding that sentences with more transformations were harder to process and remember. He inferred a close alignment between competence and performance, but later studies yielded mixed results. Moreover, it became clear that the ways of organizing a grammar that worked best for most linguistic purposes were awkward for modeling sentence production or comprehension. By 1970 cognitive scientists favorable to Chomsky had concluded that the relation of competence to performance was more abstract than originally thought, and most others found the notion of competence superfluous. In retrospect, Winograd's SHRDLU was a harbinger of numerous performance-oriented natural language processing systems implemented on computers in the 1970s (e.g., parsers using Aravind Joshi's Tree Adjoining Grammar).

Chomsky found greater uptake among the new *developmental psycholinguists* – those students of child language who signed onto the assault against behaviorism in Chomsky's (1959) review of Skinner's book *Verbal Behavior* (1957). One line of Chomsky's argument emphasized the essential creativity of language, in that there is no bound to the novel but grammatically well-formed sentences of a given language. He also argued for nature (language acquisition constrained by innate knowledge) over nurture (language acquired solely by learning – Skinner's position). David McNeill boldly brought this nativist perspective to bear on toddlers' earliest two-word utterances. More pragmatically, Roger Brown and others adapted such rules as Chomsky had proposed toward writing grammars for individual toddlers as they progressed toward more complex utterances. Views became more diverse as this field grew, and today it is a major nexus of ideas and data in cognitive science. (See Chapter 9 for more on adult and developmental psycholinguistics.)

1.5 Cognitive science gets its name and identity (1975–1980)

Cognitive science flourished for some years before it acquired a name and institutional identity. The term *cognitive science* first appeared in print in two 1975 books. The LNR group's *Explorations in Cognition* ended (p. 409) with the suggestion that the "concerted efforts of a number of people from . . . linguistics, artificial intelligence, and psychology may be creating a new field: *cognitive science*." The same term appeared in the subtitle of a book by computer scientist Daniel Bobrow and cognitive psychologist Allan Collins. The term caught on quickly, and the Alfred P. Sloan Foundation spearheaded interdisciplinary centers at selected universities. One product of its grant to UCSD was the 1979 La Jolla Conference on Cognitive Science, announced as the first annual meeting of the Cognitive Science Society. In 1980 the Society assumed ownership of the journal *Cognitive Science* (launched in 1977).

Roger Schank, who played a central role in the early days of the society and journal, constructed highly original computer simulations that offered an alternative to Chomsky's separation of syntax from meaning. He put his first major computer program through its paces in the early 1970s at Stanford, where he had a joint appointment in linguistics and computer science. MARGIE took in and produced English sentences and made inferences using semantic representations built from eleven primitive predicates and their arguments (e.g., PTRANS linked an actor, an object to be transferred, source, and goal). It worked surprisingly well but tended to license too many plausible inferences. Having already combined AI and linguistics, Schank moved to Yale in 1974 and began collaborating with psychologist Robert Abelson. They developed higher-order knowledge structures, *scripts*, which characterized common experiences. Their well-known restaurant script, for example, specified multiple roles (e.g., diner, server) and scenes (e.g., entering, ordering, eating, and exiting) and the typical sequence of primitive actions for each scene. Schank and Abelson (1977) reported that computer simulations incorporating scripts could read simple stories, infer unmentioned primitive actions to answer questions, and include such inferences in paraphrases.

Symbol-based computational models of mental representations and operations were the high-energy core of the newly named cognitive science. This bridge between psychology and artificial intelligence was constructed not only by computer scientists like Schank, but also by psychologists. We have already noted the wide-ranging models by Norman and Rumelhart (ELINOR) and by Anderson (a colleague of Newell and Simon by the time he created ACT by adding a production system to a HAM-style associative memory). These emerged amid a good deal of interdisciplinary crosstalk and occasional collaboration that was largely limited to the two disciplines. A look at the first volume of *Cognitive Science* (1977) reveals that the affiliations of the authors were either computer science (eight articles), psychology (six articles), or both (one article), and most of the articles concerned computational models.

Nonetheless, cognitive science (narrowly construed) has enjoyed a good deal of interdisciplinary crosstalk with what we might call the cognitive sciences (broadly construed). Consider the active engagement of philosophers of science who have taken cognitive science as an object of analysis. Some have assessed whether the cognitive revolution was a Kuhnian paradigm shift; others (Bechtel 2008; Thagard 2006) have examined the role of mechanistic explanation in cognitive science. A few also have made direct contributions; for example, Paul Thagard collaborated with psychologists Keith Holyoak and Richard Nisbett and computer scientist John Holland on a computer simulation of inductive learning and reasoning. Philosophers of mind have been active as well, forming interdisciplinary collaborations (e.g., Lakoff and Johnson 1980), or initiating debates that engage nonphilosophers (e.g., Fodor's language of thought and Putnam's thought experiments; see Chapter 2). Finally, certain longstanding contributions in philosophy have had an impact within cognitive science. Notably, formats for representing information were adapted from predicate logic not only by early computational modelers but also by cognitive psychologists studying knowledge representation, reasoning, and decision making (e.g., Walter Kintsch; also see Chapter 7). Moreover, philosophical proposals regarding concepts have found uptake in psychology. Eleanor Rosch (1973) propelled "west coast" research on concepts away from necessary and sufficient conditions toward a more Wittgensteinian emphasis on family resemblance and typicality (see Chapter 8). More classic "east coast" paths were pursued by Frank Keil, Elizabeth Spelke, and Susan Carey. For example, Carey (2009) credits children with a substantial core of Kantian a priori concepts, augmented with Quinean bootstrapping as a mechanism for conceptual change.

Together, these examples illustrate the complex relationship between various parts of philosophy and cognitive science from its early years to the present. At the other extreme are disciplines in which just one specialized subfield has had an ongoing participation in cognitive science. Examples include anthropology (beginning with Roy D'Andrade's cognitive treatment of kin terms) and sociology (e.g., Aaron Cicourel's reconstrual of social interaction). Cognitive scientists also monitored developments in neuroscience, but during this period made no major attempt to build bridges.

It gets tricky placing linguistics in this picture. Chomsky's earliest impact was on nonlinguists with an interdisciplinary orientation – nascent cognitive scientists – but he absorbed very little reciprocal impact as he riveted his own attention on generative grammar (and on provocative political essays). By the late 1960s he was succeeding in reshaping theoretical linguistics but also confronting a schism in his own ranks. It was the rebel "west coast" linguists who most directly interacted with and influenced the computational modelers at the core of the newly named cognitive science. Most notable (as discussed in Chapter 9) was the interlacing of semantics and syntax championed by generative semanticists such as George Lakoff and Ronald W. Langacker (giving rise later to cognitive linguistics) and the widely adopted *deep case* categories of Charles J. Fillmore, such as agent, instrument, object, and location. Lakoff and Fillmore had a strong presence in the early years of the Cognitive Science Society and were colleagues at University of California-Berkeley. By the 1980s a broader range of interdisciplinary researchers identified themselves as cognitive scientists, including many influenced by Chomskian linguistics.

Psycholinguists (of both coastal persuasions) were major contributors to cognitive science as it grew and matured, though few found their primary identity there. With the psychological reality of something akin to deep structure already well supported, attention turned to how adults parse, comprehend, and produce sentences (see Chapter 9). For example, Thomas Bever championed strategies such as (1) breaking complex sentences into simple sentoids and (2) conjecturing that a sentoid's N–V–N order corresponds to actor–action–object, which works well for active but not passive sentences. For those focusing on children rather than adults, Chomskian developmental psycholinguists confronted the emergence of semantic, cognitive, and social perspectives. Some debated, for example, whether early sentences like "Mommy eat" were produced from syntactic rules (S $\rightarrow$ NP + VP), semantic rules (actor + action), or narrow word-based formulae (Mommy + X); others sought universals by expanding inquiry beyond English. Most salient to cognitive scientists were increasingly precise data and arguments regarding acquisition processes that emphasized nurture (Catherine Snow), nature (Lila Gleitman), or a dynamic interplay with cognition (Elizabeth Bates).

Cognitive psychologists offered a variety of ingenious strategies for inferring aspects of adults' mental representations or operations. With the transition from behaviorism achieved, their battles now focused on those inferences. There was a long debate, for example, whether all mental representations were composed from discrete symbols or, as held by Alan Paivio and Stephen Kosslyn, some were visual images appropriate for analog operations such as scanning. Another dichotomy productive of research was top-down vs. bottom-up processing (e.g., to what extent is perception driven by expectation?). Also, a trend toward investigating larger units of cognitive activity yielded experimental paradigms based on Schank's scripts or the new *story grammars*.

Finally, artificial intelligence generally proceeded as its name suggests – most researchers directed to computational virtuosity rather than human simulation – but the performance of AI programs was improving only incrementally and it was a particular challenge to scale up from highly constrained domains such as the blocks micro-world. Philosopher Hubert Dreyfus pronounced symbolic AI's core strategy of symbolic rules and representations doomed to fail. Unbeknownst to him and to most cognitive scientists, their friends, and their critics, an alternative was about to shake up the field.

1.6 The connectionist challenge: artificial neural network models (1980 to present)

Information-processing models based on symbolic rules and representations opened mental life to serious inquiry and still are advantageous for many purposes. By the late 1970s, however, there had been little progress in equipping them to learn from experience or in overcoming their brittleness. A few key cognitive scientists took a new look at artificial neural networks and saw in them a promising alternative to stepwise operations on symbols. Such networks had been pioneered in the 1940s by McCulloch and Pitts, as noted above, and were a promising, active research area until the late 1960s. Frank Rosenblatt (1962) developed a training procedure for pattern classification networks in which the key components were McCulloch–Pitts neurons (linear threshold units) that provided one layer of connections with modifiable weights. His *perceptron convergence theorem* proved that if a solution existed, this procedure would find it. However, Minsky and Papert (1969) mathematically dissected important classes of perceptrons, demonstrating no solution (or no tractable solution) for whether or not a geometric figure is connected, parity is odd or even, etc. With this formal justification in place for the emerging dominance of serial, symbolic architectures, only a few dedicated researchers pursued neural network research through the 1970s (most notably Stephen Grossberg).

Within a decade artificial neural networks began their comeback (for more of this story, see Chapters 3 and 12 and Bechtel and Abrahamsen 2002). The turning point was a small, ad hoc conference in June 1979 at UCSD in which neuroscientists, cognitive psychologists, AI researchers, mathematicians, and electrical engineers became aware of common threads in their diverse projects. Visiting scholars Geoffrey Hinton (a computer scientist) and James A. Anderson (a psychologist) served as conference organizers and edited the presentations into a game-changing book (Hinton and Anderson 1981), with an introduction by conference hosts Rumelhart and Norman. UCSD assistant professor James L. McClelland had already developed an influential, transitional *interactive activation* model with Rumelhart, and by January 1982 they had reinvented the LNR research group as the PDP (*parallel distributed processing*) group. Its fluid membership included Hinton, Terrence J. Sejnowski, Paul Smolensky, and Jeffrey L. Elman – recent PhDs whose conceptual and computational virtuosity would soon help shape the new era of network modeling.

The PDP group focused on *distributed networks* in which the task-relevant information is encoded across multiple units, in contrast not only to symbolic architectures but also to the *localist networks* preferred by most other *connectionists* (the name adopted by many using artificial neural networks for cognitive modeling). One key contribution was *backpropagation*, a network learning procedure that finally made it possible to train multiple layers of connections (and hence find solutions where simple perceptrons could not). It was unveiled in a chapter in the first volume of a landmark publication by Rumelhart, McClelland and the PDP research group (1986a), titled *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. The two PDP volumes elicited a barrage of critical responses from symbolic theorists, especially Jerry Fodor and Zenon Pylyshyn. The lengthiest and best-known exchange (see Chapter 9) began with Rumelhart and McClelland's (1986b) single-network model of past tense acquisition in the second PDP volume and Pinker and Prince's (1988) defense of the classic claims that past tense forms are generated by a rule for regular verbs but retrieved from memory for irregular verbs and that two-year-olds' overregularization errors (e.g., *falled* rather than *fell*) signal that they have induced the rule.

Today there is less debate, and cognitive scientists can choose from a variety of neural network and symbolic architectures developed in the 1980s and 1990s. For example, some networks gradually increase the weights between pairs of units that become active together; this is called *Hebbian learning* in recognition of Donald Hebb's proposed synaptic modification mechanism. Others self-organize in other ways (e.g., Kohonen feature maps yield spatially organized two-dimensional sheets of units). Elman's simple recurrent networks retain traces of previous activity, and with Bates and others he developed a nuanced, connectionist perspective on the issue of innateness (Elman *et al.* 1996). For symbolic modeling, production systems continue to play a major role in Newell's SOAR and Anderson's ACT-R (see Chapter 3). Choices in linguistics and psycholinguistics include Chomsky's government binding theory (including a notion of parameter setting frequently applied to acquisition), his more recent minimalism, and alternatives better suited to processing (e.g., functional, cognitive, and construction grammars and head-driven phrase structure grammar). Of special note, *optimality theory* is a constraint-based linguistic theory that interfaces well with PDP networks as an underlying mechanism (see Smolensky and Legendre 2006). Within psychology, statistical approaches have diversified beyond artificial neural networks. Bayesian models offer a competing probabilistic framework for inductive learning and inference (Tenenbaum, Griffiths, and Kemp 2006), while language researchers have grappled with the implications of infants' knack for statistical learning of word boundaries (Saffran, Aslin, and Newport 1996).

1.7 Cognitive science expands downward and outward (1990s to present)

Since 1990 cognitive science has given increased attention to phenomena of emotion (Chapter 10), consciousness (Chapter 11), and animal cognition (Chapter 15) and incorporated new methods and perspectives from cognitive neuroscience (Chapter 12), evolutionary psychology (Chapter 13), embedded and extended cognition (Chapter 14), and dynamical systems theory. This has brought connections with a wider variety of research fields, such as clinical psychology, behavioral biology, human evolution, and artificial life. Regrettably, we cannot discuss all of these developments. Instead we highlight just two trends: the expansion of inquiry down into the brain (cognitive neuroscience) and out into the body and world (embedded and extended cognition).

To begin with the expansion downward, this would seem most naturally to involve artificial neural networks, but that came later; in fact it was a convergence between neuroscience and information processing that ignited cognitive neuroscience in the 1980s. Neuroscientist David Marr played a key transitional role by moving beyond single-cell recording to pursue neurally informed computational models of vision, especially focusing on object representation (see Chapter 4). Marr's life ended prematurely, but his former student Shimon Ullman made his own major contributions. Other neuroscientists redirected the partnership with information processing by bringing in new technologies. In particular, positron emission tomography (PET), and subsequently functional magnetic resonance imaging (fMRI), made blood flow available as a proxy in localizing neural activity while humans performed cognitive tasks. Electrophysiological studies (ERP) offered higher temporal, but lower spatial, resolution. These became the core of the new cognitive neuroscience. Initially regarded as a distraction or competitor by most cognitive scientists, by the twenty-first century these fields increasingly overlapped.

The expansion outward has been more diverse, but the transitional figure clearly is James J. Gibson (see Chapter 4 regarding both Marr and Gibson). He emphasized the rich information in the world outside the perceiver ("in the light") and argued that it was directly *picked up*, not processed step by step in the head. His successors are cognitive scientists who, in varied ways, have focused on cognition as embodied, situated, and extended beyond the individual. Embodied approaches to concepts have ranged from Lawrence Barsalou's perceptual symbol system to the more abstract image-schemas of cognitive linguistics, including Jean Mandler's (2004) nuanced treatment of their onset in infancy. Situatedness is added to embodiment, and physically realized, in robotics. For Rodney Brooks, who designs robots in which a hierarchy of controllers are coupled directly to the sensory-motor apparatus, the seamless, dynamic interaction between agents and the world demonstrates that intervening representations are unnecessary. Anthropologist Edwin Hutchins takes a different tack by adding to situatedness and embodiment the idea that cognitive activities extend beyond a single brain. He examines the coordination of multiple agents and instruments in real-world tasks such as navigating a large ship.

Overall, these avenues of inquiry have made space in cognitive science for a focus on real-time activities of embodied agents, but the more specific claims have been controversial. For example, Andy Clark advocates a philosophy of mind in which mind extends out into the world, but defends representations. In contrast, many advocates of another framework, *dynamical systems theory*, reconceptualize the mind and explicitly deny representations. They contend that coordinated interactions between the world and an agent can best be explained by identifying a small number of critical variables and capturing their evolving relation over time in differential equations. (For discussion, see Chapters 2 and 14.)

1.8 Conclusion

We have followed cognitive science from its historical roots through the cognitive revolution, symbolic rules and representations, subsymbolic artificial neural networks, and its most recent expansions down to the brain and out to the body, world, and other agents. One way of viewing this history is not as a series of polarized proposals and debates, but rather as a dynamic interplay of ideas and approaches. The claim that cognitive science is especially notable for its varied and changing *integrations* of diverse approaches both within and across disciplines is further developed and illustrated by Abrahamsen and Bechtel (2006).

Further reading

Bechtel, W., Abrahamsen, A., and Graham, G. (1998). The life of cognitive science, in W. Bechtel and G. Graham (eds.), *A Companion to Cognitive Science* (pp. 1–104). Oxford: Basil Blackwell. This opening chapter presents a detailed historical overview of cognitive science. The other chapters address a variety of research areas, methods, theoretical stances, controversies, and applications, followed by biographies of 138 early contributors to cognitive science.

Mandler, G. (2007). *A History of Modern Experimental Psychology: From James and Wundt to Cognitive Science*. Cambridge, MA: MIT Press/Bradford. A historical tour of the theoretical and experimental traditions in twentieth-century psychology that emphasizes their social and cultural context, by a leading contributor to the cognitive revolution.

Nadel, L. (ed.) (2003). *Encyclopedia of Cognitive Science*. London: Nature Publishing Group. A four-volume encyclopedia that offers detailed analysis of recent research and theoretical traditions in cognitive science.

Stainton, R. J. (ed.) (2006). *Contemporary Debates in Cognitive Science*. Oxford: Blackwell. Prominent advocates and critics of nativism, modularity, rules and representations, extended cognition, the irreducibility of consciousness and other controversial positions make their arguments accessible to graduate and advanced undergraduate students.

Thagard, P. (2005). *Mind: Introduction to Cognitive Science* (2nd edn.). Cambridge, MA: MIT Press/Bradford. This book provides a highly accessible introduction to representation and computation, including analogical reasoning and reasoning based on images. In the second edition Thagard adds discussion of how the brain and the social and material context of cognitive agents figure in current cognitive science.

Wilson, R. A. and Keil, F. C. (eds.) (1999). *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press/Bradford. This excellent one-volume resource offers approximately 450 one- to three-page articles by experts on such topics as memory, cognitive development, and dynamic approaches to cognition. It also features longer overviews of six major disciplines comprising cognitive science.

MITCogNet. http://cognet.mit.edu. A large online resource including access to many journals, books, reference works, and conference proceedings in cognitive science as well as free courseware for MIT courses in brain and cognitive science.

References

Abrahamsen, A. and Bechtel, W. (2006). Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective, in R. Stainton (ed.), *Contemporary Debates in Cognitive Science* (pp. 159–86). Oxford: Blackwell.

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Bechtel, W. (2008). *Mental Mechanisms*. London: Routledge.

Bechtel, W. and Abrahamsen, A. (2002). *Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks* (2nd edn.). Oxford: Blackwell.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

(1959). Review of *Verbal Behavior*, *Language* 35: 26–58.

(1968). *Language and Mind*. New York: Harcourt, Brace, and World.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Hinton, G. E. and Anderson, J. A. (eds.) (1981). *Parallel Models of Associative Memory*. Hillsdale, NJ: Lawrence Erlbaum.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.

Mandler, J. M. (2004). *The Foundations of Mind*. Oxford University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psychological Review* 63: 81–97.

(1979). *A Very Personal History* (Occasional paper No. 1). Cambridge, MA: Center for Cognitive Science.

Miller, G. A., Galanter, E., and Pribram, K. (1960). *Plans and the Structure of Behavior*. New York: Holt.

Minsky, M. (ed.) (1968). *Semantic Information Processing*. Cambridge, MA: MIT Press.

Minsky, M. and Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A., Rumelhart, D. E., and the LNR Research Group (1975). *Explorations in Cognition*. San Francisco: Freeman.

Pinker, S. and Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition, *Cognition* 28: 73–193.

Rosch, E. (1973). Natural categories, *Cognitive Psychology* 4: 328–50.

Rosenblatt, F. (1962). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Washington: Spartan Books.