Chapter 9

The Computational Theory of Mind

9.1 Historical Background

9.1.1 Gestation

We’ve seen that for millennia the idea has persisted that comprehensibility ultimately requires reduction to finite arrangements of discrete tokens. In mathematics this notion manifested itself first in the attempt to arithmetize geometry, that is, to reduce the continuum to discrete points. Secondly the preference for the discrete led to the attempt to formalize knowledge, that is, to reduce knowledge to formulas, finite arrangements of tokens, and inference to the mechanical rearrangement of formulas.

Traditionally there were taken to be two kinds of knowledge, analytic and synthetic. As defined by the logical positivists, analytic knowledge was exemplified by the arbitrary, factually empty, but useful truths of formal mathematics; synthetic knowledge comprised the empirical truths of “real science.” The former took the ultimate source of truth to be the (freely chosen) axioms, the latter took it to be atomic facts (whatever they might be). Both used formal logic to construct more complex ideas from more basic ones.

The idea that knowledge can be expressed as a formal system naturally suggests that thought is the mechanical rearrangement of finite structures of discrete tokens, that is, that thought is computation. We’ve seen different versions of this idea in the work of Llull, Hobbes, Leibniz, and Jevons.
We have followed the attempt to formalize knowledge in preceding chapters; in this chapter we pick up the trail of the formalization of thought, which eventuated in a new discipline, cognitive science.

Although, as we have seen, there was no lack of research into the formalization of thought up through the 1930s, the pace was accelerated by the needs of the Second World War. Increasingly sophisticated weapons, including the atomic bomb, required complex computations and “intelligent” control systems. The result was an increase both in resources – human and financial – and in emotional commitment.¹

The most tangible result was the development of the first practical electronic computers (e.g. ENIAC: 1946, EDSAC: 1949, EDVAC: 1950), which were needed for calculating numerical tables. The possibility of manipulating nonnumerical information was demonstrated by Turing’s cryptanalysis work, for which he used computers not unlike Turing machines (Sect. 6.4.2). The war effort also led to important theoretical advances, such as the development of cybernetics by Norbert Wiener (1894–1964). Cybernetics can be defined as the study of control systems in animals and machines; in Europe the term cybernetics includes artificial intelligence, though in the United States it’s used more narrowly. Further, in an influential book, Design for a Brain (1952), W. Ross Ashby () used formal methods to show how a computational system could learn.

Throughout the ‘40s and ‘50s the influence of von Neumann was pervasive; he was active in computer design, automata theory, formal mathematics, neuroscience and cybernetics. His broad viewpoint helped to keep together the many strands — practical, theoretical, technological, biological, mathematical and philosophical. As much as any individual he represents the coalition of interests that was to constitute cognitive science.

Claude Shannon, who in 1938 had shown the relation between digital circuits and Boolean logic, went on to develop information theory with Warren Weaver (1894–1978) in 1949. This theory, which perhaps stemmed from insights gained in wartime cryptographic work, is positivist in method, for it reduces information to the probability of identifying symbols, that is, to a behavioral criterion. It has been very influential in many sciences and even in the humanities, though this may be more an consequence of its suggestive

¹Much of the historical information in this section comes from Gardner (MNS), which is a comprehensive presentation of the history of Cognitive Science. Chapters 2 and 3 are especially relevant to this section.
name rather than a close connection to our everyday idea of information. For example, it has been pointed out that the book (of a given size) containing the most information (in Shannon’s sense) is a book of random numbers (Hamming, C&IT, p. 103).\(^2\) One again we see the positivist method — take a concept, such as information, which no one knows how to quantify — and replace it by a much simpler quantifiable concept, such as prediction probability; and we see the typical consequence of the method, for many investigators, even in disciplines where they should know better, no longer recognize a difference between the ordinary and Shannon notions of information. In effect, information has contracted to Shannon’s definition.

An important event was a 1948 symposium on “Cerebral Mechanisms in Behavior,” which brought together many of the new ideas. For example, von Neumann discussed the similarities between the computer and the brain. Warren McCulloch (1898–) developed a similar theme, based on prior work with Walter Pitts (), and showed how simple, neuron-like computing elements could function like the logic circuits of computers.

Most importantly, however, the neuropsychologist Karl Lashley gave convincing arguments why behaviorism was inadequate. Behaviorist theory was organized around the reflex arc, but Lashley argued that this could not account for complex, organized behavior, such as planning, reasoning, problem solving and especially language use. Thus, Lashley’s paper signalled the beginning of a shift of emphasis from “lower” cognitive functions, such as reinforcement of stimulus-response pairs, to “higher” cognitive functions, such as game playing, problem solving, abstract reason and the use of language.\(^3\) Structured behavior of this kind is, seemingly, the manipulation of discrete symbols, so the new research agenda meshed well with the trend in the philosophy of science and mathematics. These are also the tasks first attacked by artificial intelligence, which will be considered shortly.

In Chapter 8 we saw how logical positivism, quantum mechanics and relativity all gave a central theoretical role to observation. Investigations in all three areas showed that observation was not the simple thing it had been imagined, and it became apparent that the psychology of the observer was not irrelevant to physics. Therefore, J. Robert Oppenheimer (1904–1967), as

\(^2\)This is because Shannon takes a message to be less informative to the extent that it is predictable. In a book of random numbers there is no redundancy; all the digits are equally likely. Hence the digits are as unpredictable as they can be.

\(^3\)These functions are “higher” or “lower” in the sense that only “higher” animals, especially humans, possess the higher faculties, whereas all animals possess the lower.
director of the Institute for Advanced Studies, which included Einstein and von Neumann among its members, invited psychologists to visit the institute. For example, both George Miller (1920–) and Jerome Bruner who became major contributors to cognitive science, spent a year at the IAS. Bruner had investigated the effect of the attitude and expectations of observers on their observations, work with obvious relevance to the new physics. The result was a further breakdown of disciplinary barriers, which helped the new paradigm to develop. (Miller’s work will be discussed later.)

**9.1.2 Birth**

Sometimes, when pent-up frustrations with an old way of doing things have accumulated, and a number of advantages of a new way have become apparent, there will be precipitous change from the old to the new, a *phase transition* like the sudden crystallization of a supersaturated solution. This is what seems to have happened in psychology in 1956, for in that year cognitive science was born, with its distinctively formal approach to thinking in both minds and machines. Although behavioristic methods were not abandoned, cognitive science restored mental phenomena to respectibility by means of the computer model. Such a radical change of scientific methodology would later be termed a *paradigm shift* (see Sect. ??).

A second harbinger was the “Symposium on Information Theory” held at MIT in 1956. Newell and Simon were also at this conference, at which they described the “Logic Theorist,” a program that was able to prove theorems in symbolic logic from Russell and Whitehead’s *Principia Mathematica*. Here was a computer accomplishing a task that would otherwise require an intelligent person, so it suggested that artificial intelligence was a real possibility, and illustrated how very precise and specific models of higher cognitive abilities could be constructed. (Newell and Simon’s views are discussed in more detail in Sect. 9.3.1.)

Another speaker at the symposium was Noam Chomsky (1928–), a linguist who showed how grammar could be described with complete formality, by using *term rewriting rules* similar to the Markov productions we discussed
9.1. HISTORICAL BACKGROUND

in Section 6.4.2. Chomsky’s mathematical approach to linguistics superseded all others, and retains a nearly universal hegemony over the field; it’s discussed in Sect. 9.2.

George Miller, one of the major promoters of cognitive science, also presented a paper, his famous “Magical Number Seven, Plus or Minus Two.” He showed that over a wide variety of cognitive tasks the capacity of our “short-term memory” is limited to between five and nine “chunks” of information. The analogy with the “high-speed registers” of a computer (often 8 or 16 in number) was apparent and increased the attractiveness of computer models in psychology.4

Aside from these two conferences, 1956 saw the publication of a number of other key papers signaling the birth of cognitive science. For example, anthropologist Gregory Bateson () applied cybernetic principles in an innovative theory of schizophrenia. The Fundamentals of Language, which explained language in terms of atomistic features, was published by linguists Roman Jakobson (1896–) and Morris Halle () in 1956. This year also produced A Study of Thinking by Bruner and his colleagues Jacqueline Goodnow (1924–) and George Austin (1916–).

Following hard on the heels of these events came further significant publications. Von Neumann wrote a set of lectures — still worth reading — that explored the similarities and differences between brains and computers, both analog and digital. Though he was too sick to deliver the lectures, they were published posthumously (von Neumann, C&B).

In 1960 philosopher Hilary Putnam argued that traditional mind/body problems could be solved by the computer analogy: the body was like the physical computing machine — the hardware — and the mind was like the program controlling the operation of the computer — the software. Key to this solution was the observation that software is independent of its material instantiation (as the mind is supposed to be), a property we’ve named abstract formality (Sect. 6.5.1). The resulting view, that “the mind is a

4The most familiar example of this phenomenon is the way we hold an unfamiliar telephone number in our mind between when we look it up and when we dial it. Phone numbers also illustrate the effect of “chunking,” for if part of the number is familiar, such as the area code, then it counts as only one chunk, rather than the three for an unfamiliar code. We often rehearse the contents of our short-term memory to keep from losing it, a fact which must have reminded some of the conference participants of contemporary computer memories, such as mercury delay lines and cathode-ray tube storage, both of which also required regular refreshing to retain their contents.
program,” is still hotly debated (see Sect. ??)

The year 1960 also saw the publication of *Plans and the Structure of Behavior* by George Miller and Eugene Galanter (), both psychologists, and Karl Pribram (1919–), a neuroscientist. They argued that complex structured behavior, such as Lashley had emphasized, was best explained in terms of a hierarchy of “subroutines,” such as used in computer programing. They also showed that many behavioral routines could be explained in terms of a common programming technique (called a *leading-decision loop* or *while-loop*). One can see how artificial intelligence and cognitive psychology went hand-in-hand, for if human intelligence could be explained by a program, then (thanks to abstract formality) that same program could be executed on a computer, which as a result would behave intelligently. Conversely, if a computer program exhibited human-like intelligence, then that would be evidence that the human mind contains a similar program.

### 9.1.3 Growth

Work in all aspects of cognitive science accelerated in the 1960s and ‘70s. Neuroscientists David Hubel () and Thorsten Wiesel () had been studying vision since the 1950s, and in 1962 began publishing the work that would win them the Nobel Prize in 1981. They interpreted their results as showing “geometrically” organized *feature detectors* in the brain. That is, neurons in the first stages of visual processing identified points (spots of light) in the visual field. In the second stage, points were combined to form edges and lines. The subsequent stages combined edges and lines into planar shapes, and so forth. The approach was atomistic in the logical positivist tradition, since higher-order percepts (e.g., a dog percept) were constructions of invisible sense data (spots of light or color). Computer vision programs were organized along similar bottom-up lines: individual spots (pixels) of light were recognized, configurations of spots were assembled into edges or lines, configurations of edges and lines assembled into planar shapes, and so forth.

Books reporting the first successes in artificial intelligence also appeared in the 1960s — Edward Feigenbaum () and Julian Feldman () edited the

---

5Indeed, Pribram was a student of Lashley; the contribution of Lashley and Pribram to connectionism is discussed later (Sect. ??).

6So called because it repeats some action *while* some condition holds. Miller et al., however, called it a *TOTE unit* (Test-Operate-Test-Exit).
9.1. HISTORICAL BACKGROUND

collection Computers and Thought in 1963 and Minsky edited Semantic Information Processing in 1968. Not long after, the first major critique of artificial intelligence, Hubert Dreyfus’s What Computers Can’t Do, appeared; we’ll consider its arguments in Section ??.

Chomsky and his students vigorously defended their formal approach to linguistics, which was presented both in Chomsky’s books (Syntactic Structures, 1957; Aspects of the Theory of Syntax, 1965; Cartesian Linguistics, 1966; Language and Mind, 1972) and in anthologies, such as The Structure of Language (1964), edited by Jerry Fodor () and Jerrold Katz ()

The late ‘60s and early ‘70 produced several influential cognitive science books. One sign that the field was becoming recognized was the appearance of the first textbook, Cognitive Science (1967) by Ulric Neisser (). Further, in 1969 Simon published his lectures, Sciences of the Artificial, which argued that it is possible to have empirical sciences of artificial structures, such as computer programs and human organizations, such as businesses. With Newell he published Human Problem Solving in 1972; it states explicitly the physical symbol system hypothesis, which is, in essence, the hypothesis that cognition is best understood as a calculus.

Finally, we note evidence for the recognition of cognitive science that came in institutional form. For example, the Sloan Foundation began supporting cognitive science research in 1975. Also, in 1977 the journal Cognitive Science began publication, and in 1979 the Cognitive Science Society was founded. By this time the cognitive science approach was well regarded in most of the disciplines concerned with behavior: psychology, artificial intelligence, philosophy, anthropology and linguistics.

9.1.4 Characteristics of Cognitive Science

Before turning to the theory of knowledge and language deriving from cognitive science, it will be worthwhile to consider the characteristics of the cognitive science approach, as explained by Gardner (MNS, pp. 6–7, 38–45).

First, in contrast to behaviorism, which rejected any reference to the mental, cognitive science took it for granted that explanations of higher cognitive functions would have to refer to plans, goals, intentions and other mental representations of actual or possible states of the world. Further, it stressed the importance of structure at all levels of organization: neural, individual and social.
Second, cognitive science took digital computation as the best model for the thought. In part this allowed cognitive scientists to answer the objections of behaviorists: since mental representations would be modeled as specific computer data structures, and cognitive processes would be modeled as computer programs, cognitive science could promise precise testable hypotheses.

Third, cognitive science concentrated on cognitive phenomena, such as rational thought, planning, problem solving and logic. Thus, the approach neglected, on the one hand, affective phenomena, such as mood and emotion, and on the other, contextual effects, such as historical and cultural factors in cognition. This betrays an “intellectualist bias” in cognitive science, which says, in effect, that cold reason is what is most characteristic of the human mind, and that nonrational emotional and social factors are simply complications of no great significance. Just as physics postulates idealized conditions, such as frictionless surfaces and perfect vacuums, to simplify analysis, so cognitive science postulated the ideal “rational agent.” Divergence between the competence of idealized agents and the actual behavior of real people was relegated to a theoretical no-man’s land called performance.

Fourth, cognitive science took the view that it was of necessity an interdisciplinary field. This was represented in the “cognitive hexagram” (Fig. 9.1), which appeared in an unpublished 1978 report (Gardner, MNS, p. 37). The diagram showed that cognitive science comprises six disciplines; in some cases, indicated by solid lines, interdisciplinary work was already underway; in other areas, indicated by dotted lines, it needed to be encouraged. Cognitive science was seen to be very broad in scope, encompassing structures and processes from the neural level to the social level.

The fifth and last characteristic of cognitive science identified by Gardner is that its agenda is essentially that of the Western philosophical tradition. Though he thinks this claim may be controversial, I hope that the preceding chapters of Word and Flux have demonstrated that cognitive science was the natural, if not inevitable, culmination of a tradition extending back at least as far as Pythagoras.

---

7 More specifically, as already explained, the mind was viewed as a program running on the brain as a computer.

8 Since the hypotheses refer to idealized competence, but only performance can be observed, theories of this kind are hard to refute, since differences between prediction and observation can always be attributed to unanalyzed performance limitations. See Sect. ?? for more on the competence/performance distinction.
Figure 9.1: The Cognitive Hexagram. The diagram shows that cognitive science comprises six disciplines. Solid lines indicate established interdisciplinary links (as of 1978); dotted lines indicate links that needed encouragement.

Now that we’ve considered the historical background of cognitive science, we can look more closely at its approach to language and knowledge, as exemplified by two of its best-known theories.
CHAPTER 9. THE COMPUTATIONAL THEORY OF MIND

9.2 Chomsky: Formal Linguistics

9.2.1 Phrase-Structure Grammars
9.2.2 Transformational Grammars
9.2.3 Government-Binding Theory
9.2.4 The Problem of Language Acquisition

9.3 Symbolic Knowledge Representation

9.3.1 Newell & Simon: The Physical Symbol System Hypothesis
9.3.2 Schank: Conceptual Dependency Theory
9.3.3 Abelson: Scripts
9.3.4 Expert Systems
9.3.5 The Frame Problem