7. Limits of Conceptualization

No theory is fully understood until its limitations are recognized. To avoid the presumption that conceptual mechanisms completely define the human mind, this chapter surveys aspects of the mind that lie beyond (or perhaps beneath) conceptual graphs. These are the continuous aspects of the world that cannot be adequately expressed in discrete concepts and conceptual relations.

> This is the epilogue, Chapter 7 of the book *Conceptual Structures: Information Processing in Mind and Machine* by John F. Sowa. It was published by Addison-Wesley in 1984. If I were rewriting it today, I would make major additions. But it is still a good overview, since it raises many more questions than answers. For updates with some answers and even more questions, see Signs, processes, and language games (2002), The challenge of knowledge soup (2006), Worlds, models, and descriptions (2006), What is the source of fuzziness? (2013), Signs and reality (2015), and The virtual reality of the mind (2018).

7.1 Cybernetics

Norbert Wiener (1948) coined the term *cybernetics* for "the entire field of control and communication theory, whether in the machine or in the animal." Wiener's is broad enough to include artificial intelligence. But as the fields developed, cybernetics and AI proceeded in different directions. The main difference lies in the problems that they address:

- From its very beginning, cybernetics studied low-level mechanisms such as feedback loops, neural networks, and servomechanisms. It sought a bottom-up view of how such structures could interact to control animal and machine systems.
- Artificial intelligence, however, started with a top-down view of symbols, language, and knowledge and asked what kinds of mechanisms and representations are necessary to simulate human performance.
- Although AI and cybernetics both use networks, the elements in the nets belong to different levels of mental processing. The nets in cybernetics represent individual neurons, but the nets in AI represent high-level concepts and relations, each of which may depend on millions of interacting neurons.

The difference between cybernetics and AI is exactly parallel to the difference between neuroscience and cognitive psychology. In interdisciplinary studies, AI and cognitive psychology have had a great deal of fruitful interchange. Likewise, neuroscience and cybernetics have had a strong influence on each other. But despite some attempts to bridge the gap, the two top-down fields have little contact with the two bottom-up fields.

Besides the differences in subject matter, AI and cybernetics inherited outlooks from their parent disciplines. Cybernetics is descended from analog systems, but AI is based on digital computers. Because of their heritage, they use different formalisms: cybernetics uses continuous mathematics, especially calculus and differential equations; but AI uses discrete mathematics, especially symbolic logic and formal grammars. AI has strong ties to programming language design and natural language processing, but cybernetics has no contact with those fields. Although Wiener's definition claimed the entire subject matter of AI, his work has had little influence on AI.

Among the pioneers of AI, Marvin Minsky is one of the few who has continued to search for low-level neural mechanisms that can support high-level symbolic operations. His widely used textbook on automata theory (1967) contains a major section on neural nets. Yet many teachers who use the book skip that section in their reading assignments. Minsky and Papert (1969) wrote a definitive treatment of *perceptrons*, a class of network machines that can learn to recognize simple patterns. That book was, perhaps, too definitive — after Minsky and Papert showed exactly what the machines could do, people lost interest in them as a field for further research. Minsky's recent work (1981) has been on *K-lines* or *knowledge lines* for passing messages among a society of interacting agents or actors.

Research on K-lines and other kinds of neural networks may someday support symbolic processes as firmly as modern chemistry is supported by physics. But the human brain is vastly more complex than the simple atoms and molecules studied in physics, and many decades of work remain before psychology can be reduced to neuroscience. In fact, the analogy with chemistry suggests that cognitive science is more likely to guide neuroscience than to be derived from it: the atomic hypothesis, the theory of valences, and the periodic table were all accepted in chemistry for many years before atoms, electrons, and electronic orbits were discovered in physics. When physics finally caught up with chemistry, the foundations of chemical theory were reinterpreted, but experimental practice remained unchanged.

Attempts to derive psychology from physiology have often been misleading. One source of inspiration for the arc was the view of the brain as a telephone switchboard: each arc would correspond to a nerve fiber linking a point in a sensory region to a point in a motor region. Yet Lashley (1950) showed that even a simple conditioned reflex must involve the concerted action of millions of neurons. Wolfgang Köhler, who had a background in physics, tried to explain Gestalt principles in terms of electrical fields in the brain. Yet the field theory never proved fruitful, although the Gestalt phenomena remain significant. For a long time to come, psychology and neurophysiology will remain independent, but related disciplines. Advances in one field will often clarify problems in the other field, but neither will be wholly derived from the other.

7.2 Vagueness and Expressive Power

Anything that can be stated precisely in any artificial language can be stated equally precisely in English or any other natural language. In fact, most artificial languages may be viewed as extensions or abbreviated forms of natural languages: the equation 2+2=4 is an abbreviation for the sentence *Two and two is four.* More esoteric notations are defined in terms of simpler notations, which in turn are defined in terms of other notations, until finally one reaches a definition in a natural language. If English did not have the ability to be precise, languages defined in terms of it could never be precise.

Artificial languages eliminate vagueness by reducing their range of expression. They are defined for a special purpose, have a constrained syntax, and limit the vocabulary to a few hundred symbols. Even more important than the constraints on syntax are the constraints on concepts. With only a small number of concepts, a restricted language can express each one in a single character instead of a full word; if each symbol is limited to a single meaning, ambiguities and metaphors are impossible. Natural languages, by contrast, have the vocabulary and means of expression for talking about anything within the scope of human experience. As the range of experience grows, they can quickly accommodate new definitions and metaphorical extensions. Whatever can only be stated vaguely in English cannot be stated at all in a formal language.

Varying shades of gray instead of black and white dichotomies, family resemblances instead of necessary and sufficient conditions, and issues with more sides to them than there are people to think about them — all these are characteristic of human life. Because natural languages can express anything within human experience, they must accommodate all the variability and vagueness of human emotions and the external world. Odell (1981) summarized the characteristics of natural language in the following ten principles:

- *Context.* "Communication through a natural language is, in large part, a function of context. *Where* and *when* something is said largely determines *what* is said."
- *Emphasis.* "What we mean is also a function of *how* we say it. Where or upon what word or words we place an emphasis (*intonation contour*) as well as how we move various parts of our bodies (*body language*) will frequently affect what we mean."
- *Multiple speech acts.* "The range of things (speech acts) a given sentence can be used to accomplish is limitless."
- *Intentionality.* "What a sentence *means* (a proposition) is often quite different from what we *mean by* it, which is sometimes a statement, sometimes a warning, sometimes a request, and sometimes something else."
- *Nonfunctionality.* "What a given string of words means is not a function of the formal characteristics those strings possess. "Why not?" can be used to make a request, even though its *form* is that of a question."
- *Family resemblance.* "What most, if not all, general empirical terms *mean* in a natural language, as opposed to what we might *mean by* them on some specific occasion, cannot be specified *formally*, that is, in terms of necessary and sufficient conditions. They are family resemblant in nature."
- *Overlapping and criss-crossing definitions.* "Since most of the general empirical terms of a natural language are family resemblant in nature, it follows that in order to get at their meanings, i.e. the concepts they express, one must specify the set of overlapping and criss-crossing characteristics which determine the similarities and differences relevant to the question of whether or not some imagined or existing case falls under the concept in question."
- *Open texture.* "Even if we legislate sets of necessary and sufficient conditions to govern what they mean, we can't be sure that our legislations will preclude the existence of contexts where we will be uncertain what our words mean, that is, we can still imagine cases where we wouldn't know whether or not a given word applied."
- *Continuity.* "The concept expressed by any given word in a natural language is inextricably tied to the concepts expressed by nearly every other word in the language. While the words themselves are no doubt *discrete*, the concepts they involve, or are tied to, are *continuous* with other concepts."
- *Sincerity.* "A very large number of speech acts which can be implemented in a natural language involve expressing one's emotions. A natural language incorporates the distinction between a genuine and a nongenuine expression of an emotion. Expressing concern and expressing genuine concern are recognizably quite different."

Odell presented these ten characteristics of natural language in arguing against the possibility of processing unrestricted natural language by computer. They also represent ten limitations on the expressive power of artificial languages. No programming language ever designed has been able to deal with any of them. Natural language processors available today cannot satisfy them with the full expressive power that people use, and perhaps they never will. Yet some language processors can approximate some of them. What they do today does not qualify as unrestricted natural language, but it can provide a more natural and flexible environment than the formal programming and query languages.

Critics of natural language emphasize its vagueness and ambiguity. Hill (1972) quoted traffic regulations, cooking recipes, instructions on a shampoo bottle, and the *Book of Common Prayer*. His examples show ambiguities, misleading phrases, and incomplete or inaccurate statements. The examples, however, all show problems with communication in English. None of them show that human-to-computer communication would be different. Hill's final point is that programming languages are more precise and that people ought to learn ALGOL to communicate better with each other. Yet people do not communicate well in ALGOL: the overwhelming majority of programs written in any programming language are wrong when first submitted to a computer. Mathematics is just as error prone: textbooks by first-rate mathematicians often go through several editions without having all their errors detected and corrected. Artificial languages are designed to be unambiguous, but what they so unambiguously say may bear no resemblance to what the author intended.

Some criticisms arise from the misconception that special symbols and abbreviations are not a part of natural language. One study (Sheppard et al. 1981) concluded that English was harder to understand than a special program design language (PDL). As an example, they compared the PDL form,

SET TOTAL = DELIV * PRICE

to a supposedly natural form,

CALCULATE THE TOTAL PRICE FOR THE ITEM BY MULTIPLYING
THE QUANTITY DELIVERED BY THE PRICE PER ITEM.

This sentence is highly unnatural. In an accounting textbook, natural language includes symbols like "=" and "×":

Total price = Quantity delivered × Price per item.

What is natural depends on the topic. In chemistry, natural language includes such sentences as *Add 125 ml of conc.* H_2SO_4 *and 125 ml of 85%* H_3PO_4 *and dilute to 1 liter*. In prescribing drugs, a physician might write *penicillamine 250 mg PO QD*. For any subject, natural language is the form of expression that two experts in the field commonly use in speaking or writing to each other. If both persons are mathematicians, their natural language normally includes symbols and formulas. If they are programmers, their natural language may include computer terms and even fragments of programming languages.

Correctness in any language is not possible without a dialog: debugging sessions on a computer, conversations between people, or peer review of scientific literature. The ability to be vague or incomplete lets a person talk about ideas while they are still in a half-formed state. One person can begin a problem-solving session with a tentative, partial statement. Through a dialog, others can analyze, refine, and complete the problem statement. One of the most challenging uses for natural language processing — and potentially the most fruitful — is in working with the user to formulate and complete a newly emerging idea. Several AI prototypes are exploring the use of English as an aid to the analysis and formulation of a problem statement: Heidorn's NLPQ for writing programs (Section 1.1), Codd's RENDEZVOUS for formulating database queries (Section 6.2), and Haas and Hendrix's NANOKLAUS for knowledge acquisition (Section 6.6). With such systems, people do not write programs in English. Instead, they carry on analysis and discussion in English, and the computer itself writes the programs. Even Hill admitted, "The main difficulty in programming lies in deciding *exactly* what is the right thing to do. To put it into a programming language is relatively trivial." Better than any programming language, natural languages are suited to analysis, discussion, design, and planning. Once the analysis is done, the "relatively trivial" task of writing code can be left to the machine.

7.3 Conceptual Relativity

Concepts are inventions of the human mind used to construct a model of the world. They package reality into discrete units for further processing, they support powerful mechanisms for doing logic, and they are indispensable for precise, extended chains of reasoning. But concepts and percepts cannot form a perfect model of the world — they are abstractions that select features that are important for one purpose, but they ignore details and complexities that may be just as important for some other purpose. Leech (1974) noted that “bony structured” concepts form an imperfect match to a fuzzy world. People make black and white distinctions when the world consists of a continuum of shadings.

For many aspects of the world, a discrete set of concepts is adequate: plants and animals are grouped into species that usually do not interbreed; most substances can quickly be classified as solid, liquid, or gas; the dividing line between a person’s body and the rest of the world is fairly sharp. Yet such distinctions break down when pushed to extremes. Many species do interbreed, and the distinctions between variety, subspecies, and species are often arbitrary. Tar, glass, quicksand, and substances under high heat or pressure violate common distinctions between the states of matter. Even the border between the body and the rest of the world is not clear: Are non-living appendages such as hair and fingernails part of the body? If so, then what is the status of fingernail polish, hair dye, and make-up? What about fillings in the teeth or metal reinforcements embedded in a bone? Are tattoos, contact lenses, braces on the teeth, or clothes part of the body? Even the borderline between life and death is vague, to the embarrassment of doctors, lawyers, politicians, and clergymen.

These examples show that concepts are *ad hoc*: they are defined for specific purposes; they may be generalized beyond their original purposes, but they soon come into conflict with other concepts defined for other purposes. This point is not merely a philosophical puzzle; it is a major problem in designing databases and natural language processors. Section 6.3, for example, cited the case of an oil company that could not merge its geological database with its accounting database because the two systems used different definitions of *oil well*. A database system for keeping track of computer production would have a similar problem: the distinctions between minicomputer and mainframe, between microcomputer and minicomputer, between computer and pocket calculator, are all vague. Attempts to draw a firm boundary have become obsolete as big machines become more compact and small machines adopt features from the big ones.

If an oil company can’t give a precise definition of oil well, a computer firm can’t define computer, and doctors can’t define death, can anything be defined precisely? The answer is that the only things that can be represented accurately in concepts are man-made structures that once originated as concepts in some person’s mind. The rules of chess, for example, are unambiguous and can be programmed on a digital computer. But a chess piece carved out of wood cannot be described completely because it is partly the product of discrete concepts in the mind of the carver and partly the result of continuous processes in growing the wood and applying the chisel to it. The crucial problem is that the world is a continuum and concepts are discrete. For any specific purpose, a discrete model can form a workable approximation to a continuum, but it is always an approximation that must leave out features that may be essential for other purposes.

Since the world is a continuum and concepts are discrete, a network of concepts can never be a perfect model of the world. At best, it can only be a workable approximation. The psychologist Jaensch (1930) stressed the need for a *principle of tolerance* to accommodate different systems of percepts, concepts, and relations:

> Our investigations show that, like the perceptual world, our world of thought and knowledge is decisively determined by the structure of our consciousness. The kind of structure differs in the various fundamental types. The systems of knowledge of the

> different sciences are also based to a large extent on the different type of mind-structure operating in them. Different categories correspond to each. Each structure of consciousness separates out different aspects of reality, by reproducing certain categories of reality through the medium of categories of consciousness that are related to them. Those categories of reality, in which corresponding categories of consciousness are not present, remain unapproachable and are apprehended through different structures. The danger of one-sidedness, subjectivity and error in the fundamental questions of knowledge, is chiefly due to the fact that every structure of consciousness claims unlimited validity; but in truth each makes very wide negative abstractions of reality. We can, therefore, only penetrate reality and approach the ideal of "pure experience" by successively taking up the standpoints of different mental structures. (p. 118)

Jaensch emphasized the need for tolerance. A closed, rigid system maintains a sense of security by giving instant answers to all perplexities. But it is a false security that is threatened by any incompatible viewpoint. When the good is defined in absolute terms, anything that differs from the definition is automatically evil.

By drawing distinctions and giving names to the things distinguished, language separates figure from ground. Consider a tree. It has no sharp boundaries between parts; yet words divide the tree into trunk, roots, branches, bark, twigs, leaves, buds, knots, flowers, seeds, fruit, and even finer subparts such as veins in the leaves and pistils in the flowers. Even the boundary between the tree and the environment may be indistinct: the tree may have started as a sprout from the root of another tree and may still share a root system with its parent and siblings; insects and animals may be living in and on the tree; a vine may be climbing up the trunk, moss may be on the bark, fungus may be growing on a dead branch, and bacteria in root nodules may be supplying nutrients. The arbitrary way that words cut up the world was emphasized by the linguist Benjamin Lee Whorf (1956):

> We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds — and this means largely by the linguistic systems in our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way — an agreement that holds throughout our speech community and is codified in the patterns of our language. (p. 213)

The division of the world into distinct things is a result of language. The philosopher Searle (1978) elaborated that point:

> I am not saying that language creates reality. Far from it. Rather, I am saying that *what counts* as reality — what counts as a glass of water or a book or a table, what counts as the same glass or a different book or two tables — is a matter of the categories that we impose on the world; and those categories are for the most part linguistic. And furthermore, when we experience the world, we experience it *through* linguistic categories that help to shape the experiences themselves. The world doesn't come to us already sliced up into objects and experiences: what counts as an object is already a function of our system of representation, and how we perceive the world in our experiences is influenced by that system of representation. The mistake is to suppose that the application of language to the world consists of attaching labels to objects that are, so to speak, self identifying. On my view, the world divides the way we divide it, and our main way of dividing things up is in language. Our concept of reality is a matter of our linguistic categories.

The biologist Maturana (1978) summarized the issue in a pithy slogan, "Human beings can talk about things because they generate things by talking about them."

Whorf emphasized the influence of language on thought, but much if not most of the influence results from cultural distinctions that have been codified in language. The concepts of earlier generations, fossilized in language, then shape the perceptions of later generations who learn that language. Polish and Russian illustrate the relative effects of language and culture. Linguistically, the two languages have similar grammars and a large body of cognate words. Culturally, however, Polish has had closer economic, literary, and religious ties to western Europe. In doing conceptual analysis for machine translation, von Glasersfeld et al. (1961) found that the concept types expressed in Polish are more nearly commensurate with the concepts of English or German than with Russian. The Polish word for *write* is *pisać*, which is cognate with the Russian *pisat'*. Yet Russian uses the same verb *pisat'* for writing a letter and painting a picture. Polish, however, borrowed the German word *malen* for *paint* and added a Polish ending to form *malować*. As a result, English, German, and Polish make the same distinction: *write/paint*, *schreiben/malen*, and *pisacć/malować*. For many other verbs as well, English, German, and Polish are semantically closer to each other than any of them is to Russian. As a result of cultural contact, the word forms do not change, but the underlying concepts are calibrated so that people (or machines) can translate them more directly.

Although Russian was more isolated from western Europe than Polish, it is still a European language that embodies a European world view. Hungarian and Finnish are not structurally Indo-European, but they are so firmly bound to European culture that their concepts are semantically commensurate with other European languages. Chinese, however, is structurally unlike the European languages, and it was culturally isolated from them for many millennia. Bloom (1981) noted differences of structure in Chinese and English that have a major influence on the modes of thinking in those languages:

- *Morphology.* English and other Indo-European languages can add an ending to a word that converts it into another part of speech: *approve, approval; discuss, discussion; white, whiteness; soft, soften, softness; general, generality, generalize, generalization*. Since Chinese words lack inflections, they behave more like the English words *talk* or *hand*, which can be used as a noun, verb, or adjective without a change of form.

- *Syntax.* Along with the difference in morphology, Chinese lacks the syntactic means for transforming the sentence *Congress approved the measure* into the noun phrase *the approval of the measure by Congress*. As a result, a sentence containing such constructions is difficult to translate into Chinese: *We will put off to next week discussion of the further implications of the new method for calculating the relationship between the rate of economic development and the individual standard of living*. Bloom found that a direct translation of that sentence strikes native Chinese speakers as "Westernized Chinese speech." Furthermore, they found it difficult to understand: 58% of them were unsure about what was to be discussed. A more natural Chinese form would break up the English sentence into three separate sentences with simple, active verbs instead of the more abstract noun phrases.

- *Counterfactual conditionals.* English speakers can easily talk about hypothetical situations that have not happened, but Chinese has no syntactic form for expressing them. Bilingual speakers report that in English they feel comfortable in saying *If the lecture had ended earlier, Bill would have had a chance to prepare for the exam*; but in Chinese they would say *The lecture ended too late, so Bill did not have a chance to prepare for the exam*. Chinese speakers learning English find the most difficulty with sentences containing *would have*. Speakers of other European languages learn the English form without difficulty.

Bloom showed the effects of these differences with both informal anecdotes and controlled psychological experiments. In analyzing a Chinese newspaper over a period, he found only one example of a counterfactual argument, and that was in a translation of a speech by Henry Kissinger. The writings of Mao Tse-Tung, who had studied Western political writings for many years, do contain many such Westernized constructions. Bloom noted, "while Westerners find Mao's writings relatively easier to read than typical Chinese prose, and his logic relatively more accessible, I have been told on repeated occasions by people with extensive experience in mainland China that, for the Chinese, the opposite is very much the case." With its abstract nominalizations and hypothetical expressions, English readily leads to abstract theories, while Chinese tends to be more concrete. For science and philosophy, the two languages have opposite advantages and disadvantages. English leads the speaker (or thinker) to a general principle, but Chinese keeps the speaker closer to the facts and encourages a more thorough search of alternatives.

Different conceptual systems may be internally consistent, but incompatible with one another. In *Either/Or*, Kierkegaard presented the world views of two men: the first man had an esthetic view of the world, and the second an ethical view. Each view contained a comprehensive set of mutually compatible concepts. In terms of them, each man could give a coherent interpretation of his experience. Yet communication between the two broke down because their concepts were incompatible. Even when they used the same words, their concepts were oriented in conflicting directions. Compatible concepts form self-contained systems, and knowing one leads to the discovery of others. Every concept is compatible with its opposite: good and evil, beauty and ugliness, justice and injustice. In the *Book of the Tao*, Lao Tzu said, "When everyone recognizes the good as good, there is beginning of evil. When everyone recognizes the beautiful as beautiful, there is the beginning of ugliness." This seemingly paradoxical statement refers to the interdependence of all the concepts in a compatible set. When smooth, tender skin is classified as beautiful, then coarse, wrinkled skin becomes ugly. In such terms, an elephant may become ugly, even though an elephant, on its own terms, is a very beautiful animal.

Kierkegaard illustrated incompatible conceptual systems through a dialog between two persons. One individual may use incompatible systems to express different aspects of the same situation. An ecological approach to a forest views it as a system of flora and fauna with complex interdependencies. An esthetic approach emphasizes the beauty and variety of the sights, sounds, and smells. A business approach views it as a source of pulp for paper mills. And a religious approach views it as a manifestation of the creativity and harmony of the universe. All these views are true to a certain extent, but none of them is absolutely true. Whenever a concept expresses an aspect of reality, it is partly true, but it is also partly false because the very point of view it espouses forces the exclusion of other, equally valid points of view.

Since concepts are inventions of the mind imposed upon experience, there is no reason to suppose that one set of concepts is more natural or fundamental than another. As Kierkegaard showed in *Either/Or*, two radically different views of life can give equally comprehensive interpretations of experience. There is no form of logic by which one man can refute the other's position because logic can only develop the implications of a compatible set of concepts. It cannot form a bridge between two incompatible conceptual systems. This idea is the starting point for a conceptual theory of relativity. In physics, relativity makes all coordinate systems equally fundamental, although the mathematical equations may be simpler in one system than in another. In philosophy, the analogous principle of conceptual relativity treats all self-consistent conceptual systems as equally fundamental. Some systems may be more comprehensive than others or better attuned to human needs and aspirations. But no single system of concepts can ever capture all of human experience.

Conceptual relativity sets limitations on the generality of conceptual analysis. The concept types and schemata discovered by the analysis hold only for a single culture, language, or domain of discourse.

Leibniz's dream of a universal lexicon of concepts that would be fixed for all time is doomed to failure. For a special application or range of applications, it is still possible to have restricted lexicons that are adequate to support knowledge-based systems. A universal expert system, however, would require a method for freely inventing new concepts for any possible domain. Such a system would require learning and discovery techniques that are far beyond present capabilities.

Discrete concepts divide the world into discrete things. The arbitrariness of this division is a common theme of Oriental philosophers. Lao Tzu said, "The Nameless is the origin of heaven and earth, the Named is the mother of all things." The world flows according to the unnamed Tao, but the differentiation of the world into discrete objects is a consequence of the discreteness of the conceptual mechanisms and the words that reflect them. By meditation on paradoxical sayings or *koans*, Zen Buddhism seeks to undermine a person's conceptual system and promote a direct experience of conceptual relativity. The process cuts through many years of cherished beliefs and automatic ways of thinking and acting. It requires a painful letting go of familiar habits. But the result is a blissful state of Enlightenment where the anxieties based on the old system of concepts melt into insignificance. The most detailed statement of the Buddhist theory of knowledge comes from the *Lankavatara Sutra* (Goddard 1938):

- Appearance knowledge gives names to things. It "belongs to the word mongers who revel in discriminations, assertions, and negations."
- Relative knowledge does more than classifying. "It rises from the mind's ability to arrange, combine, and analyze these relations by its powers of discursive logic and imagination, by reason of which it is able to peer into the meanings and significance of things."
- Perfect knowledge "is the pathway and the entrance into the exalted state of self-realization of Noble Wisdom." Perfect knowledge does not rule out the use of words and concepts, but it goes beyond them to a state of nonattachment to any particular conceptual system.

Concepts are useful fictions that are not absolute. There is a Buddhist saying that words are like a finger pointing to the moon: one who focuses only on the words and the concepts they symbolize will miss the reality they express, just as one who looks only at the finger will not see the moon it points to. Nonattachment to any system does not mean ignorance of all systems; appearance knowledge and relative knowledge are important for everyday life. The enlightened one is free to use concepts, but is not bound to them as absolute. Yet the path to enlightenment requires a painful abandonment of the comfortable old ways of thinking before any assurance is offered that the new way is better.

7.4 Creativity and Intelligence

The most obvious way to determine whether a computer is intelligent is to give it an IQ test. Evans (1968) designed a program called ANALOGY to solve geometric analogy problems of the kind that appear on IQ tests. It scored about as well as a typical high-school student. Evans made no claim that his program was as smart as a high-school student, but Good (1965) proposed the following definition:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the *last* invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.

That statement is merely a definition and, by itself, makes no claim about existence. Good went on, however, to make a further claim:

> It is more probable than not that, within the twentieth century, an ultraintelligent machine will be built and that it will be the last invention that man need make.

Half the time available for Good's prediction has elapsed, and the chance of its coming to pass in the twentieth century is zero. Progress has been made since 1965, many practical applications of AI have been implemented, but deeper analysis of the problems has led to some solutions and to many new problems.

Good's prediction was based on the assumption that intelligence, once simulated, would increase as a direct function of computer speed and storage capacity. IQ scores expressed in simple numbers foster that illusion. Yet IQ scores do not represent the full range of human mental abilities. Besides the issue of cultural influence, there are fundamental questions about the existence of other factors that cannot be measured by IQ tests. Creativity is one such factor that leads to speculation, but seldom to precise theoretical constructs. Most often, psychological testers seem to regard creativity as an annoying residue that spoils the correlation between IQ scores and scholastic achievement.

Getzels and Jackson (1962) carried out a study that clearly shows the importance of mental talents other than IQ. They studied a group of adolescents in a private school where the average IQ was quite high — 132. To measure creativity, they administered various tests such as ability to discover multiple associations for words, invent uses for things, find hidden shapes, and make up stories. To get as sharp a distinction as possible between the talents measured by IQ and creativity scores, the experimenters chose not to study those students who scored high on both types of tests or those who scored low on both. Instead, they selected two groups: the high IQ group scored high on the IQ tests, but low on creativity; and the high creatives scored comparatively low on IQ, but high on creativity.

Although one might question what the creativity tests really measure, there is no doubt that they measure a type of mental aptitude quite different from IQ. The average score for the high IQ group was 150, and for the high creative group was 127, five points below the school average. Yet both groups scored significantly better than the school average in scholastic achievement. In fact, the high creatives did slightly better than the high IQ students. The high creatives were not typical overachievers: they did not score higher on tests of motivation, they did not study as much, they were not as well liked by their teachers, and they were sometimes considered lazy. Yet despite their handicap of 23 points in IQ, their poorer study habits, their alleged laziness, and their lower estimation in the eyes of the teachers, they did as well as or better than the high IQ students in scholastic achievement.

Several personality traits correlate strongly with the two kinds of ability. The most prominent trait of the high creatives was a strong sense of humor. They rated it much higher than the high IQ group in traits they considered desirable, and they displayed it in their stories and drawings. The students in the high IQ group were generally better organized, less tolerant of chaos, more predictable, and more conventional — but they wrote boring stories. Wallach and Kogan (1965) explicitly characterized each of the four groups:

- *High creativity, high intelligence.* "These children can experience within themselves both control and freedom, both adultlike and childlike kinds of behavior."
- *High creativity, low intelligence.* "These children are in angry conflict with themselves and with their school environment and are beset by feelings of unworthiness and inadequacy. In a stress-free context, however, they can blossom forth cognitively."
- *Low creativity, high intelligence.* "These children can be described as "addicted" to school achievement. Academic failure would be perceived by them as catastrophic, so that they must

continually strive for academic excellence in order to avoid the possibility of pain."

- *Low creativity, low intelligence.* "Basically bewildered, these children engage in various defensive maneuvers ranging from useful adaptations such as intensive social activity to regression such as passivity or psychosomatic symptoms."

Correlations of mental aptitudes with personality traits do not explain the underlying mechanisms. Bruner (1960) distinguished the processes of creative intuition from the analytic thinking measured by IQ tests:

> Analytic thinking characteristically proceeds a step at a time. Steps are explicit and usually can be adequately reported by the thinker to another individual. Such thinking proceeds with relatively full awareness of the information and operations involved. It may involve careful and deductive reasoning, often using mathematics or logic and an explicit plan of attack. Or it may involve a step-by-step process of induction and experiment.... Intuitive thinking characteristically does not advance in careful, well-planned steps. Indeed, it tends to involve maneuvers based seemingly on an implicit perception of the total problem. The thinker arrives at an answer, which may be right or wrong, with little if any awareness of the process by which he reached it. (pp. 57-58)

When Bruner wrote this passage, the physiological work on left-brain vs. right-brain processes had not yet been done. Analytic thinking has sometimes been characterized as a typically left-brain function, and intuition as a right-brain function. That distinction, however, is too simplistic, and the physiological basis of creativity vs. intelligence or analytic thinking vs. intuition is still unknown.

Creativity and intelligence are apparently unrelated mental aptitudes that are not correlated with each other. Neither one is well defined or well understood in human beings, and the assumption of a single, unified faculty for either one is premature. Instead of just two independent mental aptitudes, Lowen (1982) distinguished sixteen aptitudes. He started with Jung's four categories of feeling, sensation, intuition, and thought, then divided each of those categories according to an extroverted vs. an introverted approach, and finally subdivided each of those categories according to a detailed vs. a global processing style. The resulting sixteen categories are a finer classification of cognitive abilities than a gross dichotomy of analytic vs. intuitive or intelligent vs. creative. Whether or not Lowen's categories prove to be more fruitful than other classifications, they emphasize the complexity of the processes taking place in the brain and the inadequacy of a single factor like IQ as a measure of mental ability.

Because of the different kinds of mental aptitudes, one should not expect to find a computer that uniformly simulates human abilities in all areas. From their earliest beginnings, computers far surpassed people in arithmetic. Today, they surpass all but the masters in chess playing. In pattern recognition, they can read printed material, but they cannot match a child in reading handwriting or recognizing speech. Computers are overwhelmingly superior to people for rote memory, acceptable for certain specialized kinds of learning, but very poor at tasks that require insight and complex analysis. They are good at applying a standard problem-solving method to many different cases, but they are poor at discovering new methods. Current AI simulations behave like people with high IQ, but exceptionally low creativity:

- IQ tests appear to measure speed and accuracy in using a given set of conceptual structures.
- Creativity is the ability to recognize the inadequacy of the old conceptual structures and to invent new ones to replace them.

A conceptual processor that has an encyclopedia of information might achieve an IQ score higher than any human being, yet not have the creativity of a two year old child. Such a machine would not qualify as ultraintelligent according to Good's definition; a better name for it would be Superclerk. Limited precursors of Superclerk are available today, and better ones will be built within the next two decades.

In Lowen's categories, Superclerk would be strongly biased towards verbal, analytic, discursive modes of thought and would have poorly developed intuition. It might not understand a joke, but it would be able to follow directions in English and give helpful explanations. It would not propose new plans on its own initiative, but it would be good at detecting inconsistent and incomplete plans that were given to it. It could read English texts on subjects about which it already had a well-developed set of concepts, but it would be confused by most literary texts and by scientific texts for which its predefined concepts were inadequate. Its novels and poetry would be boring, but it would excel at reading and writing form letters. Because of its uneven simulation of human abilities, Superclerk would not rival human intelligence; but like computers in general, it would make an excellent assistant for routine chores.

7.5 Science as a Mythology

Constructing a scientific theory means forging a new system of concepts for interpreting the world. Such a construction has a great deal in common with the process of creating a fictional world, as described by the novelist Vladimir Nabokov (1980):

> The material of this world may be real enough (as far as reality goes) but does not exist at all as an accepted entirety: it is chaos, and to this chaos the author says "Go!" allowing the world to flicker and to fuse. It is now recombined in its very atoms, not merely in its visible and superficial parts. The writer is the first man to map it and to name the natural objects it contains. Those berries there are edible. That speckled creature that bolted across my path might be tamed. That lake between those trees will be called Lake Opal or, more artistically, Dishwater Lake. That mist is a mountain — and that mountain must be conquered.

Like the artist, the scientist is confronted with a chaos of data. Out of that chaos, he or she must define the atoms that make up the world, recombine them "not merely in the visible and superficial parts," map the world, and name the natural objects it contains.

In either a work of fiction or a work of science, the task of creation is the same, and as Nabokov continued, "the boundary line between the two is not as clear as is generally believed." One might object that fiction is not true. It just tells stories — popular entertainment that is hardly comparable to precise, experimental science. Yet a work of fiction may contain as much truth as a work of nonfiction; and much of nonfiction may be mistaken, misleading, or wrong-headed. As this chapter has emphasized, the concepts in which supposed truths are expressed are essentially fictions. They may capture some aspect of truth, but there are infinitely many possible truths that they ignore, obscure, or distort.

Truth is only a measure of how well our mental models fit our observations of the world. Since finite, discrete concepts can never form a perfect model of continuous reality, a truly precise, truly objective science is not possible even as an ideal. The truth of any model must be limited to those few aspects of the world that the designer of the model chose to represent. Even for those aspects that the model covers, its truth is limited by the accuracy of the measuring instruments used to test its correspondence with the world: if the only standard of weight is a butcher scale, it is meaningless to say that a particular steak weighs 12.0695 ounces. According to the measuring instruments available five millennia ago, the ancient myths corresponded quite well with reality. Since they covered all aspects of all human concerns, one might even say that they were more true for their society than modern science is for ours.

There is no discontinuity between the thinking processes underlying modern science and the thinking represented in the ancient myths. They both stem from the same impulse — to speculate about phenomena in order to find explanations that make sense out of experience. But to be a science, a mythology must also satisfy some stringent criteria:

- *Predictive.* It must make clearly defined predictions under specific conditions.
- *Empirically testable.* The predictions must be testable by experiments that can be repeated by anyone who has suitable equipment and technique.
- *Cumulative.* The result of one person's theories and experiments must be stated precisely enough that other people can devise further theories and experiments that build upon them.

The scientist is a mythopoet who constructs a system of concepts for interpreting experience and weaves them into a coherent story. But science adds the discipline of prediction, testing, and building upon the results of others. Science is mythology plus discipline.

Without the discipline of scientific technique, speculation becomes mere fantasy. Parapsychology, for example, is speculation supported by an impressive amount of anecdotal evidence. Jahn (1982), the dean of engineering at Princeton, published a sympathetic review of psychic phenomena in the *Proceedings of the IEEE*. Yet no one has found a way of formulating precise, testable, cumulative hypotheses about such phenomena. Perhaps someone may do so in the future, but until then, parapsychology is not a science. By the same criteria, some of the work in AI barely qualifies as a science: people write a program, show some interesting output, and then describe the program in such a vague, qualitative way that no one else can build on their results without duplicating nearly all the work that went into the original program. The difference between AI and parapsychology, however, is that much of AI has been formulated in a precise, testable, cumulative way. All science may indeed be a mythology, but not all mythology qualifies as a science.

The concepts and schemata that direct human thought also determine how scientists carry out their research and interpret their results. Kuhn (1970) called those schemata *paradigms*. He emphasized that there is no such thing as "neutral data." The tracks in a bubble chamber are data only because of complex theories that led to the construction of the chamber, to the design of experiments for producing the tracks, and to the selection of "significant" tracks from commonplace events and background noise. Once a set of conceptual categories has been defined, scientific methodology is purely objective. But the selection of problems to study and the choice of concepts for describing them is a subjective judgment based on a scientist's personal preferences and the currently fashionable trends. The fashionable paradigms determine what questions are asked, what experiments are performed, and what books and articles are published and read. Scientific revolutions occur when one paradigm is replaced by another: associationism by behaviorism, and behaviorism by cognitive psychology.

Even physics, the standard of precision for all experimental science, is a mythology created by human minds guided by the paradigms of the day. Whitehead (1954) expressed the shock that he and other scientists felt when they realized that physics is a fallible human creation:

> I had a good classical education, and when I went up to Cambridge early in the 1880's my mathematical training was continued under good teachers. Now nearly everything was supposed to be known about physics that could be known — except a few spots, such as electromagnetic phenomena, which remained (or so it was thought) to be coordinated with the Newtonian principles. But for the rest, physics was supposed to be nearly a closed subject. Those investigations to coordinate went on through the next dozen years. By the middle of the 1890's there were a few tremors, a slight shiver as of all not being quite

> secure, but no one sensed what was coming. By 1900 the Newtonian physics was demolished, done for! Still speaking personally, it had a profound effect on me; I have been fooled once, and I'll be damned if I'll be fooled again! Einstein is supposed to have made an epochal discovery. I am respectful and interested, but also skeptical. There is no more reason to suppose that Einstein's relativity is anything final, than Newton's *Principia*. The danger is dogmatic thought; it plays the devil with religion, and science is not immune from it. (p. 277)

Frederick Thompson noted that students at Cal Tech experience a similar shock in their junior year. Bright students with high scores in mathematical aptitude arrive eager to learn all the wonderful truths of science. For the first two years, they really believe it. But in the third year, they study issues in the philosophy of science and discover that it is all a myth. It is a myth with high predictive value, and no other myth has been found to be more accurate. Yet it is not an Eternal Verity, but simply our best guess about how the universe works. When they come to that realization, many of the students go through a profound emotional crisis. Some of them never recover. But the best ones emerge with a deeper understanding of science and a better ability to do original research.

Not only do science and mythology stem from the same urge to speculate and explain, they also serve many of the same functions. According to Campbell (1968), a mythology serves four functions (p. 609):

- *Metaphysical-mystical*. A mythology awakens and maintains "an experience of awe, humility, and respect" in recognition of the ultimate mysteries of life and the universe.
- *Cosmological*. It provides an image of the universe and an explanation of how it operates.
- *Social*. It validates and maintains an established order.
- *Psychological*. It supports "the centering and harmonization of the individual."

In modern western culture, science has largely taken the place of traditional religions in serving these functions. Many millennia ago, orally recited myths represented the best scientific thought of the time. Today, "objective" science is one of our most widely accepted myths.

7.6 Minds and Machines

The subtitle of this book is *Information Processing in Mind and Machine*. A lot has been said about representing and processing information in machines, a lot has been said about processes in the human brain, but mind itself has never been defined as a formal construct in this theory. Various schools of philosophy have held differing opinions about mind and its relationship to the body:

- *Dualistic.* Mind is a substance separate from the body; it monitors the state of the body and directs its actions.
- *Epiphenomenal.* Only bodies exist, and the mental activity that is accessible to introspection is merely a by-product of neural processes.
- *Mentalistic.* Only minds exist, and bodies are merely illusions or by-products of intercommunicating minds.
- *Conceptual.* The concept of mind belongs to a complete system for talking about people and their ways of knowing, believing, understanding, and intending. Neurophysiology provides a totally different system of concepts for describing how the brain works. Although the mind depends on the brain, mental concepts are not definable in neural terms.

The conceptual position, which is the one adopted in this book, derives from the principle of conceptual relativity: a mental description and a physiological description of behavior are internally consistent, but mutually incompatible views, and neither one can be reduced to or be explained in terms of the other. Like Kierkegaard's dichotomy between an esthetic and an ethical view of life, mental terms and neural terms involve concepts and assumptions that are not interdefinable: anger may be correlated with an increase of adrenaline in the blood, but anger is not defined as the level of the adrenaline nor as the concomitant brain states. Just as the beauty of a symphony depends on the sequence of tones and harmonies, the activities of the mind depend on processes in the brain. But it is no more possible to define mind in terms of the brain than it is to define beauty in terms of tones and harmonies.

To illustrate the conceptual position, consider the concept COURAGE. Suppose that someone claims that a computer is courageous because it shows a firmness of purpose and lack of fear in executing any program given to it. But such a claim is silly. Since a computer is not capable of experiencing fear, a lack of fear means nothing. Furthermore, its firmness of purpose is nothing more than mechanical plodding. A horse may show courage in plodding through an icy storm because it is capable of having fear and, by its training, has learned to overcome that fear. Besides the ability to experience fear, courage presupposes honorable intentions, an understanding of the situation, and a reasonable belief that the benefits of the action are worth the risk. If any of these attributes are missing, a dangerous act is not courageous: without honorable intentions, it is despicable; without understanding, it is mistaken; and without a reasonable belief in the benefits, it is foolhardy. Outward signs can show that a person is undertaking a dangerous act without showing fear; but by themselves, they do not prove that the act is courageous. Only a knowledge of the person's mental state can distinguish a courageous act from a despicable, mistaken, or foolhardy act. This discussion illustrates two points: first, the concept of courage presupposes the related concepts of fear, understanding, intentionality, and rational assessment of risks (at least at the level of the higher mammals); second, observable behavior can suggest, but not prove the presence of a mental trait.

Like courage, the concepts of mind, thought, intentions, needs, and understanding belong to a unified, consistent system for interpreting the behavior of people and higher animals. Those terms are interdependent: a speaker cannot apply one of them to a person, animal, or computer system without admitting that the others apply as well. Applying those terms to computer systems raises serious philosophical problems. There are AI programs that simulate emotions, with a *happiness scale* that goes from -10 to +10. But a computer (or person) that simulates an emotion does not experience that emotion. Many AI systems can map an English sentence to a logical formula, derive other formulas from it, and then map those formulas back into English sentences. To say that such a system understands English is equivalent to saying that a program understands symbolic logic. Yet a person can understand an English sentence without realizing all its implications; conversely, a theorem prover can derive the implications of a proposition without being said to understand the proposition.

Even though behavior is governed by the brain, a detailed physiological description would obscure rather than explain normal behavior. Concepts of mind and intentions are a better basis for understanding human behavior than a complete wiring diagram of the brain. As computer systems increase in complexity, the engineers and programmers who design them work on only a small part at a time; complete flowcharts and wiring diagrams of current systems are beyond the capacity of any human being to comprehend as a whole. Large AI systems increase the levels of complexity even further. For such systems, mental analogies may become necessary to make them intelligible to people. Boden (1981) described hypothetical cases where the behavior of a robot could be more easily explained in terms of intentions than in conventional programming terms. In such a case, mental terms may be justified, but they lose their significance if they are casually applied to any AI system that the designer wishes to glorify.

Until one is prepared to say that a computer system has emotions, needs, and intentions, one cannot say that it has understanding. When the qualifier *simulated* can be dropped from the term *emotion*, then it can be dropped from *understanding, thinking*, and *mind* as well. For the AI systems available today, a casual use of terms like *thinking* or *understanding* is a sloppy and misleading practice. Those terms may have a dramatic effect, but they lead to confusion, especially for novices and people who are outside the AI field. McDermott (1976) maintained that they even have a mind-numbing effect on experts within the field.

Wittgenstein (1953) and Ryle (1949) deliberately avoided talk about the neural processes that support language and thought because such processes are irrelevant to understanding the meaning of language. Since Wittgenstein never talked about mental processes, some people have confused his position with a behavioristic denial of such processes. Yet Wittgenstein himself was careful to avoid that confusion. He criticized idle speculation about "the yet uncomprehended process in the yet unexplored medium," but he added, "And now it looks as if we had denied mental processes. And naturally we don't want to deny them." The assumption Wittgenstein denied was that an understanding of neural processes is needed to understand the normal use of language. Children learn the meaning of a word by seeing how adults use it. Philologists determine the meaning of a Latin word by analyzing Latin texts, even though the brains that composed those texts have long since crumbled into dust. Conceptual analysis, either formal or informal, is the basic method of determining meanings, and it requires no peering into the inner workings of brains or machines.

Wittgenstein declined to speculate about mental processes, partly because they were unobservable, but primarily because they cannot, in principle, explain meaning. But some of his followers, such as Malcolm (1977), went much further than the mas- ter in dismissing AI, theoretical linguistics, and cognitive psychology as a "mythology of inner guidance systems" (p. 169). Although a study of the human brain cannot explain the meaning of a word, it might uncover the neural mechanisms that process words. In the years since Wittgenstein wrote the *Philosophical Investigations*, new techniques have been found for studying those mechanisms:

- Measuring instruments for observing the electrical and chemical activity in various parts of the brain,
- Reaction-time experiments that can discriminate processes that are almost instantaneous from those that take a few additional milliseconds,
- Formal analyses of language and the mechanisms that generate it or interpret it,
- Detailed hypotheses about language and thought that can be simulated on a digital computer.

What these studies analyze are mechanisms, not meanings. Malcolm is justified in doing conceptual analysis of language without considering "inner guidance systems." But he is wrong in dismissing those systems as unimportant, uninteresting, or even nonexistent. Cognitive psychology may suggest an optimal form for representing and processing meaning, but conceptual analysis is also needed to determine the content that is represented in those forms.

The conceptual position on mind undermines the dualistic distinction between body and mind and thereby removes one prop in the argument for an immortal soul as a substrate for mind. Yet the existence or nonexistence of immortal souls is irrelevant to the question of whether artificial intelligence is possible. If extrasensory perception, telekinesis, and life after death require some sort of psychic ether, then that ether might be just another substrate for representing information and would be computationally indistinguishable from a material substance. If someone should find incontrovertible evidence that conceptual graphs are not represented in the brain but in a psychic ether, that would simply mean that AI would be simulating psychic ether on a computer rather than neural processes. On the other hand, it may turn out that the human brain or some psychic ether relies on continuous processes that are intrinsically different from anything that can be simulated on a digital computer. No one knows for sure, and the only way to see where the limitations might be is to try as hard as possible to see what can be done with the means available.

To return to the question of whether the word *mind* is justified in the title of this book, one can say that the primary justification is that *mind* is a short English word that nicely balances the word *machine*. It tells the reader that this book says something about the way people think and the way computers can simulate thinking. Nowhere, however, does *mind* appear in the formal definitions, assumptions, and theorems. Although this book discusses neural processes that are correlated with activities of the human mind, the word *mind* itself does not refer to any single one of them. An analysis of the concept of mind is an important philosophical issue, but the analysis cannot be reduced to programming or physiological terms.

7.7 Problems for Cognitive Science

Until the nineteenth century, psychology was considered a part of philosophy. The pioneers in the field were philosophers such as Aristotle, Locke, and Kant. Theoretical study of language was also part of philosophy, while the study of word forms was part of philology. When psychology and linguistics became distinct subjects, psychologists developed more precise experimental techniques, and linguists devoted more attention to the details of language. With the improvement in methodology, however, came an increasing specialization. Philosophers, psychologists, and linguists all studied aspects of language, thought, and reasoning, but with little awareness of the results and techniques of the related disciplines. Purity of method became a barrier to collaboration.

With the advent of artificial intelligence, a new methodology, computer simulation of thought processes, became possible. AI is an engineering discipline that uses results from all three fields: knowledge representation from philosophy, grammar rules from linguistics, and mental phenomena from psychology. In response to AI, the new field of cognitive science emerged as a re-integration of philosophy, psychology, and linguistics. What characterizes the field is not a common theory, a common methodology, or an established body of facts. Instead, it has three characteristics:

- Receptiveness to philosophy, psychology, linguistics, and AI,
- Study of behavior, not as an end in itself, but as indirect evidence for the mental processes that produce the behavior,
- Theoretical analysis of phenomena instead of the antitheoretical bias of behaviorism.

These characteristics are so general that they impose few constraints on what cognitive scientists do or how they talk about what they do. On a narrowly defined topic, people working with the same methodology can share insights on issues that are uppermost in everyone’s mind. Cognitive science, however, suffers from the problem of many interdisciplinary fields: AI researchers have a lot to say to other AI people; psychologists have a lot to say to other psychologists; but when they try to talk to each other, they continue to use the terminology and research paradigms of their parent disciplines. The difficulty in communication is one more example of conceptual relativity: a given set of phenomena can be described in many different conceptual systems, and the concepts in one system may be incompatible, or at least incommensurate, with those in another.

Cognitive science will not become a clearly defined discipline until it develops its own research paradigms and methodologies. Some of the methodologies may be inherited from the parent disciplines. Others will evolve out of collaboration on serious problems that researchers in all disciplines recognize as important. The following kinds of problems would benefit from a collaboration of two or more disciplines:

- *Ecological psychology.* Laboratory experiments in psychology set up artificial tasks like memorizing lists of nonsense syllables. The attempt to study the way people behave outside the laboratory is called *ecological psychology.* Yet the wealth of behavior in natural settings is so complex that it almost defies classification. Is that complexity, as Simon (1969) suggested, the result of a simple organism interacting with a complex environment, or is it the result of an inherent complexity in the human brain? The 3 billion neurons in the brain give an enormous scope for complexity, but is that complexity built up by replicating a few simple forms many times over, or are the underlying forms themselves complex? Are the simplified laboratory experiments fundamentally flawed because they cannot deal with the full complexity of human behavior? If so, how can the complexity contributed by the environment be distinguished from the simplicity or complexity inherent in the brain?
- *Strata of mentation.* The human brain is based on a three-level hierarchy: pathic, iconic, and noetic (Section 2.7). Yet except for some isolated attempts, AI simulations are almost exclusively devoted to a noetic approach. What is the relation between emotions and thought? Is the three-level brain a relic of evolutionary history, or is it essential to a human-like intelligence?
- *Combining syntax, semantics, and pragmatics.* The most complete grammars for English, such as Quirk et al. (1972), are written as informal descriptions. The most complete formal grammars have been developed by computational linguists, such as Sager (1981), but they are primarily syntactic. Can such grammars be extended with a semantic component based on conceptual graphs, or must they be completely rewritten to accommodate semantics? Pragmatics has been handled by some programs, but often in an *ad hoc* way. Can pragmatic rules be generalized and expressed in a formalism that is compatible with conceptual graphs, APSG rules, and other notations?
- *Lexicography.* Natural language processing requires dictionaries. Tapes for some of the standard English dictionaries are available, but their formats are designed for people to use, not computer systems. For computational linguistics, a more formal notation, such as conceptual graphs, is necessary. Appendix B suggests the form of such dictionaries. Exactly what information should be contained in them? How should they be constructed? Can they be generated automatically from conventional dictionaries?
- *Ontology.* A catalog of concept and relation types is an essential basis for formal lexicography. But since every generic concept has an implicit existential quantifier, a catalog of concept types is also an *ontology* — a catalog of modes of existence. Formulating such a catalog for concrete types is difficult enough, but the more abstract concepts involve complex philosophical issues. Lexicography must be done as a concerted effort of philosophy, logic, linguistics, philology, and AI.
- *Conceptual relativity.* A definition of all concepts by necessary and sufficient conditions is impossible. Except for artificially constructed concepts in formal systems, essentially no concepts can be so defined. Because of that limitation, a complete dictionary of English that captures the full meaning of every word (or even the full meaning of just a single word) is impossible. Is it possible, however, to develop a series of specialized dictionaries, each of which would provide a workable approximation to English for a specialized topic? How could a system recognize when the bounds of its dictionary had been exceeded? Could an open-ended family of specialized dictionaries ever grow to approximate the full richness of truly natural

language?

- *Humor.* A good sense of humor is one of the characteristics of a creative person, but it is present to some extent in nearly everybody. Humor depends on and promotes insights that lead to new ways of looking at a situation. How is humor related to conceptual relativity? Instead of being a frivolous adjunct to intelligence, could it hold the key to a creative, adaptive form of intelligence? Is there some relationship between humor and metaphor? Between humor and learning? How could one design a program that would invent original, truly funny jokes? What about a program that could appreciate jokes?
- *Historical perspective.* Feigenbaum (1980) deplored "the lack of cumulation of AI methods and techniques. The AI field tends to reward scientifically irresponsible pseudo-innovation. That is, it tends to reward individuals for reinventing and renaming concepts and methods that are well explored." Ideas generated at other research institutions or in related disciplines at the same institution tend to be ignored. People writing on AI often cite an unpublished remark by a colleague who rediscovered a point that Aristotle analyzed 2300 years earlier. The history of philosophy is important, not just as cultural background, but often as the most thorough way of gaining breadth and depth of insight into the full range of issues about mind, meaning, and language.
- *History of linguistics.* AI is not the only field that has neglected its historical roots. Wierzbicka (1980) noted that transformational linguists "behave as if Chomsky's *Syntactic Structures* was the first document of theoretical reflection on language in the entire history of human thought. An abyss has opened between theoretical linguistics as it is currently practiced in many linguistic centers and most linguistic ideas of the past. More than two thousand years of history, in the course of which (particularly in the later Middle Ages and in the seventeenth century) the best minds occupied themselves extensively with language and thought, are blithely or contemptuously dismissed. Aristotle, Boethius, Abelard, Peter of Spain, Roger Bacon, Wilkins, Leibniz, Arnauld, Peirce, Frege, Wittgenstein, Sapir, Jespersen, Bally, etc. are either ignored or solemnly, unwittingly and usually only partially replicated" (pp. 32-33). Studying that history is interesting in its own right, but even more importantly, current theories can be built on more solid ground if they incorporate the centuries of analysis and thinking that have gone before.
- *History of logic.* Standard symbolic logic is based on an extensional approach that has proved useful for foundational studies in mathematics. Yet for over two millennia, logic was dominated by semantic issues in natural language. The medieval Scholastics, in particular, elaborated complex and subtle intensional approaches that have never been analyzed and reinterpreted in modern terms. The various schools of Indian philosophy also elaborated complex logical schemes that have never been reformulated in modern notation. Can those approaches express aspects of meaning that are not captured by standard logic? What light can they shed on foundational issues for conceptual graphs and other AI systems?
- *Fuzzy logic.* Standard logic has only two quantifiers, *for all* and *there exists*. Natural languages, however, express a wide variety of intermediate forms, such as *many, most, almost all, few, very few,* and *practically none*. Zadeh (1982) claimed "almost everything that relates to natural language is a matter of degree." To represent those degrees, he developed *fuzzy logic*. Yet AI has been most successful in dealing with discrete logic. The question remains whether fuzziness occurs in human thought or in the mapping between discrete concepts and the outside world. What is the ultimate source of fuzziness in natural language? Is it in the continuous variability of the real world, the human thinking processes, or both? If a system should permit fuzziness in reasoning, at what stage should it occur?

- *Modes of communication.* Natural languages evolved under two constraints: the linear form of speech and the powers and limitations of the human conceptual system. If a better spoken form were possible, the course of language evolution over the past 30,000 years would have found it. But computer terminals can combine graphics, menus, pushbuttons, and sound in a single interface. Can such systems support a more efficient or accurate form of communication? What is the significance of the gestures and hand waving that usually accompany speech? Could joy sticks or other devices be used as an accompaniment to speech — possibly to emphasize, punctuate, or disambiguate the spoken form?

Each of these problems leads to dozens or hundreds of subtasks, each of which may require many years of research. Not all of them need to be solved in order to support useful applications of AI. But they represent the work that must be done before any system that approaches the versatility of human intelligence can be designed. Whether such systems will ever be possible is still an open question, but no such system will appear within the twentieth century.

Exercises

Imagine that you have been invited to contribute an article to a volume of collected papers on conceptual graphs. The articles will be grouped in seven areas — philosophical issues, psychological evidence, formalism, logic and computation, linguistic studies, knowledge engineering, and limits of conceptualization. Choose one of these seven areas, and write a 10 to 20 page article that would be suitable for publication in the volume. Sample topics include the following:

1. Map some English sentences into conceptual graphs, and look for constructions that are especially difficult to represent. Develop a systematic way of representing those constructions, and propose it as a general method.
2. Follow the guidelines of Section 6.3 for doing a conceptual analysis of some particularly difficult area. Describe your approach, the results of your analysis, and possibilities for simplifying or systematizing the analysis of similar areas.
3. Compare conceptual graphs with another formalism you may happen to know, such as transformational grammar, Montague grammar, or diagrams for database design. Map some extended example expressed in that formalism into the conceptual graph notation. Compare the two approaches for perspicuity, expressive power, completeness, naturalness, and ease of use.
4. Find some area where conceptual graphs are not expressive enough. Propose some extensions to handle that area, and generalize them into universal principles.
5. Explore the formal properties of conceptual graphs or the operations that may be performed on them, and prove some interesting theorems about them.
6. Write a computer program to implement some aspect of the theory. Describe the novel features of the program and show some sample output.
7. Select some topic in this book that was mentioned in only a brief sentence or paragraph. Develop that topic in depth, citing evidence from the published literature and experiments or analyses that you have done yourself.
8. Select some point in the book with which you disagree and develop counterarguments against it. Propose an alternative solution and show how it would either improve the theory or refute some aspect of it.
9. Select some psychological, linguistic, or philosophical point that was not mentioned in this

book. Relate it to conceptual graphs and show whether it helps to confirm the current theory or requires some extension or modification of it.

10. Design a programming language for processing conceptual graphs or the language interface to them. It may, for example, include the canonical formation rules as primitive operations and have a convenient notation for defining new types, schemata, and actions. Implement the language or at least describe how it might be implemented. What kinds of hardware and software would be needed to support it?

Suggested Readings

Haugeland (1981) presents a collection of readings on all points of view about AI. Dreyfus (1979) gleefully and unsympathetically pounces on and debunks rash pronouncements about AI. *The Mind's I*, edited by Hofstadter and Dennet, leans toward mystical visions of minds, both natural and artificial. Weizenbaum (1976) raises ethical questions about the desirability of artificial intelligence. Sloman (1978), Dennet (1978), and Boden (1981) are philosophers who have turned their attentions to AI.

The original book on cybernetics is by Wiener (1948). Ashby (1956) is still one of the best introductions to the subject. *The Metaphorical Brain* by Arbib seeks to combine AI with cybernetics and brain theory. *Engineering Intelligent Systems* by Glorioso and Osorio presents a survey of both AI and cybernetics. General systems theory (von Bertalanffy 1968, Klir 1978) is an abstract approach that has close ties to cybernetics. Fuzzy set theory and fuzzy logic are more closely related to cybernetics than to standard AI approaches. The collection of papers edited by Yager (1982) presents some of the recent developments in fuzzy set theory and its applications.

The classic statement of conceptual relativity is by Whorf (1956). In following Whorf, Bloom (1981) shows some important differences in the structure of the Chinese and English languages and their effects on thought and expression. *About Chinese* by Newnham describes the structure of Chinese and compares its forms to English. The notion of *perspectivism* by Ortega y Gasset is very close to the notion of conceptual relativity; for an introduction to Ortega's philosophy, see his book *The Modern Theme*. The distinction between an esthetic and an ethical viewpoint in Kierkegaard's *Either/Or* is a clear illustration of conceptual relativity. *Psychological Types* by Jung is a classic study of various styles of thinking and understanding; Lowen (1982) presents a finer subdivision of Jung's types.

Oriental philosophy is an important source of insights into conceptual relativity, but it is such a vast field that no short reading list can do justice to it. *Three Pillars of Zen* by Kapleau is good starting point for Zen Buddhism. For a study of Zen koans, see Miura and Sasaki (1965). The two volumes of *Buddhist Logic* by Stcherbatsky are a classic treatment of Buddhist philosophy and its relationships to Western philosophy and logic. The *Source Book in Chinese Philosophy* by Chan is a good collection of original sources; the *Book of the Tao* by Lao Tzu is especially important.