Philosophy

Robert A. Wilson

The areas of philosophy that contribute to and draw on the cognitive sciences are various; they include the philosophy of mind, science, and language; formal and philosophical logic; and traditional metaphysics and epistemology. The most direct connections hold between the philosophy of mind and the cognitive sciences, and it is with classical issues in the philosophy of mind that I begin this introduction (section 1). I then briefly chart the move from the rise of materialism as the dominant response to one of these classic issues, the mind-body problem, to the idea of a science of the mind. I do so by discussing the early attempts by introspectionists and behaviorists to study the mind (section 2). Here I focus on several problems with a philosophical flavor that arise for these views, problems that continue to lurk backstage in the theater of contemporary cognitive science. Between these early attempts at a science of the mind and today's efforts lie two general, influential philosophical traditions, ordinary language philosophy and logical positivism. In order to bring out, by contrast, what is distinctive about the contemporary naturalism integral to philosophical contributions to the cognitive sciences, I sketch the approach to the mind in these traditions (section 3). And before getting to contemporary naturalism itself I take a quick look at the philosophy of science, in light of the legacy of positivism (section 4).

In sections 5 through 7 I get, at last, to the mind in cognitive science proper. Section 5 discusses the conceptions of mind that have dominated the contemporary cognitive sciences, particularly that which forms part of what is sometimes called "classic" cognitive science and that of its connectionist rival. Sections 6 and 7 explore two specific clusters of topics that have been the focus of philosophical discussion of the mind over the last 20 years or so, folk psychology and mental content. The final sections gesture briefly at the interplay between the cognitive sciences and logic (section 8) and biology (section 9).

1 Three Classic Philosophical Issues About the Mind

i. The Mental-Physical Relation

The relation between the mental and the physical is the deepest and most recurrent classic philosophical topic in the philosophy of mind, one very much alive today. In due course, we will come to see why this topic is so persistent and pervasive in thinking about the mind. But to convey something of the topic's historical significance let us begin with a classic expression of the puzzling nature of the relation between the mental and the physical, the MIND-BODY PROBLEM.

This problem is most famously associated with RENÉ DESCARTES, the preeminent figure of philosophy and science in the first half of the seventeenth century. Descartes combined a thorough-going mechanistic theory of nature with a *dualistic* theory of the nature of human beings that is still, in general terms, the most widespread view held by ordinary people outside the hallowed halls of academia. Although nature, including that of the human body, is material and thus completely governed by basic principles of mechanics, human beings are special in that they are composed both of material and nonmaterial or mental stuff, and so are not so governed. In Descartes's own

terms, people are essentially a combination of mental substances (minds) and material substances (bodies). This is Descartes's *dualism*. To put it in more commonsense terms, people have both a mind and a body. Although dualism is often presented as a possible solution to the mind-body problem, a possible position that one might adopt in explaining how the mental and physical are related, it serves better as a way to bring out why there is a "problem" here at all. For if the mind is one type of thing, and the body is another, how do these two types of things interact? To put it differently, if the mind really is a nonmaterial substance, lacking physical properties such as spatial location and shape, how can it be both the cause of effects in the material world—like making bodies move—and itself be causally affected by that world—as when a thumb slammed with a hammer (bodily cause) causes one to feel pain (mental effect)? This problem of causation between mind and body has been thought to pose a largely unanswered problem for Cartesian dualism. It would be a mistake, however, to assume that the mind-body problem in its most general form is simply a consequence of dualism. For the general question as to how the mental is related to the physical arises squarely for those convinced that some version of materialism or PHYSICALISM must be true of the mind. In fact, in the next section, I will suggest that one reason for the resilience and relevance of the mind-body problem has been the *rise* of materialism over the last fifty years.

Materialists hold that all that exists is material or physical in nature. Minds, then, are somehow or other composed of arrangements of physical stuff. There have been various ways in which the "somehow or other" has been cashed out by physicalists, but even the view that has come closest to being a consensus view among contemporary materialists—that the mind *supervenes* on the body—remains problematic. Even once one adopts materialism, the task of articulating the relationship between the mental and the physical remains, because even physical minds have special properties, like intentionality and consciousness, that require further explanation. Simply proclaiming that the mind is not made out of distinctly mental substance, but is material like the rest of the world, does little to explain the features of the mind that seem to be distinctively if not uniquely features of physical minds.

ii. The Structure of the Mind and Knowledge

Another historically important cluster of topics in the philosophy of mind concerns what is in a mind. What, if anything, is distinctive of the mind, and how is the mind structured? Here I focus on two dimensions to this issue. One dimension stems from the RATIONALISM VS. EMPIRICISM debate that reached a high point in the seventeenth and eighteenth centuries. Rationalism and empiricism are views of the nature of human knowledge. Broadly speaking, empiricists hold that all of our knowledge derives from our sensory, experiential, or empirical interaction with the world. Rationalists, by contrast, hold the negation of this, that there is some knowledge that does not derive from experience. Since at least our paradigms of knowledge—of our immediate environments, of common physical objects, of scientific kinds—seem obviously to be based on sense experience, empiricism has significant intuitive appeal. Rationalism, by contrast, seems to require further motivation: minimally, a list of knowables that represent a prima facie challenge to the empiricist's global claim about the foundations of knowledge. Classic rationalists, such as Descartes, Leibniz, Spinoza, and perhaps more contentiously KANT, included knowledge of God, substance, and abstract ideas (such as that of a triangle, as opposed to ideas of particular triangles). Empiricists over the last three hundred years

or so have either claimed that there was nothing to know in such cases, or sought to provide the corresponding empiricist account of how we could know such things from experience.

The different views of the sources of knowledge held by rationalists and empiricists have been accompanied by correspondingly different views of the mind, and it is not hard to see why. If one is an empiricist and so holds, roughly, that there is nothing in the mind that is not first in the senses, then there is a fairly literal sense in which *ideas*, found in the mind, are complexes that derive from *impressions* in the senses. This in turn suggests that the processes that constitute cognition are themselves elaborations of those that constitute perception, that is, that cognition and perception differ only in degree, not kind. The most commonly postulated mechanisms governing these processes are *association* and *similarity*, from Hume's laws of association to featureextraction in contemporary connectionist networks. Thus, the mind tends to be viewed by empiricists as a *domain-general* device, in that the principles that govern its opera tion are constant across various types and levels of cognition, with the common empirical basis for all knowledge providing the basis for parsimony here. By contrast, in denying that all knowledge derives from the senses, rationalists are faced with the question of what other sources there are for knowledge. The most natural candidate is the mind itself, and for this reason rationalism goes hand in hand with NATIVISM about both the source of human knowledge and the structure of the human mind. If some ideas are innate (and so do not need to be derived from experience), then it follows that the mind already has a relatively rich, inherent structure, one that in turn limits the malleability of the mind in light of experience. As mentioned, classic rationalists made the claim that certain ideas or CONCEPTS were innate, a claim occasionally made by contemporary nativists—most notably Jerry Fodor (1975) in his claim that *all* concepts are innate. However, contemporary nativism is more often expressed as the view that certain implicit knowledge that we have or principles that govern how the mind works—most notoriously, linguistic knowledge and principles—are innate, and so not learned. And because the types of knowledge that one can have may be endlessly heterogeneous, rationalists tend to view the mind as a *domainspecific* device, as one made up of systems whose governing principles are very different. It should thus be no surprise that the historical debate between rationalists and empiricists has been revisited in contemporary discussions of the INNATENESS OF LANGUAGE, the MODULARITY OF MIND, and CONNECTIONISM.

A second dimension to the issue of the structure of the mind concerns the place of CONSCIOUSNESS among mental phenomena. From WILLIAM JAMES's influential analysis of the phenomenology of the stream of consciousness in his *The Principles of Psychology* (1890) to the renaissance that consciousness has experienced in the last ten years (if publication frenzies are anything to go by), consciousness has been thought to be the most puzzling of mental phenomena. There is now almost universal agreement that conscious mental states are a part of the mind. But how large and how important a part? Consciousness has sometimes been thought to exhaust the mental, a view often attributed to Descartes. The idea here is that everything mental is, in some sense, conscious or available to consciousness. (A version of the latter of these ideas has been recently expressed in John Searle's [1992: 156] *connection principle*: "all unconscious intentional states are in principle accessible to consciousness.") There are two challenges to the view that everything mental is conscious or even available to consciousness. The first is posed by the *unconscious*. SIGMUND FREUD's extension of our common-sense attributions of belief and

desire, our folk psychology, to the realm of the unconscious played and continues to play a central role in PSYCHOANALYSIS.
The second arises from the conception of cognition as information processing that has been and remains focal in contemporary cognitive science, because such information processing is mostly *not* available to consciousness. If cognition so conceived is mental, then most mental processing is not available to consciousness.

iii. The First- and Third-Person Perspectives
Occupying center stage with the mind-body problem in traditional philosophy of mind is the *problem of other minds,* a problem that, unlike the mind-body problem, has all but disappeared from philosophical contributions to the cognitive sciences. The problem is often stated in terms of a contrast between the relatively secure way in which I "directly" know about the existence of *my own* mental states, and the far more epistemically risky way in which I must infer the existence of the mental states of others. Thus, although I can know about my own mental states simply by introspection and self-directed reflection, because this way of finding out about mental states is peculiarly first-person, I need some other type of evidence to draw conclusions about the mental states of others. Naturally, an agent's behavior is a guide to what mental states he or she is in, but there seems to be an epistemic gap between this sort of evidence and the attribution of the corresponding mental states that does not exist in the case of self-ascription. Thus the problem of other minds is chiefly an *epistemological* problem, sometimes expressed as a form of skepticism about the justification that we have for attributing mental states to others. There are two reasons for the waning attention to the problem of other minds *qua problem* that derive from recent philosophical thought sensitive to empirical work in the cognitive sciences. First, research on introspection and SELF-KNOWLEDGE has raised questions about how "direct" our knowledge of our own mental states and of the SELF is, and so called into question traditional conceptions of first-person knowledge of mentality. Second, explorations of the THEORY OF MIND, ANIMAL COMMUNICATION, and SOCIAL PLAY BEHAVIOR have begun to examine and assess the sorts of attribution of mental states that are actually justified in empirical studies, suggesting that third-person knowledge of mental states is not as limited as has been thought. Considered together, this research hints that the contrast between first- and thirdperson knowledge of the mental is not as stark as the problem of other minds seems to intimate. Still, there is something distinctive about the first-person perspective, and it is in part as an acknowledgment of this, to return to an earlier point, that consciousness has become a hot topic in the cognitive sciences of the 1990s. For whatever else we say about consciousness, it seems tied ineliminably to the first-person perspective. It is a state or condition that has an irreducibly *subjective* component, something with an essence to be experienced, and which presupposes the existence of a subject of that experience. Whether this implies that there are QUALIA that resist complete characterization in materialist terms, or other limitations to a science of the mind, remain questions of debate.

2 From Materialism to Mental Science

In raising issue *i.*, the mental-physical relation, in the previous section, I implied that materialism was the dominant ontological view of the mind in contemporary

philosophy of mind. I also suggested that, if anything, general convergence on this issue has intensified interest in the mind-body problem. For example, consider the large and lively debate over whether contemporary forms of materialism are compatible with genuine MENTAL CAUSATION, or, alternatively, whether they commit one to EPIPHENOMENALISM about the mental (Kim 1993; Heil and Mele 1993; Yablo 1992). Likewise, consider the fact that despite the dominance of materialism, some philosophers maintain that there remains an EXPLANATORY GAP between mental phenomena such as consciousness and any physical story that we are likely to get about the workings of the brain (Levine 1983; cf. Chalmers 1996). Both of these issues, very much alive in contemporary philosophy of mind and cognitive science, concern the mind-body problem, even if they are not always identified in such old-fashioned terms. I also noted that a healthy interest in the first-person perspective persists within this general materialist framework. By taking a quick look at the two major initial attempts to develop a systematic, scientific understanding of the mind—late nineteenth-century introspectionism and early twentieth-century behaviorism—I want to elaborate on these two points and bring them together.

Introspectionism was widely held to fall prey to a problem known as the *problem of the homunculus*. Here I argue that behaviorism, too, is subject to a variation on this very problem, and that both versions of this problem continue to nag at contemporary sciences of the mind. Students of the history of psychology are familiar with the claim that the roots of contemporary psychology can be dated from 1879, with the founding of the first experimental laboratory devoted to psychology by WILHELM WUNDT in Leipzig, Germany. As an *experimental* laboratory, Wundt's laboratory relied on the techniques introduced and refined in physiology and psychophysics over the preceding fifty years by HELMHOLTZ, Weber, and Fechner that paid particular attention to the report of SENSATIONS.

What distinguished Wundt's as a laboratory of *psychology* was his focus on the data reported in consciousness via the first-person perspective; psychology was to be the science of immediate experience and its most basic constituents. Yet we should remind ourselves of how restricted this conception of psychology was, particularly relative to contemporary views of the subject. First, Wundt distinguished between mere INTROSPECTION, first-person reports of the sort that could arise in the everyday course of events, and experimentally manipulable self-observation of the sort that could only be triggered in an experimental context. Although Wundt is often thought of as the founder of an introspectionist methodology that led to a promiscuous psychological ontology, in disallowing mere introspection as an appropriate method for a science of the mind he shared at least the sort of restrictive conception of psychology with *both* his physiological predecessors and his later behaviorist critics. Second, Wundt thought that the vast majority of ordinary thought and cognition was *not* amenable to acceptable first-person analysis, and so lay beyond the reach of a scientific psychology. Wundt thought, for example, that belief, language, personality, and SOCIAL COGNITION could be studied systematically only by detailing the cultural mores, art, and religion of whole societies (hence his four-volume

Völkerpsychologie

of 1900–1909). These studies belonged to the humanities (*Geisteswissenshaften*) rather than the experimental sciences (*Naturwissenschaften*), and were undertaken by anthropologists inspired by Wundt, such as BRONISLAW MALINOWSKI. Wundt himself took one of his early contributions to be a solution of the mind-body problem,

for that is what the data derived from the application of the experimental method to distinctly psychological phenomena gave one: correlations between the mental and the physical that indicated how the two were systematically related. The discovery of psychophysical laws of this sort showed how the mental was related to the physical. Yet with the expansion of the domain of the mental amenable to experimental investigation over the last 150 years, the mind-body problem has taken on a more acute form: just how do we get all that mind-dust from merely material mechanics? And it is here that the problem of the homunculus arises for introspectionist psychology after Wundt. The problem, put in modern guise, is this. Suppose that one introspects, say, in order to determine the location of a certain feature (a cabin, for example) on a map that one has attempted to memorize (Kosslyn 1980). Such introspection is typically reported in terms of exploring a mental image with one's *mind's eye*. Yet we hardly want our psychological story to end there, because it posits a process (introspection) and a processor (the mind's eye) that themselves cry out for further explanation. The problem of the homunculus is the problem of leaving undischarged homunculi ("little men" or their equivalents) in one's *explanantia*, and it persists as we consider an elaboration on our initial introspective report. For example, one might well report forming a mental image of the map, and then scanning around the various features of the map, zooming in on them to discern more clearly what they are to see if any of them is the sought-after cabin. To take this introspective report seriously as a guide to the underlying psychological mechanisms would be to posit, minimally, an *imager* (to form the initial image), a *scanner* (to guide your mind's eye around the image), and a *zoomer* (to adjust the relative sizes of the features on the map). But here again we face the problem of the homunculus, because such "mechanisms" themselves require further psychological decomposition.

To be faced with the problem of the homunculus, of course, is not the same as to succumb to it. We might distinguish two understandings of just what the "problem" is here. First, the problem of the homunculus could be viewed as a problem specifically for introspectionist views of psychology, a problem that was never successfully met and that was principally responsible for the abandonment of introspectionism. As such, the problem motivated BEHAVIORISM in psychology. Second, the problem of the homunculus might simply be thought of as a challenge that *any* view that posits internal mental states must respond to: to show how to discharge all of the homunculi introduced in a way that is acceptably materialistic. So construed, the problem remains one that has been with us more recently, in disputes over the psychological reality of various forms of GENERATIVE GRAMMAR (e.g., Stabler 1983); in the nativism that has been extremely influential in post-Piagetian accounts of COGNITIVE DEVELOPMENT (Spelke 1990; cf. Elman et al. 1996); and in debates over the significance of MENTAL ROTATION and the nature of IMAGERY (Kosslyn 1994; cf. Pylyshyn 1984: ch.8).

With Wundt's own restrictive conception of psychology and the problem of the homunculus in mind, it is with some irony that we can view the rise and fall of behaviorism as the dominant paradigm for psychology subsequent to the introspectionism that Wundt founded. For here was a view so deeply indebted to materialism and the imperative to explore psychological claims only by reference to what was acceptably experimental that, in effect, in its purest form it appeared to do away with the distinctively mental altogether! That is, because objectively observable behavioral responses to objectively measurable stimuli are all that could be rigorously explored, experimental psychological investigations would need to be significantly curtailed, relative to those of introspectionists such as Wundt and Titchener. As J. B.

Watson said in his early, influential “Psychology as the Behaviorist Views It” in 1913, “Psychology as behavior will, after all, have to neglect but few of the really essential problems with which psychology as an introspective science now concerns itself. In all probability even this residue of problems may be phrased in such a way that refined methods in behavior (which certainly must come) will lead to their solution” (p. 177).

Behaviorism brought with it not simply a global conception of psychology but specific methodologies, such as CONDITIONING, and a focus on phenomena, such as that of LEARNING, that have been explored in depth since the rise of behaviorism. Rather than concentrate on these sorts of contribution to the interdisciplinary sciences of the mind that behaviorists have made, I want to focus on the central problem that faced behaviorism as a research program for reshaping psychology. One of the common points shared by behaviorists in their philosophical and psychological guises was a commitment to an *operational* view of psychological concepts and thus a suspicion of any reliance on concepts that could not be operationallycharacterized. Construed as a view of scientific *definition* (as it was by philosophers),operationalism is the view that scientific terms must be defined in terms of observableand measurable operations that one can perform. Thus, an operational definition of“length,” as applied to ordinary objects, might be: “the measure we obtain by laying astandard measuring rod or rods along the body of the object.” Construed as a view ofscientific *methodology* (as it was by psychologists), operationalism claims that thesubject matter of the sciences should be objectively observable and measurable, byitself a view without much content.The real bite of the insistence on operational definitions and methodology for psychology came via the application of operationalism to unobservables, for the various feelings, sensations, and other internal states reported by introspection, themselves unobservable, proved difficult to operationalize adequately. Notoriously, the introspective reports from various psychological laboratories produced different listings of the basic feelings and sensations that made up consciousness, and the lack of agreement here generated skepticism about the reliability of introspection as a method for revealing the structure of the mind. In psychology, this led to a focus on behavior, rather than consciousness, and to its exploration through observable stimulus and response: hence, behaviorism. But I want to suggest that this reliance on operationalism itself created a version of the problem of the homunculus for behaviorism. This point can be made in two ways, each of which offers a reinterpretation of a standard criticism of behaviorism. The first of these criticisms is usually called “philosophical behaviorism,” the attempt to provide conceptual analyses of mental state terms exclusively in terms of behavior; the second is “psychological behaviorism,” the research program of studying objective and observable behavior, rather than subjective and unobservable inner mental episodes.

First, as Geach (1957: chap. 4) pointed out with respect to belief, behaviorist analyses of individual folk psychological states are bound to fail, because it is only in concert with many other propositional attitudes that any given such attitude has behavioral effects. Thus, to take a simple example, we might characterize the belief that it is raining as the tendency to utter “yes” when asked, “Do you believe that it is raining?” But one reason this would be inadequate is that one will engage in this verbal behavior only if one *wants* to answer truthfully, and only if one *hears* and *understands* the question asked, where each of the italicized terms above refers to some other mental state. Because the problem recurs in *every* putative analysis, this implies that a behavioristically acceptable construal of folk psychology is not possible. This point would seem to generalize beyond folk psychology to representational psychology

more generally. So, in explicitly attempting to do without internal mental representations, behaviorists themselves are left with mental states that must simply be assumed. Here we are not far from those undischarged homunculi that were the bane of introspectionists, especially once we recognize that the metaphorical talk of "homunculi" refers precisely to internal mental states and processes that themselves are not further explained.

Second, as Chomsky (1959: esp. p. 54) emphasized in his review of Skinner's *Verbal Behavior*, systematic attempts to operationalize psychological language invariably smuggle in a reference to the very mental processes they are trying to do without. At the most general level, the behavior of interest to the linguist, Skinner's "verbal behavior," is difficult to characterize adequately without at least an implicit reference to the sorts of psychological mechanism that generate it. For example, linguists are not interested in mere noises that have the same physical properties—"harbor" may be pronounced so that its first syllable has the same acoustic properties as an exasperated grunt—but in parts of speech that are taxonomized at least partially in terms of the surrounding mental economy of the speaker or listener. The same seems true for *all* of the processes introduced by behaviorists—for example, stimulus control, reinforcement, conditioning—insofar as they are used to characterize complex, human behavior that has a natural psychological description (making a decision, reasoning, conducting a conversation, issuing a threat). What marks off their instances as behaviors *of the same kind* is not exclusively their physical or behavioral similarity, but, in part, the common, internal psychological processes that generate them, and that they in turn generate. Hence, the irony: behaviorists, themselves motivated by the idea of reforming psychology so as to generalize about objective, observable behavior and so avoid the problem of the homunculus, are faced with undischarged homunculi, that is, irreducibly mental processes, in their very own alternative to introspectionism. The two versions of the problem of the homunculus are still with us as a Scylla and Charybdis for contemporary cognitive scientists to steer between. On the one hand, theorists need to avoid building the very cognitive abilities that they wish to explain into the models and theories they construct. On the other, in attempting to side-step this problem they also run the risk of masking the ways in which their "objective" taxonomic categories presuppose further internal psychological description of precisely the sort that gives rise to the problem of the homunculus in the first place.

See also BEHAVIORISM; COGNITIVE DEVELOPMENT; CONDITIONING; EPIPHENOMENALISM; EXPLANATORY GAP; GENERATIVE GRAMMAR; HELMHOLTZ, HERMANN; IMAGERY; INTROSPECTION; LEARNING; MALINOWSKI, BRONISLAW; MENTAL CAUSATION; MENTAL ROTATION; SENSATIONS; SOCIAL COGNITION; SOCIAL COGNITION IN ANIMALS; WUNDT, WILHELM

3 A Detour Before the Naturalistic Turn

Given the state of philosophy and psychology in the early 1950s, it is surprising that within twenty-five years there would be a thriving and well-focused interdisciplinary unit of study, cognitive science, to which the two are central. As we have seen, psychology was dominated by behaviorist approaches that were largely skeptical of positing internal mental states as part of a serious, scientific psychology. And Anglo-American philosophy featured two distinct trends, each of which made philosophy more insular with respect to other disciplines, and each of which served to reinforce

the behaviorist orientation of psychology. First, ordinary language philosophy, particularly in Great Britain under the influence of Ludwig Wittgenstein and J. L. Austin, demarcated distinctly philosophical problems as soluble (or dissoluble) chiefly by reference to what one would ordinarily say, and tended to see philosophical views of the past and present as the result of confusions in how philosophers and others come to use words that generally have a clear sense in their ordinary contexts. This approach to philosophical issues in the post-war period has recently been referred to by Marjorie Grene (1995: 55) as the "Bertie Wooster season in philosophy," a characterization I suspect would seem apt to many philosophers of mind interested in contemporary cognitive science (and in P. G. Wodehouse). Let me illustrate how this approach to philosophy served to isolate the philosophy of mind from the sciences of the mind with perhaps the two most influential examples pertaining to the mind in the ordinary language tradition. In *The Concept of Mind*, Gilbert Ryle (1949: 17) attacked a view of the mind that he referred to as "Descartes' Myth" and "the dogma of the Ghost in the Machine"—basically, dualism—largely through a repeated application of the objection that dualism consisted of an extended *category mistake*: it "represents the facts of mental life as if they belonged to one logical type or category . . . when they actually belong to another." Descartes' Myth represented a category mistake because in supposing that there was a special, inner theater on which mental life is played out, it treated the "facts of mental life" as belonging to a special category of facts, when they were simply facts about how people can, do, and would behave in certain circumstances. Ryle set about showing that for the range of mental concepts that were held to refer to private, internal mental episodes or events according to Descartes' Myth—intelligence, the will, emotion, self-knowledge, sensation, and imagination—an appeal to what one would ordinarily say both shows the dogma of the Ghost in the Machine to be false, and points to a positive account of the mind that was behaviorist in orientation. To convey why Ryle's influential views here turned philosophy of mind away from science rather than towards it, consider the opening sentences of *The Concept of Mind*: "This book offers what may with reservations be described as a theory of the mind. But it does not give new information about minds. We possess already a wealth of information about minds, information which is neither derived from, nor upset by, the arguments of philosophers. The philosophical arguments which constitute this book are intended not to increase what we know about minds, but to rectify the logical geography of the knowledge which we already possess" (Ryle 1949: 9). The "we" here refers to ordinary folk, and the philosopher's task in articulating a theory of mind is to draw on what we already know about the mind, rather than on arcane, philosophical views or on specialized, scientific knowledge.

The second example is Norman Malcolm's *Dreaming*, which, like *The Concept of Mind*, framed the critique it wished to deliver as an attack on a Cartesian view of the mind. Malcolm's (1959: 4) target was the view that "dreams are the activity of the mind during sleep," and associated talk of DREAMING as involving various mental acts, such as remembering, imagining, judging, thinking, and reasoning. Malcolm argued that such dream-talk, whether it be part of commonsense reflection on dreaming (How long do dreams last?; Can you work out problems in your dreams?) or a contribution to more systematic empirical research on dreaming, was a confusion arising from the failure to attend to the proper "logic" of our ordinary talk about dreaming. Malcolm's argument proceeded by appealing to how one would *use* various expressions and sentences that contained the word "dreaming." (In looking back at Malcolm's book, it is striking that nearly every one of the eighteen short chapters

begins with a paragraph about words and what one would say with or about them.) Malcolm's central point was that there was no way to *verify* any given claim about such mental activity occurring while one was asleep, because the commonsense criteria for the application of such concepts were incompatible with saying that a person was asleep or dreaming. And because there was no way to tell whether various attributions of mental states to a sleeping person were correct, such attributions were meaningless .

These claims not only could be made without an appeal to any empirical details about dreaming or SLEEP, but implied that the whole enterprise of investigating dreaming empirically itself represented some sort of *logical* muddle. Malcolm's point became more general than one simply about dreaming (or the word "dreaming"). As he said in a preface to a later work, written after "the notion that thoughts, ideas, memories, sensations, and so on 'code into' or 'map onto' neural firing patterns in the brain" had become commonplace: "I believe that a study of our psychological concepts can show that [such] psycho-physical isomorphism is not a coherent assumption" (Malcolm 1971: x). Like Ryle's straightening of the logical geography of our knowledge of minds, Malcolm's appeal to the study of our psychological concepts could be conducted without any knowledge gleaned from psychological science (cf. Griffiths 1997: chap. 2 on the emotions).

Quite distinct from the ordinary language tradition was a second general perspective that served to make philosophical contributions to the study of the mind "distinctive" from those of science. This was logical positivism or empiricism, which developed in Europe in the 1920s and flourished in the United States through the 1930s and 1940s with the immigration to the United States of many of its leading members, including Rudolph Carnap, Hans Reichenbach, Herbert Feigl, and Carl Hempel. The logical empiricists were called "empiricists" because they held that it was via the senses and observation that we came to know about the world, deploying this empiricism with the logical techniques that had been developed by Gottlob Frege, Bertrand Russell, and Alfred Whitehead. Like empiricists in general, the logical positivists viewed the sciences as the paradigmatic repository of knowledge, and they were largely responsible for the rise of philosophy of science as a distinct subdiscipline within philosophy. As part of their reflection on science they articulated and defended the doctrine of the UNITY OF SCIENCE, the idea that the sciences are, in some sense, essentially unified, and their empiricism led them to appeal to PARSIMONY AND SIMPLICITY as grounds for both theory choice within science and for preferring theories that were ontological Scrooges. This empiricism came with a focus on *what could be verified*, and with it scepticism about traditional metaphysical notions, such as God, CAUSATION, and essences, whose instances could not be verified by an appeal to the data of sense experience. This emphasis on verification was encapsulated in the verification theory of meaning, which held that the meaning of a sentence was its method of verification, implying that sentences without any such method were *meaningless*. In psychology, this fueled skepticism about the existence of internal mental representations and states (whose existence could not be objectively verified), and offered further philosophical backing for behaviorism. In contrast to the ordinary language philosophers (many of whom would have been professionally embarrassed to have been caught knowing anything about science), the positivists held that philosophy was to be informed about and sensitive to the results of science. The distinctive task of the philosopher, however, was not simply to describe scientific practice, but to offer a *rational reconstruction* of it, one that made clear the logical structure of science. Although the term "*rational reconstruction*" was used first by

Carnap in his 1928 book *The Logical Construction of the World*, quite a general epistemological tract, the technique to which it referred came to be applied especially to scientific concepts and theories. This played out in the frequent appeal to the distinction between the *context of discovery* and the *context of justification*, drawn as such by Reichenbach in *Experience and Prediction* (1938) but with a longer history in the German tradition. To consider an aspect of a scientific view in the context of discovery was essentially to raise psychological, sociological, or historical questions about how that view originated, was developed, or came to be accepted or rejected. But properly philosophical explorations of science were to be conducted in the context of justification, raising questions and making claims about the logical structure of science and the concepts it used.

Rational reconstruction was the chief way of divorcing the relevant scientific theory from its mere context of discovery. A story involving Feigl and Carnap nicely illustrates the divorce between philosophy and science within positivism. In the late 1950s, Feigl visited the University of California, Los Angeles, to give a talk to the Department of Philosophy, of which Carnap was a member. Feigl's talk was aimed at showing that a form of physicalism, the mind-brain identity theory, faced an empirical problem, since science had little, if anything, to say about the "raw feel" of consciousness, the WHAT-IT'S-LIKE of experience. During the question period, Carnap raised his hand, and was called on by Feigl. "Your claim that current neurophysiology tells us nothing about raw feels is wrong! You have overlooked the discovery of alpha-waves in the brain," exclaimed Carnap. Feigl, who was familiar with what he thought was the relevant science, looked puzzled:"Alpha-waves? What are they?" Carnap replied: "My dear Herbert. You tell me what raw feels are, and I will tell you what alpha-waves are." Of the multiple readings that this story invites (whose common denominator is surely Carnap's savviness and wit), consider those that take Carnap's riposte to imply that he thought that one could defend materialism by, effectively, making up the science to fit whatever phenomena critics could rustle up. A rather extreme form of rational reconstruction, but it suggests one way in which the positivist approach to psychology could be just as a priori and so divorced from empirical practice as that of Ryle and Malcolm.

See also CAUSATION; DREAMING; PARSIMONY AND SIMPLICITY; SLEEP; UNITY OF SCIENCE; WHAT-IT'S-LIKE

4 The Philosophy of Science

The philosophy of science is integral to the cognitive sciences in a number of ways. We have already seen that positivists held views about the overall structure of science and the grounds for theory choice in science that had implications for psychology. Here I focus on three functions that the philosophy of science plays vis-à-vis the cognitive sciences: it provides a perspective on the place of psychology among the sciences; it raises questions about what any science can tell us about the world; and it explores the nature of knowledge and how it is known. I take these in turn.

One classic way in which the sciences were viewed as being unified, according to the positivists, was via reduction. REDUCTIONISM, in this context, is the view that intuitively "higher-level" sciences can be reduced, in some sense, to "lower-level" sciences. Thus, to begin with the case perhaps of most interest to MITECS readers, psychology was held to be reducible in principle to biology, biology to chemistry, chemistry to physics. This sort of reduction presupposed the existence of *bridge laws*,

laws that exhaustively characterized the concepts of any higher-level science, and the generalizations stated using them, in terms of those concepts and generalizations at the next level down. And because reduction was construed as relating theories of one science to those of another, the advocacy of reductionism went hand-in-hand with a view of EXPLANATION that gave lower-level sciences at least a usurpatory power over their higher-level derivatives.

This view of the structure of science was opposed to EMERGENTISM, the view that the properties studied by higher-level sciences, such as psychology, were not mere aggregates of properties studied by lower-level sciences, and thus could not be completely understood in terms of them. Both emergentism and this form of reductionism were typically cast in terms of the relationship between laws in higher- and lowerlevel sciences, thus presupposing that there were, in the psychological case, PSYCHOLOGICAL LAWS in the first place. One well-known position that denies this assumption is Donald Davidson's ANOMALOUS MONISM, which claims that while mental states *are* strictly identical with physical states, our descriptions of them as mental states are neither definitionally nor nomologically reducible to descriptions of them as physical states. This view is usually expressed as denying the possibility of the bridge laws required for the reduction of psychology to biology.

Corresponding to the emphasis on scientific laws in views of the relations between the sciences is the idea that these laws state relations between NATURAL KINDS. The idea of a natural kind is that of a type or kind of thing that exists in the world itself, rather than a kind or grouping that exists because of our ways of perceiving, thinking about, or interacting with the world. Paradigms of natural kinds are biological kinds—species, such as the domestic cat (*Felis domesticus*)—and chemical kinds—such as silver (Ag) and gold (Au). Natural kinds can be contrasted with *artifactual* kinds (such as chairs), whose members are artifacts that share common functions or purposes relative to human needs or designs; with *conventional* kinds (such as marriage vows), whose members share some sort of conventionally determined property; and from purely arbitrary groupings of objects, whose members have nothing significant in common save that they belong to the category.

Views of what natural kinds are, of how extensively science traffics in them, and of how we should characterize the notion of a natural kind vis-à-vis other metaphysic notions, such as essence, intrinsic property, and causal power, all remain topics of debate in contemporary philosophy of science (e.g., van Fraassen 1989; Wilson 1999). There is an intuitive connection between the claims that there are natural kinds, and that the sciences strive to identify them, and *scientific realism,* the view that the entities in mature sciences, whether they are observable or not, exist and our theories about them are at least approximately true. For realists hold that the sciences strive to "carve nature at its joints," and natural kinds are the pre-existing joints that one's scientific carving tries to find. The REALISM AND ANTIREALISM issue is, of course, more complicated than suggested by the view that scientific realists think there are natural kinds, and antirealists deny this—not least because there are a number of ways to deny either this realist claim or to diminish its significance. But such a perspective provides one starting point for thinking about the different views one might have of the relationship between science and reality.

Apart from raising issues concerning the relationships between psychology and other sciences and their respective objects of study, and questions about the relation between science and reality, the philosophy of science is also relevant to the cognitive sciences as a branch of epistemology or the theory of knowledge, studying a particular

type of knowledge, scientific knowledge. A central notion in the general theory of knowledge is JUSTIFICATION, because being justified in what we believe is at least one thing that distinguishes knowledge from mere belief or a lucky guess. Since scientific knowledge is a paradigm of knowledge, views of justification have often been developed with scientific knowledge in mind.

The question of what it is for an individual to have a justified belief, however, has remained contentious in the theory of knowledge. Justified beliefs are those that we are entitled to hold, ones for which we have reasons, but how should we understand such entitlement and such reasons? One dichotomy here is between *internalists* about justification, who hold that having justified belief exclusively concerns facts that are "internal" to the believer, facts about his or her internal cognitive economy; and *externalists* about justification, who deny this. A second dichotomy is between *naturalists,* who hold that what cognitive states are justified may depend on facts about cognizers or about the world beyond cognizers that are uncovered by empirical science; and *rationalists,* who hold that justification is determined by the relations between one's cognitive states that the agent herself is in a special position to know about. Clearly part of what is at issue between internalists and externalists, as well as between naturalists and rationalists, is the role of the first-person perspective in accounts of justification and thus knowledge (see also Goldman 1997).

These positions about justification raise some general questions about the relationship between EPISTEMOLOGY AND COGNITION, and interact with views of the importance of first- and third-person perspectives on cognition itself. They also suggest different views of RATIONAL AGENCY, of what it is to be an agent who acts on the basis of justified beliefs. Many traditional views of rationality imply that cognizers have LOGICAL OMNISCIENCE, that is, that they believe all the logical consequences of their beliefs. Since clearly we are not logically omniscient, there is a question of how to modify one's account of rationality to avoid this result.

5 The Mind in Cognitive Science

At the outset, I said that the relation between the mental and physical remains the central, general issue in contemporary, materialist philosophy of mind. In section 2, we saw that the behaviorist critiques of Cartesian views of the mind and behaviorism themselves introduced a dilemma that derived from the problem of the homunculus that any mental science would seem to face. And in section 3 I suggested how a vibrant skepticism about the scientific status of a distinctively psychological science and philosophy's contribution to it was sustained by two dominant philosophical perspectives.

It is time to bring these three points together as we move to explore the view of the mind that constituted the core of the developing field of cognitive science in the 1970s, what is sometimes called *classic* cognitive science, as well as its successors. If we were to pose questions central to each of these three issues—the mentalphysical relation, the problem of the homunculus, and the possibility of a genuinely cognitive science, they might be:

a. What is the relation between the mental and the physical?

b. How can psychology avoid the problem of the homunculus?

c. What makes a genuinely *mental* science possible?

Strikingly, these questions received standard answers, in the form of three "isms," from the nascent naturalistic perspective in the philosophy of mind that accompanied the rise of classic cognitive science. (The answers, so you don't have to peek ahead, are, respectively, functionalism, computationalism, and representationalism.).

The answer to (a) is FUNCTIONALISM, the view, baldly put, that mental states are functional states. Functionalists hold that what really matters to the identity of types of mental states is not what their instances are made of, but how those instances are causally arranged: what causes them, and what they, in turn, cause. Functionalism represents a view of the mental-physical relation that is compatible with materialism or physicalism because even if it is the functional or causal *role* that makes a mental state the state it is, every *occupant* of any particular role could be physical. The role-occupant distinction, introduced explicitly by Armstrong (1968) and implicitly in Lewis (1966), has been central to most formulations of functionalism.

A classic example of something that is functionally identified or individuated is *money:* it's not what it's made of (paper, gold, plastic) that makes something Money but, rather, the causal role that it plays in some broader economic system. Recognizing this fact about money is not to give up on the idea that money is material or physical. Even though material composition is not what determines whether something is money, every instance of money is material or physical: dollar bills and checks are made of paper and ink, coins are made of metal, even money that is stored solely as a string of digits in your bank account has *some* physical composition. There are at least two related reasons why functionalism *about the mind* has been an attractive view to philosophers working in the cognitive sciences.

The first is that functionalism at least appears to support the AUTONOMY OF PSYCHOLOGY, for it claims that even if, as a matter of fact, our psychological states are realized in states of our brains, their status as *psychological* states lies in their functional organization, which can be abstracted from this particular material stuff. This is a *nonreductive* view of psychology. If functionalism is true, then there will be distinctively psychological natural kinds that cross-cut the kinds that are determined by a creature's material composition. In the context of materialism, functionalism suggests that creatures with very different material organizations could not only have mental states, but have *the same kinds* of mental states. Thus functionalism makes sense of comparative psychological or neurological investigations across species.

The second is that functionalism allows for *nonbiological* forms of intelligence and mentality. That is, because it is the "form" not the "matter" that determines psychological kinds, there could be entirely artifactual creatures, such as robots or computers, with mental states, provided that they have the right functional organization. This idea has been central to traditional artificial intelligence (AI), where one ideal has been to create programs with a functional organization that not only allows them to behave in some crude way like intelligent agents but to do so in a way that instantiates at least some aspects of intelligence itself.

Both of these ideas have been criticized as part of attacks on functionalism. For example, Paul and Patricia Churchland (1981) have argued that the "autonomy" of psychology that one gains from functionalism can be a cover for the emptiness of the science itself, and Jaegwon Kim (1993) has argued against the coherence of the nonreductive forms of materialism usually taken to be implied by functionalism. Additionally, functionalism and AI are the targets of John Searle's much-discussed CHINESE ROOM ARGUMENT.

Consider (c), the question of what makes a distinctively mental science possible. Although functionalism gives one sort of answer to this in its basis for a defense of the autonomy (and so distinctness) of psychology, because there are more functional kinds than those in psychology (assuming functionalism), this answer does not explain what is distinctively *psychological* about psychology. A better answer to this question is *representationalism,* also known as the representational theory of mind.

This is the view that mental states are relations between the bearers of those states and internal mental representations. Representationalism answers (c) by viewing psychology as the science concerned with the forms these mental representations can take, the ways in which they can be manipulated, and how they interact with one another in mediating between perceptual input and behavioral output. A traditional version of representationalism, one cast in terms of Ideas, themselves often conceptualized as images, was held by the British empiricists John Locke, George Berkeley, and DAVID HUME. A form of representationalism, the LANGUAGE OF THOUGHT (LOT) hypothesis, has more recently

been articulated and defended by Jerry Fodor (1975, 1981, 1987, 1994). The LOT hypothesis is the claim that we are able to cognize in virtue of having a mental language, *mentalese*, whose symbols are combined systematically by syntactic rules to form more complex units, such as thoughts.

Because these mental symbols are intentional or representational (they are about things), the states that they compose are representational; mental states inherit their intentionality from their constituent mental representations. Fodor himself has been particularly exercised to use the language of thought hypothesis to chalk out a place for the PROPOSITIONAL ATTITUDES and our folk psychology within the developing sciences of the mind. Not all proponents of the representational theory of mind, however, agree with Fodor's view that the system of representation underlying thought is a *language*, nor with his defense of folk psychology. But even forms of representationalism that are less committal than Fodor's own provide an answer to the question of what is distinctive about psychology: psychology is not mere neuroscience because it traffics in a range of mental representations and posits internal processes that operate on these representations.

Representationalism, particularly in Fodoresque versions that see the language of thought hypothesis as forming the foundations for a defense of both cognitive psychology and our commonsense folk psychology, has been challenged within cognitive science by the rise of connectionism in psychology and NEURAL NETWORKS within computer science. Connectionist models of psychological processing might be taken as an existence proof that one does not need to assume what is sometimes called the RULES AND REPRESENTATIONS approach to understand cognitive functions: the language of thought hypothesis is no longer "the only game in town." Connectionist COGNITIVE MODELING of psychological processing, such as that of the formation of past tense (Rumelhart and McClelland 1986), face recognition (Cottrell and Metcalfe 1991), and VISUAL WORD RECOGNITION (Seidenberg and McClelland 1989), typically does not posit discrete, decomposable representations that are concatenated through the rules of some language of thought. Rather, connectionists posit a COGNITIVE ARCHITECTURE made up of simple neuron-like nodes, with activity being propagated across the units proportional to the weights of the connection strength between them. Knowledge lies not in the nodes themselves but in the values of the weights connecting nodes. There seems to be nothing of a propositional form within such connectionist networks, no place for the internal sentences that are the objects of folk psychological states and other subpersonal psychological states posited in accounts of (for example) memory and reasoning.

The tempting idea that "classicists" accept, and connectionists reject, representationalism is too simple, one whose implausibility is revealed once one shifts one's focus from folk psychology and the propositional attitudes to cognition more generally. Even when research in classical cognitive science—for example, that on KNOWLEDGE-BASED SYSTEMS and on BAYESIAN NETWORKS—is cast in terms of "beliefs" that a system has, the connection between "beliefs" and the beliefs of folk psychology has been underexplored. More importantly, the notion of representation itself has not been abandoned across-the-board by connectionists, some of whom have sought to salvage and adapt the notion of mental representation, as suggested by the continuing debate over DISTRIBUTED VS. LOCAL REPRESENTATION and the exploration of sub-symbolic forms of representation within connectionism (see Boden 1990; Haugeland 1997; Smolensky 1994).

What perhaps better distinguishes classic and connectionist cognitive science here is not the issue of whether some form of representationalism is true, but whether the question to which it is an answer needs answering at all. In classical cognitive science, what makes the idea of a genuinely *mental* science possible is the idea that psychology describes representation crunching. But in starting with the idea that neural representation occurs from single neurons up through circuits to modules and more nebulous, distributed neural systems, connectionists are less likely to think that psychology offers a distinctive level of explanation that deserves some identifying characterization.

This rejection of question (c) is clearest, I think, in related DYNAMIC APPROACHES TO COGNITION, since such approaches investigate psychological states as dynamic systems that

need not posit distinctly *mental* representations. (As with connectionist theorizing about cognition, dynamic approaches encompass a variety of views of mental representation and its place in the study of the mind that make representationalism itself a live issue within such approaches; see Haugeland 1991; van Gelder 1998.)

Finally, consider (b), the question of how to avoid the problem of the homunculus in the sciences of the mind. In classic cognitive science, the answer to (b) is *computationalism* the view that mental states are computational, an answer which integrates and strengthens functionalist materialism and representationalism as answers to our previous two questions. It does so in the *way* in which it provides a more precise characterization of the nature of the functional or causal relations that exist between mental states: these are *computational relations between mental representations*. The traditional way to spell this out is the COMPUTATIONAL THEORY OF MIND, according to which the mind is a digital computer, a device that stores symbolic representations and performs operations on them in accord with *syntactic* rules, rules that attend only to the "form" of these symbols. This view of computationalism has been challenged not only by relatively technical objections (such as that based on the FRAME PROBLEM), but also by the development of neural networks and models of SITUATED COGNITION AND LEARNING, where (at least some) informational load is shifted from internal codes to organism-environment interactions (cf. Ballard et al. 1997).

The computational theory of mind avoids the problem of the homunculus because digital computers that exhibit some intelligence exist, and they do not contain undischarged homunculi. Thus, if *we* are fancy versions of such computers, then we can understand our intelligent capacities without positing undischarged homunculi. The way this works in computers is by having a series of programs and languages, each compiled by the one beneath it, with the most basic language directly implemented in the hardware of the machine. We avoid an endless series of homunculi because the capacities that are posited at any given level are typically simpler and more numerous than those posited at any higher level, with the lowest levels specifying instructions to perform actions that require no intelligence at all. This strategy of FUNCTIONAL DECOMPOSITION solves the problem of the homunculus if we are digital computers, assuming that it solves it for digital computers.

Like representationalism, computationalism has sometimes been thought to have been superseded by either (or both) the connectionist revolution of the 1980s, or the Decade of the Brain (the 1990s). But as with proclamations of the death of representationalism this notice of the death of computationalism is premature. In part this is because the object of criticism is a specific version of computationalism, not computationalism per se (cf. representationalism), and in part it is because neural networks and the neural systems in the head they model are both themselves typically claimed to be computational in some sense. It is surprisingly difficult to find an answer within the cognitive science community to the question of whether there is a univocal notion of COMPUTATION that underlies the various different computational approaches to cognition on offer. The various types of AUTOMATA postulated in the 1930s and 1940s—particularly TURING machines and the "neurons" of MCCULLOCH and PITTS, which form the intellectual foundations, respectively, for the computational theory of mind and contemporary neural network theory—have an interwoven history, and many of the initial putative differences between classical and connectionist cognitive science have faded into the background as research in artificial intelligence and cognitive modeling has increasingly melded the insights of each approach into more sophisticated hybrid models of cognition (cf. Ballard 1997).

While dynamicists (e.g., Port and van Gelder 1995) have sometimes been touted as providing a noncomputational alternative to both classic and connectionist cognitive science (e.g., Thelen 1995: 70), as with claims about the nonrepresentational stance of such approaches, such a characterization is not well founded (see Clark 1997, 1998). More generally, the relationship between dynamical approaches to both classical and connectionist views remains a topic for further discussion (cf. van Gelder and Port 1995; Horgan and Tienson 1996; and Giunti 1997).

6 A Focus on Folk Psychology

Much recent philosophical thinking about the mind and cognitive science remains preoccupied with the three traditional philosophical issues I identified in the first section: the mental-physical relation, the structure of the mind, and the first-person perspective. All three issues arise in one of the most absorbing discussions over the last twenty years, that over the nature, status, and future of what has been variously called commonsense psychology, the propositional attitudes, or FOLK PSYCHOLOGY.

The term *folk psychology* was coined by Daniel Dennett (1981) to refer to the systematic knowledge that we "folk" employ in explaining one another's thoughts, feelings, and behavior; the idea goes back to Sellars's Myth of Jones in "Empiricism and the Philosophy of Mind" (1956). We all naturally and without explicit instruction engage in psychological explanation by attributing beliefs, desires, hopes, thoughts, memories, and emotions to one another. These patterns of folk psychological explanation are "folk" as opposed to "scientific" since they require no special training and are manifest in everyday predictive and explanatory practice; and genuinely "psychological" because they posit the existence of various states or properties that seem to be paradigmatically mental in nature. To engage in folk psychological explanation is, in Dennett's (1987) terms, to adopt the INTENTIONAL STANCE.

Perhaps the central issue about folk psychology concerns its relationship to the developing cognitive sciences. ELIMINATIVE MATERIALISM, or eliminativism, is the view that folk psychology will find no place in any of the sciences that could be called "cognitive" in orientation; rather, the fortune of folk psychology will be like that of many other folk views of the world that have found themselves permanently out of step with scientific approaches to the phenomena they purport to explain, such as folk views of medicine, disease, and witchcraft.

Eliminativism is sometimes motivated by adherence to reductionism (including the thesis of EXTENSIONALITY) and the ideal of the unity of science, together with the recognition that the propositional attitudes have features that set them off in kind from the types of entity that exist in other sciences. For example, they are intentional or representational, and attributing them to individuals seems to depend on factors beyond the boundary of those individuals, as the TWIN EARTH arguments suggest. These arguments and others point to a prima facie conflict between folk psychology and INDIVIDUALISM (or *internalism*) in psychology (see Wilson 1995). The apparent conflict between folk psychology and individualism has provided one of the motivations for developing accounts of NARROW CONTENT, content that depends solely on an individual's intrinsic, physical properties. (The dependence here has usually been understood in terms of the technical notion of SUPERVENIENCE; see Horgan 1993.)

There is a spin on this general motivation for eliminative materialism that appeals more directly to the issue of the how the mind is structured. The claim here is that whether folk psychology is defensible will turn in large part on how compatible its ontology—its list of what we find in a folk psychological mind—is with the developing ontology of the cognitive sciences. With respect to classical cognitive science, with its endorsement of both the representational and computational theories of mind, folk psychology is on relatively solid ground here. It posits representational states, such as belief and desire, and it is relatively easy to see how the causal relations between such states could be modeled computationally. But connectionist

models of the mind, with what representation there is lying in patterns of activity rather than in explicit representations like propositions, seem to leave less room in the structure of the mind for folk psychology.

Finally, the issue of the place of the first-person perspective arises with respect to folk psychology when we ask how people deploy folk psychology. That is, what sort of psychological machinery do we folk employ in engaging in folk psychological explanation? This issue has been the topic of the SIMULATION VS. THEORY-THEORY debate, with proponents of the simulation view holding, roughly, a "first-person first" account of how folk psychology works, and theory-theory proponents viewing folk psychology as essentially a third-person predictive and explanatory tool. Two recent volumes by Davies and Stone (1995a, 1995b) have added to the literature on this debate, which has developmental and moral aspects, including implications for MORAL PSYCHOLOGY.

7 Exploring Mental Content

Although BRENTANO's claim that INTENTIONALITY is the "mark of the mental" i problematic and has few adherents today, intentionality has been one of the flagship topics in philosophical discussion of the mental, and so at least a sort of mark of that discussion. Just what the puzzle about intentionality is and what one might say about it are topics I want to explore in more detail here. To say that something is intentional is just to say that it is *about something,* or that it *refers to something*. In this sense, tatements of fact are paradigmatically intentional, since they are about how things are in the world. Similarly, a highway sign with a picture of a gas pump on it is intentional because it conveys the information that there is gas station ahead at an exit: it is, in some sense, about that state of affairs.

The beginning of chapter 4 of Jerry Fodor's *Psychosemantics* provides one lively expression of the problem with intentionality: I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin, charm,* and *charge* will perhaps appear upon their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see, in face of this consideration, how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or maybe of their supervenience on?) properties that are themselves *neither* intentional *nor* semantic. If aboutness is real, it must be really something else. (p. 97, emphases in original) Although there is much that one could take issue with in this passage, my reason for introducing it here is not to critique it but to try to capture some of the worries about intentionality that bubble up from it. The most general of these concerns the *basis* of intentionality in the natural order: given that only special parts of the world (like our minds) have intentional properties, what is it about those things that gives them (and not other things) intentionality? Since not only mental phenomena are intentional (for example, spoken and written natural language and systems of signs and codes are as well), one might think that a natural way to approach this question would be as follows. Consider all of the various sorts of "merely material" things that at least seem to have intentional properties. Then proceed to articulate why each of them is intentional, either taking the high road of specifying something like the "essence of intentionality"—something that all and only things with intentional properties have—or taking the low road of doing so for

each phenomenon, allowing these accounts to vary across disparate intentional phenomena.

Very few philosophers have explored the problem of intentionality in this way. I think this is chiefly because they do not view all things with intentional properties as having been created equally. A common assumption is that even if lots of the nonmental world is intentional, its intentionality is *derived,* in some sense, from the intentionality of the mental. So, to take a classic example, the sentences we utter and write are intentional all right (they are about things). But their intentionality derives from that of the corresponding thoughts that are their causal antecedents. To take another oftentouted example, computers often produce intentional output (even photocopiers can do this), but whatever intentionality lies in such output is not inherent to the machines that produce it but is derivative, ultimately, from the mental states of those who design, program, and use them and their products. Thus, there has been a focus on mental states as a sort of paradigm of intentional state, and a subsequent narrowing of the sorts of intentional phenomena discussed. Two points are perhaps worth making briefly in this regard.

First, the assumption that not all things with intentional properties are created equally is typically shared even by those who have not focused almost exclusively on mental states as paradigms of intentional states, but on languages and other public and conventional forms of representation (e.g., Horst 1996). It is just that their paradigm is different. Second, even when mental states *have* been taken as a paradigm here, those interested in developing a "psychosemantics"—an account of the basis for the semantics of psychological states—have often turned to decidedly nonmental systems of representation in order to theorize about the intentionality of the mental. This focus on what we might think of as *proto-intentionality* has been prominent within both Fred Dretske's (1981) informational semantics and the biosemantic approach pioneered by Ruth Millikan (1984, 1993).

The idea common to such views is to get clear about the grounds of simple forms of intentionality before scaling up to the case of the intentionality of human minds, an instance of a research strategy that has driven work in the cognitive sciences from early work in artificial intelligence on KNOWLEDGE REPRESENTATION and cognitive modeling through to contemporary work in COMPUTATIONAL NEUROSCIENCE. Exploring simplified or more basic intentional systems in the hope of gaining some insight into the more full-blown case of the intentionality of human minds runs the risk, of course, of focusing on cases that leave out precisely that which is crucial to full-blown intentionality. Some (for example, Searle 1992) would claim that consciousness and phenomenology are such features.

As I hinted at in my discussion of the mind in cognitive science in section 5, construed one way the puzzle about the grounds of intentionality has a general answer in the hypothesis of computationalism. But there is a deeper problem about the grounds of intentionality concerning *just how* at least some mental stuff could be about other stuff in the world, and computationalism is of little help here. Computationalism does not even pretend to answer the question of what it is about specific mental states (say, my belief that trees often have leaves) that gives them the content that they have—for example, that makes them *about trees*. Even if we *were* complicated Turing machines, what would it be about *my* Turing machine table that implies that I have the belief that trees often have leaves? Talking about the correspondence between the semantic and syntactic properties that symbol structures in computational systems have, and of how the former are "inherited" from the latter is well and good. But it leaves open the "just how" question, and so fails to address

what I am here calling the deeper problem about the grounds of intentionality. This problem is explored in the article on MENTAL REPRESENTATION, and particular proposals for a psychosemantics can be found in those on INFORMATIONAL SEMANTICS and FUNCTIONAL ROLE SEMANTICS.

It would be remiss in exploring mental content to fail to mention that much thought about intentionality has been propelled by work in the philosophy of language: on INDEXICALS AND DEMONSTRATIVES, on theories of REFERENCE and the propositional attitudes, and on the idea of RADICAL INTERPRETATION. Here I will restrict myself to some brief comments on theories of reference, which have occupied center stage in the philosophy of language for much of the last thirty years.

One of the central goals of theories of reference has been to explain in virtue of what parts of sentences of natural languages refer to the things they refer to. What makes the name "Miranda" refer to my daughter? In virtue of what does the plural noun "dogs" refer to dogs? Such questions have a striking similarity to my above expression of the central puzzle concerning intentionality. In fact, the application of causal theories of reference (Putnam 1975, Kripke 1980) developed principally for natural languages has played a central role in disputes in the philosophy of mind that concern intentionality, including those over individualism, narrow content, and the role of Twin Earth arguments in thinking about intentionality. In particular, applying them not to the meaning of natural language terms but to the content of thought is one way to reach the conclusion that *mental* content does not supervene on an individual's physical properties, that is, that mental content is not individualistic.

GOTTLOB FREGE is a classic source for contrasting descriptivist theories of reference, according to which natural language reference is, in some sense, mediated by a speaker's descriptions of the object or property to which she refers. Moreover, Frege's notion of sense and the distinction between SENSE AND REFERENCE are often invoked in support of the claim that there is much to MEANING—linguistic or mental—that goes beyond the merely referential. Frege is also one of the founders of modern logic, and it is to the role of logic in the cognitive sciences that I now turn.

8 Logic and the Sciences of the Mind

Although INDUCTION, like deduction, involves drawing inferences on the basis of one or more premises, it is *deductive* inference that has been the focus in LOGIC, what is often simply referred to as "formal logic" in departments of philosophy and linguistics. The idea that it is possible to abstract away from deductive arguments given in natural language that differ in the content of their premises and conclusions goes back at least to Aristotle in the fourth century B.C. Hence the term "Aristotelian syllogisms" to refer to a range of argument forms containing premises and conclusions that begin with the words "every" or "all," "some," and "no." This abstraction makes it possible to talk about argument *forms* that are valid and invalid, and allows one to describe two arguments as being of the same *logical* form. To take a simple example, we know that any argument of the form:

All A are B.
No B are C.

No A are C.

is *formally* valid, where the emphasis here serves to highlight reference to the preservation of truth from premises to conclusion, that is, the validity, solely in virtue of the forms of the individual sentences, together with the form their arrangement constitutes.

Whatever plural noun phrases we substitute for "A," "B," and "C," the resulting natural language argument will be valid: if the two premises are true, the conclusion must also be true. The same general point applies to arguments that are formally *invalid,* which makes it possible to talk about formal *fallacies,* that is, inferences that are invalid because of the forms they instantiate. Given the age of the general idea of LOGICAL FORM, what is perhaps surprising is that it is only in the late nineteenth century that the notion was developed so as to apply to a wide range of natural language constructions through the development of the *propositional* and *predicate* logics. And it is only in the late twentieth century that the notion of logical form comes to be appropriated within linguistics in the study of SYNTAX. I focus here on the developments in logic.

Central to propositional logic (sometimes called "sentential logic") is the idea of a propositional or sentential *operator,* a symbol that acts as a function on propositions or sentences. The paradigmatic propositional operators are symbols for negation ("$\neg$"), conjunction ("$\wedge$"), disjunction ("$\vee$"), and conditional ("$\Rightarrow$"). And with the development of formal languages containing these symbols comes an ability to represent a richer range of formally valid arguments, such as that manifest in the following thought:

If Sally invites Tom, then either he will say "no," or cancel his game with Bill. But there's no way he'd turn Sally down. So I guess if she invites him, he'll cancel with Bill. In predicate or quantificational logic, we are able to represent not simply the relations between propositions, as we can in propositional logic, but also the structure within propositions themselves through the introduction of QUANTIFIERS and the terms and predicates that they bind. One of the historically more important applications of predicate logic has been its widespread use in linguistics, philosophical logic, and the philosophy of language to formally represent increasingly larger parts of natural languages, including not just simple subjects and predicates, but adverbial constructions, tense, indexicals, and attributive adjectives (for example, see Sainsbury 1991).

These fundamental developments in logical theory have had perhaps the most widespread and pervasive effect on the foundations of the cognitive sciences of *any* contributions from philosophy or mathematics. They also form the basis for much contemporary work across the cognitive sciences: in linguistic semantics (e.g., through MODAL LOGIC, in the use of POSSIBLE WORLDS SEMANTICS to model fragments of natural language, and in work on BINDING); in metalogic (e.g., on FORMAL SYSTEMS and results such as the CHURCH-TURING THESIS and GÖDEL'S THEOREMS); and in artificial intelligence (e.g., on LOGICAL REASONING SYSTEMS, TEMPORAL REASONING, and METAREASONING).

Despite their technical payoff, the relevance of these developments in logical theory for thinking more directly about DEDUCTIVE REASONING in human beings is, ironically, less clear. Psychological work on human reasoning, including that on JUDGMENT HEURISTICS, CAUSAL REASONING, and MENTAL MODELS, points to ways in which human reasoning may be governed by structures very different from those developed in formal logic, though this remains an area of continuing debate and discussion

9 Two Ways to Get Biological

By the late nineteenth century, both evolutionary theory and the physiological study of mental capacities were firmly entrenched. Despite this, these two paths to a biological view of cognition have only recently been re-explored in sufficient depth to warrant the claim that contemporary cognitive science incorporates a truly biological perspective on the mind. The neurobiological path, laid down by the tradition of physiological psychology that developed from the mid-nineteenth century, is certainly the better traveled of the two. The recent widening of this path by those dissatisfied with the distinctly nonbiological approaches adopted within traditional artificial intelligence has, as we saw in our discussion of computationalism, raised new questions about COMPUTATION AND THE BRAIN, the traditional computational theory of the mind, and the rules and representations approach to understanding the mind. The evolutionary path, by contrast, has been taken only occasionally and half-heartedly over the last 140 years. I want to concentrate not only on why but on the ways in which evolutionary theory is relevant to contemporary interdisciplinary work on the mind.

The theory of EVOLUTION makes a claim about the *patterns* that we find in the biological world—they are patterns of *descent*—and a claim about the predominant cause of those patterns—they are caused by the mechanism of natural selection. None of the recent debates concerning evolutionary theory—from challenges to the focus on ADAPTATION AND ADAPTATIONISM in Gould and Lewontin (1979) to more recent work on SELF-ORGANIZING SYSTEMS and ARTIFICIAL LIFE—challenges the substantial core of the theory of evolution (cf. Kauffman 1993, 1995; Depew and Weber 1995).

The vast majority of those working in the cognitive sciences both accept the theory of evolution and so think that a large number of traits that organisms possess are adaptations to evolutionary forces, such as natural selection. Yet until the last ten years, the scattered pleas to apply evolutionary theory to the mind (such as those of Ghiselin 1969 and Richards 1987) have come largely from those outside of the psychological and behavioral sciences.

Within the last ten years, however, a distinctive EVOLUTIONARY PSYCHOLOGY has developed as a research program, beginning in Leda Cosmides's (1989) work on human reasoning and the Wason selection task, and represented in the collection of papers *The Adapted Mind* (Barkow, Cosmides, and Tooby 1992) and, more recently and at a more popular level, by Steven Pinker's *How the Mind Works* (1997). Evolutionary psychologists view the mind as a set of "Darwinian algorithms" designed by natural selection to solve adaptive problems faced by our hunter-gatherer ancestors.

The claim is that this basic Darwinian insight can and should guide research into the cognitive architecture of the mind, since the task is one of discovering and understanding the *design* of the human mind, in all its complexity. Yet there has been more than an inertial resistance to viewing evolution as central to the scientific study of human cognition. One reason is that evolutionary theory in general is seen as answering different questions than those at the core of the cognitive sciences. In terms of the well-known distinction between *proximal* and *ultimate* causes, appeals to evolutionary theory primarily allow one to specify the latter, and cognitive scientists are chiefly interested in the former: they are interested in the *how* rather than the *why* of the mind. Or to put it more precisely, central to cognitive science is an

understanding of the *mechanisms* that govern cognition, not the various histories—evolutionary or not—that produced these mechanisms. This general perception of the concerns of evolutionary theory and the contrasting conception of cognitive science, have both been challenged by evolutionary psychologists. The same general challenges have been issued by those who think that the relations between ETHICS AND EVOLUTION and those between cognition and CULTURAL EVOLUTION have not received their due in contemporary cognitive science.

Yet despite the skepticism about this direct application of evolutionary theory to human cognition, its implicit application is inherent in the traditional interest in the minds of *other* animals, from *aplysia* to (nonhuman) apes. ANIMAL NAVIGATION, PRIMATE LANGUAGE, and CONDITIONING AND THE BRAIN, while certainly topics of interest in their own right, gain some added value from what their investigation can tell us about *human* minds and brains. This presupposes something like the following: that there are natural kinds in psychology that transcend species boundaries, such that there is a general way of exploring how a cognitive capacity is structured, independent of the particular species of organism in which it is instantiated (cf. functionalism).

Largely on the basis of research with non-human animals, we know enough now to say, with a high degree of certainty, things like this: that the CEREBELLUM is the central brain structure involved in MOTOR LEARNING, and that the LIMBIC SYSTEM plays the same role with respect to at least some EMOTIONS. This is by way of returning to (and concluding with) the neuroscientific path to biologizing the mind, and the three classic philosophical issues about the mind with which we began. As I hope this introduction has suggested, despite the distinctively philosophical edge to all three issues—the mental-physical relation, the structure of the mind, and the first-person perspective—discussion of each of them is elucidated and enriched by the interdisciplinary perspectives provided by empirical work in the cognitive sciences. It is not only a priori arguments but complexities revealed by empirical work (e.g., on the neurobiology of consciousness, or ATTENTION and animal and human brains) that show the paucity of the traditional philosophical "isms" (dualism, behaviorism, type-type physicalism) with respect to the mental-physical relation. It is not simply general, philosophical arguments against nativism or against empiricism about the structure of the mind that reveal limitations to the global versions of these views, but ongoing work on MODULARITY AND LANGUAGE, on cognitive architecture, and on the innateness of language. And thought about introspection and self-knowledge, to take two topics that arise when one reflects on the first-person perspective on the mind, is both enriched by and contributes to empirical work on BLINDSIGHT, the theory of mind, and METAREPRESENTATION. With some luck, philosophers increasingly sensitive to empirical data about the mind will have paved a two-way street that encourages psychologists, linguists, neuroscientists, computer scientists, social scientists and evolutionary theorists to venture more frequently and more surely into philosophy.

References

Armstrong, D. M. (1968). *A Materialist Theory of the Mind.* London: Routledge and Kegan Paul.

Ballard, D. (1997). *An Introduction to Natural Computation.* Cambridge, MA: MIT Press.

Ballard, D., M. Hayhoe, P. Pook, and R. Rao. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences* 20: 723–767.

Barkow, J. H., L. Cosmides, and J. Tooby, Eds. (1992). *The Adapted Mind.* New York: Oxford University Press.

Boden, M., Ed. (1990). *The Philosophy of Artificial Intelligence.* Oxford: Oxford University Press.

Carnap, R. (1928). *The Logical Construction of the World.* Translated by R. George (1967). Berkeley: University of California Press.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory.* New York: Oxford University Press.

Chomsky, N. (1959). Review of B. F. Skinner's *Verbal Behavior. Language* 35 : 26–58.

Churchland, P. M. (1979). *Scientific Realism and the Plasticity of Mind.* New York: Cambridge University Press.

Churchland, P. M., and P. S. Churchland. (1981). Functionalism, qualia, and intentionality. *Philosophical Topics* 12: 121–145.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again.* Cambridge, MA: MIT Press.

Clark, A. (1998). Twisted tales: Causal complexity and cognitive scientific explanation. *Minds and Machines* 8: 79–99.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason Selection Task. *Cognition* 31: 187–276.

Cottrell, G., and J. Metcalfe. (1991). EMPATH: Face, Emotion, and Gender Recognition Using Holons. In R. Lippman, J. Moody, and D. Touretzky, Eds., *Advances in Neural Information Processing Systems*, vol. 3. San Mateo, CA: Morgan Kaufmann.

Davies, M., and T. Stone, Eds. (1995a). *Folk Psychology: The Theory of Mind Debate.* Oxford: Blackwell.

Davies, M., and T. Stone, Eds. (1995b). *Mental Simulation: Evaluations and Applications.* Oxford: Blackwell.

Dennett, D. C. (1981). Three kinds of intentional psychology. Reprinted in his 1987.

Dennett, D. C. (1987). *The Intentional Stance.* Cambridge, MA: MIT Press.

Depew, D., and B. Weber. (1995). *Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection.* Cambridge, MA: MIT Press.

Dretske, F. (1981). *Knowledge and the Flow of Information.* Cambridge, MA: MIT Press.

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

Fodor, J. A. (1975). *The Language of Thought.* Cambridge, MA: Harvard University Press.

Fodor, J. A. (1981). *Representations: Philosophical Essays on the Foundations of Cognitive Science.* Sussex: Harvester Press.

Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.* Cambridge, MA: MIT Press.

Fodor, J. A. (1994). *The Elm and the Expert.* Cambridge, MA: MIT Press.

Geach, P. (1957). *Mental Acts.* London: Routledge and Kegan Paul.

Ghiselin, M. (1969). *The Triumph of the Darwinian Method.* Berkeley: University of California Press.

Giunti, M. (1997). *Computation, Dynamics, and Cognition.* New York: Oxford University Press.

Goldman, A. (1997). Science, Publicity, and Consciousness. *Philosophy of Science* 64: 525–545.

Gould, S. J., and R. C. Lewontin. (1979). The spandrels of San Marco and the panglossian paradigm: A critique of the adaptationist programme. Reprinted in E. Sober, Ed., *Conceptual Issues in Evolutionary Biology*, 2nd ed. (1993.) Cambridge, MA: MIT Press.

Grene, M. (1995). *A Philosophical Testament.* Chicago: Open Court.

Griffiths, P. E. (1997). *What Emotions Really Are.* Chicago: University of Chicago Press.

Haugeland, J. (1991). Representational genera. In W. Ramsey and S. Stich, Eds., *Philosophy and Connectionist Theory.* Hillsdale, NJ: Erlbaum.

Haugeland, J., Ed. (1997). *Mind Design 2: Philosophy, Psychology, and Artificial Intelligence.* Cambridge, MA: MIT Press.

Heil, J., and A. Mele, Eds. (1993). *Mental Causation.* Oxford: Clarendon Press. Horgan, T. (1993). From supervenience to superdupervenience: Meeting the demands of a material world. *Mind* 102: 555–586.

Horgan, T., and J. Tienson. (1996). *Connectionism and the Philosophy of Psychology.* Cambridge, MA: MIT Press.

Horst, S. (1996). *Symbols, Computation, and Intentionality.* Berkeley: University of California Press.

James, W. (1890). *The Principles of Psychology.* 2 vol. Dover reprint (1950). New York: Dover.

Kauffman, S. (1993). *The Origins of Order.* New York: Oxford University Press.

Kauffman, S. (1995). *At Home in the Universe.* New York: Oxford University Press.

Kim, J. (1993). *Supervenience and Mind.* New York: Cambridge University Press.

Kosslyn, S. (1980). *Image and Mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. (1994). *Image and Brain.* Cambridge, MA: MIT Press.

Kripke, S. (1980). *Naming and Necessity.* Cambridge, MA: Harvard University Press.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64: 354–361.

Lewis, D. K. (1966). An argument for the identity theory. *Journal of Philosophy* 63: 17–25.

Malcolm, N. (1959). *Dreaming.* London: Routledge and Kegan Paul.

Malcolm, N. (1971). *Problems of Mind: Descartes to Wittgenstein.* New York: Harper and Row.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories.* Cambridge, MA: MIT Press.

Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice.* Cambridge, MA: MIT Press.

Pinker, S. (1997). *How the Mind Works.* New York: Norton.

Port, R., and T. van Gelder, Eds. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition.* Cambridge, MA: MIT Press.

Putnam, H. (1975). The meaning of "meaning." Reprinted in *Mind, Language, and Reality: Collected Papers,* vol. 2. Cambridge: Cambridge University Press.

Pylyshyn, Z. (1984). *Computation and Cognition.* Cambridge, MA: MIT Press.

Reichenbach, H. (1938). *Experience and Prediction.* Chicago: University of Chicago Press.

Richards, R. (1987). *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior.* Chicago: University of Chicago Press.

Rumelhart, D., and J. McClelland. (1986). On Learning the Past Tenses of English Verbs. In J. McClelland, D. Rumelhart, and the PDP Research Group, Eds., *Parallel Distributed Processing,* vol. 2. Cambridge, MA: MIT Press.

Ryle, G. (1949). *The Concept of Mind.* New York: Penguin.

Sainsbury, M. (1991). *Logical Forms.* New York: Blackwell.

Searle, J. (1992). *The Rediscovery of the Mind.* Cambridge, MA: MIT Press.

Seidenberg, M. S., and J. L. McClelland. (1989). A distributed, developmental model of visual word recognition and naming. *Psychological Review* 96: 523–568.

Sellars, W. (1956). Empiricism and the philosophy of mind. In H. Feigl and M. Scriven, Eds., *Minnesota Studies in the Philosophy of Science,* vol. 1. Minneapolis: University of Minnesota Press.

Skinner, B. F. (1957). *Verbal Behavior.* New York: Appleton-Century-Crofts.

Smolensky, P. (1994). Computational models of mind. In S. Guttenplan, Ed., *A Companion to the Philosophy of Mind.* Cambridge, MA: Blackwell.

Spelke, E. (1990). Principles of object perception. *Cognitive Science* 14: 29–56.

Stabler, E. (1983). How are grammars represented? *Behavioral and Brain Sciences* 6: 391–420.

Thelen, E. (1995). Time-scale dynamics and the development of an embodied cognition. In R. Port and T. van Gelder, Eds., *Mind as Motion: Explorations in the Dynamics of Cognition.* Cambridge, MA: MIT Press.

van Fraassen, B. (1989). *Laws and Symmetry*. New York: Oxford University Press.

van Gelder, T. J. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences* 21: 1–14.

van Gelder, T., and R. Port. (1995). It's about time: An overview of the dynamical approach to cognition. In R. Port and T. van Gelder, Eds., *Mind as Motion: Explorations in the Dynamics of Cognition.* Cambridge, MA: MIT Press.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review* 20: 158–177.

Wilson, R. A. (1995). *Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind.* New York: Cambridge University Press.

Wilson, R. A., Ed. (1999). *Species: New Interdisciplinary Essays.* Cambridge, MA: MIT Press.

Wundt, W. (1900–1909). *Völkerpsychologie*. Leipzig: W. Engelmann.

Yablo, S. (1992). Mental causation. *Philosophical Review* 101: 245–280.

Psychology

Keith J. Holyoak

Psychology is the science that investigates the representation and processing of information by complex organisms. Many animal species are capable of taking in information about their environment, forming internal representations of it, and manipulating these representations to select and execute actions. In addition, many animals are able to adapt to their environments by means of learning that can take place within the lifespan of an individual organism. Intelligent information processing implies the ability to acquire and process information about the environment in order to select actions that are likely to achieve the fundamental goals of survival and propagation. Animals have evolved a system of capabilities that collectively provide them with the ability to process information. They have sensory systems such as TASTE and HAPTIC PERCEPTION (touch), which provide information about the immediate environment with which the individual is in direct contact; proprioception, which provides information about an animal's own bodily states; and SMELL, AUDITION, and VISION, which provide information about more distant aspects of the environment. Animals are capable of directed, self-generated motion, including EYE MOVEMENTS and other motoric behaviors such as MANIPULATION AND GRASPING, which radically increase their ability to pick up sensory information and also to act upon their environments.

The central focus of psychology concerns the information processing that intervenes between sensory inputs and motoric outputs. The most complex forms of intelligence, observed in birds and mammals, and particularly primates (especially great apes and humans) require theories that deal with the machinery of thought and inner experience. These animals have minds and EMOTIONS; their sensory inputs are interpreted to create perceptions of the external world, guided in part by selective ATTENTION; some of the products of perception are stored in MEMORY, and may in turn influence subsequent perception. Intellectually sophisticated animals perform DECISION MAKING and PROBLEM SOLVING, and in the case of humans engage in LANGUAGE AND COMMUNICATION. Experience coupled with innate constraints results in a process of COGNITIVE DEVELOPMENT as the infant becomes an adult, and also leads to LEARNING over the lifespan, so that the individual is able to adapt to its environment within a vastly shorter time scale than that required for evolutionary change. Humans are capable of the most complex and most domain-general forms of information processing of all species; for this reason (and because those who study psychology are humans), most of psychology aims directly or indirectly to understand the nature of human information processing and INTELLIGENCE. The most general characteristics of the human system for information processing are described as the COGNITIVE ARCHITECTURE.

1 The Place of Psychology within Cognitive Science

As the science of the representation and processing of information by organisms, psychology (particularly cognitive psychology) forms part of the core of cognitive science. Cognitive science research conducted in other disciplines generally has actual or potential implications for psychology. Not all research on intelligent information processing is relevant to psychology. Some work in artificial intelligence, for example, is based on representations and algorithms with no apparent connection to biological intelligence. Even though such work may be highly successful at achieving high levels of competence on cognitive tasks, it does not fall within the scope of cognitive science. For example, the Deep Blue II program that defeated the xl Psychology human CHESS champion Gary Kasparov is an example of an outstanding artificialintelligence program that has little or no apparent psychological relevance, and hence would not be considered to be part of cognitive science. In contrast, work on adaptive PRODUCTION SYSTEMS and NEURAL NETWORKS, much of which is conducted by computer scientists, often has implications for psychology. Similarly, a great deal of work in such allied disciplines as neuroscience, linguistics, anthropology, and philosophy has psychological implications. At the same time, work in psychology often has important implications for research in other disciplines. For example, research in PSYCHOLINGUISTICS has influenced developments in linguistics, and research in PSYCHOPHYSICS has guided neurophysiological research on the substrates of sensation and perception.

In terms of MARR's tripartite division of levels of analysis (computational theory, representation and algorithm, and hardware implementation), work in psychology tends to concentrate on the middle level, emphasizing how information is represented and processed by humans and other animals. Although there are many important exceptions, psychologists generally aim to develop process models that specify more than the input-output functions that govern cognition (for example, also specifying timing relations among intervening mental processes), while abstracting away from the detailed neural underpinnings of behavior. Nonetheless, most psychologists do not insist in any strict sense on the AUTONOMY OF PSYCHOLOGY, but rather focus on important interconnections with allied disciplines that comprise cognitive science.

Contemporary psychology at the information-processing level is influenced by research in neuroscience that investigates the neural basis for cognition and emotion, by work on representations and algorithms in the fields of artificial intelligence and neural networks, and by work in social sciences such as anthropology that places the psychology of individuals within its cultural context. Research on the psychology of language (e.g., COMPUTATIONAL PSYCHOLINGUISTICS and LANGUAGE AND THOUGHT) is influenced by the formal analyses of language developed in linguistics. Many areas of psychology make close contact with classical issues in philosophy, especially in EPISTEMOLOGY (e.g., CAUSAL REASONING; INDUCTION; CONCEPTS).

The field of psychology has several major subdivisions, which have varying degrees of connection to cognitive science. Cognitive psychology deals directly with the representation and processing of information, with greatest emphasis on cognition in adult humans; the majority of the psychology entries that appear in this volume reflect work in this area. Developmental psychology deals with the changes in cognitive, social, and emotional functioning that occur over the lifespan of humans and other animals (see

in particular COGNITIVE DEVELOPMENT, PERCEPTUAL DEVELOPMENT, and INFANT COGNITION). Social psychology investigates the cognitive and emotional factors involved in interactions between people, especially in small groups.
One subarea of social psychology, SOCIAL COGNITION, is directly concerned with the manner in which people understand the minds, emotions, and behavior of themselves and others (see also THEORY OF MIND; INTERSUBJECTIVITY). Personality psychology deals primarily with motivational and emotional aspects of human experience (see FREUD for discussion of the ideas of the famous progenitor of this area of psychology), and clinical psychology deals with applied issues related to mental health. COMPARATIVE PSYCHOLOGY investigates the commonalities and differences in cognition and behavior between different animal species (see PRIMATE COGNITION; ANIMAL NAVIGATION; CONDITIONING; and MOTIVATION), and behavioral neuroscience provides the interface between research on molar cognition and behavior and their underlying neural substrate.

2 Capsule History of Psychology

Until the middle of the nineteenth century the nature of the mind was solely the concern of philosophers. Indeed, there are a number of reasons why some have argued that the scientific investigation of the mind may prove to be an impossible undertaking. One objection is that thoughts cannot be measured; and without measurement, science cannot even begin. A second objection is to question how humans could objectively study their own thought processes, given the fact that science itself depends on human thinking. A final objection is that our mental life is incredibly complex and bound up with the further complexities of human social interactions; perhaps cognition is simply too complex to permit successful scientific investigation. Despite these reasons for skepticism, scientific psychology emerged as a discipline separate from philosophy in the second half of the nineteenth century. A science depends on systematic empirical methods for collecting observations and on theories that interpret these observations. Beginning around 1850, a number of individuals, often trained in philosophy, physics, physiology, or neurology, began to provide these crucial elements.
The anatomist Ernst Heinrich Weber and the physicist and philosopher Gustav Fechner measured the relations between objective changes in physical stimuli, such as brightness or weight, and subjective changes in the internal sensations the stimuli generate. The crucial finding of Weber and Fechner was that subjective differences were not simply equivalent to objective differences. Rather, it turned out that for many dimensions, the magnitude of change required to make a subjective difference (“just noticeable difference,” or “jnd”) increased as overall intensity increased, often following an approximately logarithmic function, known as the Weber-Fechner Law. Weber and Fechner's contribution to cognitive psychology was much more general than identifying the law that links their names. They convincingly demonstrated that, contrary to the claim that thought is inherently impossible to measure, it is in fact possible to measure mental concepts, such as the degree of sensation produced by a stimulus.
Fechner called this new field of psychological measurement PSYCHOPHYSICS: the interface of psychology and physics, of the mental and the physical. A further

foundational issue concerns the speed of human thought. In the nineteenth century, many believed that thought was either instantaneous or else so fast that it could never be measured. But HERMANN VON HELMHOLTZ, a physicist and physiologist, succeeded in measuring the speed at which signals are conducted through the nervous system. He first experimented on frogs by applying an electric current to the top of a frog's leg and measuring the time it took the muscle at the end to twitch in response. Later he used a similar technique with humans, touching various parts of a person's body and measuring the time taken to press a button in response. The response time increased with the distance of the stimulus (i.e., the point of the touch) from the finger that pressed the button, in proportion to the length of the neural path over which the signal had to travel. Helmholtz's estimate of the speed of nerve signals was close to modern estimates—roughly 100 meters per second for large nerve fibers.

This transmission rate is surprisingly slow—vastly slower than the speed of electricity through a wire. Because our brains are composed of neurons, our thoughts cannot be generated any faster than the speed at which neurons communicate with each other. It follows that the speed of thought is neither instantaneous nor immeasurable. Helmholtz also pioneered the experimental study of vision, formulating a theory of color vision that remains highly influential today. He argued forcefully against the commonsensical idea that perception is simply a matter of somehow "copying" sensory input into the brain. Rather, he pointed out that even the most basic aspects of perception require major acts of construction by the nervous system. For example, it is possible for two different objects—a large object far away, and a small object nearby—to create precisely the same image on the retinas of a viewer's eyes. Yet normally the viewer will correctly perceive the one object as being larger, but further away, than the other. The brain somehow manages to unconsciously perform some basic geometrical calculations. The brain, Helmholtz argued, must construct this unified view by a process of "unconscious inference"—a process akin to reasoning without awareness.

Helmholtz's insight was that the "reality" we perceive is not simply a copy of the external world, but rather the product of the constructive activities of the brain. Another philosopher, HERMANN EBBINGHAUS, who was influenced by Fechner's ideas about psychophysical measurements, developed experimental methods tailored to the study of human memory. Using himself as a subject, Ebbinghaus studied memory for nonsense syllables—consonant-vowel-consonant combinations, such as "zad," "bim," and "sif." He measured how long it took to commit lists of nonsense syllables to memory, the effects of repetition on how well he could remember the syllables later, and the rate of forgetting as a function of the passage of time. Ebbinghaus made several fundamental discoveries about memory, including the typical form of the "forgetting curve"—the gradual, negatively accelerated decline in the proportion of items that can be recalled as a function of time. Like Weber, Fechner, and Helmholtz, Ebbinghaus provided evidence that it is indeed possible to measure mental phenomena by objective experimental procedures.

Many key ideas about possible components of cognition were systematically presented by the American philosopher WILLIAM JAMES in the first great psychology textbook, *Principles of Psychology*, published in 1890. His monumental work included topics that remain central in psychology, including brain function, perception, attention, voluntary movement, habit, memory, reasoning, the SELF, and hypnosis. James discussed the nature of "will," or mental effort, which remains one of the basic aspects of attention. He

also drew a distinction between different memory systems: *primary* memory, which roughly corresponds to the current contents of consciousness, and *secondary* memory, which comprises the vast store of knowledge of which we are not conscious at any single time, yet continually draw upon. Primary memory is closely related to what we now term *active*, *short-term*, or WORKING MEMORY, while secondary memory corresponds to what is usually called *long-term* memory.

James emphasized the *adaptive* nature of cognition: the fact that perception, memory, and reasoning operate not simply for their own sake, but to allow us to survive and prosper in our physical and social world. Humans evolved as organisms skilled in tool use and in social organization, and it is possible (albeit a matter of controversy) that much of our cognitive apparatus evolved to serve these basic functions (see EVOLUTIONARY PSYCHOLOGY). Thus, human cognition involves intricate systems for MOTOR CONTROL and MOTOR LEARNING; the capacity to understand that other people have minds, with intentions and goals that may lead them to help or hinder us; and the ability to recognize and remember individual persons and their characteristics. Furthermore, James (1890:8) recognized that the hallmark of an intelligent being is its ability to link ends with means—to select actions that will achieve goals: "The pursuance of future ends and the choice of means for their attainment are thus the mark and criterion of the presence of mentality in a phenomenon." This view of goal-directed thinking continues to serve as the foundation of modern work on PROBLEM SOLVING, as reflected in the views of theorists such as ALAN NEWELL and Herbert Simon.

Another pioneer of psychology was Sigmund Freud, the founder of psychoanalysis, whose theoretical ideas about cognition and consciousness anticipated many key aspects of the modern conception of cognition. Freud attacked the idea that the "self" has some special status as a unitary entity that somehow governs our thought and action. Modern cognitive psychologists also reject (though for different reasons) explanations of intelligent behavior that depend upon postulating a "homunculus"—that is, an internal mental entity endowed with all the intelligence we are trying to explain. Behavior is viewed not as the product of a unitary self or homunculus, but as the joint product of multiple interacting subsystems. Freud argued that the "ego"—the information-processing system that modulates various motivational forces—is not a unitary entity, but rather a complex system that includes attentional bottlenecks, multiple memory stores, and different ways of representing information (e.g., language, imagery, and physiognomic codes, or "body language"). Furthermore, as Freud also emphasized, much of information processing takes place at an unconscious level. We are aware of only a small portion of our overall mental life, a tip of the cognitive iceberg.

For example, operating beneath the level of awareness are attentional "gates" that open or close to selectively attend to portions of the information that reaches our senses, memory stores that hold information for very brief periods of time, and inaccessible memories that we carry with us always but might never retrieve for years at a time. Given the breadth and depth of the contributions of the nineteenth-century pioneers to what would eventually become cognitive science, it is ironic that early in the twentieth century the study of cognition went into a steep decline. Particularly in the United States, psychology in the first half of the century came to be dominated by BEHAVIORISM, an approach characterized by the rejection of theories that depended on

“mentalistic” concepts such as goals, intentions, or plans. The decline of cognitive psychology was in part due to the fact that a great deal of psychological research had moved away from the objective measurement techniques developed by Fechner, Helmholtz, Ebbinghaus, and others, and instead gave primacy to the method of INTROSPECTION, promoted by WILHELM WUNDT, in which trained observers analyzed their own thought processes as they performed various cognitive tasks. Not surprisingly, given what is now known about how expectancies influence the way we think, introspectionists tended to find themselves thinking in more or less the manner to which they were theoretically predisposed. For example, researchers who believed thinking always depended on IMAGERY usually found themselves imaging, whereas those who did not subscribe to such a theory were far more likely to report “imageless thought.”

The apparent subjectivity and inconstancy of the introspective method encouraged charges that all cognitive theories (rather than simply the method itself, as might seem more reasonable) were “unscientific.” Cognitive theories were overshadowed by the behaviorist theories of such leading figures as John Watson, Edward Thorndike, Clark Hull, and B. F. Skinner. Although there were major differences among the behaviorists in the degree to which they actually avoided explanations based on assumptions about unobservable mental states (e.g., Hull postulated such states rather freely, whereas Watson was adamant that they were scientifically illicit), none supported the range of cognitive ideas advanced in the nineteenth century.

Cognitive psychology did not simply die out during the era of behaviorism. Working within the behaviorist tradition, Edward Tolman pursued such cognitive issues as how animals represented spatial information internally as COGNITIVE MAPS of their environment. European psychologists were far less captivated with behaviorism than were Americans. In England, Sir FREDERICK BARTLETT analyzed the systematic distortions that people exhibit when trying to remember stories about unfamiliar events, and introduced the concept of “schema” (see SCHEMATA) as a mental representation that captures the systematic structural relations in categories of experience. In Soviet Russia, the neuropsychologist Aleksandr LURIA provided a detailed portrait of links between cognitive functions and the operation of specific regions of the brain. Another Russian, LEV VYGOTSKY, developed a sociohistorical approach to cognitive development that emphasized the way in which development is constructed through social interaction, cultural practices, and the internalization of cognitive tools. Vygotsky emphasized social interaction through language in the development of children’s concepts. The Swiss psychologist JEAN PIAGET spent decades refining a theory of cognitive development. Piaget's theory emphasizes milestones in the child’s development including decentration, the ability to perform operations on concrete objects, and finally the ability to perform operations on thoughts and beliefs. Given its emphasis on logical thought, Piaget's theory is closely related to SCIENTIFIC THINKING AND ITS DEVELOPMENT.

In addition, the great German tradition in psychology, which had produced so many of the nineteenth-century pioneers, gave rise to a new cognitive movement in the early twentieth century: GESTALT PSYCHOLOGY. The German word *Gestalt* translates roughly as “form,” and the Gestalt psychologists emphasized that the whole form is something different from the mere sum of its parts, due to emergent properties that arise

as new relations are created. Gestalt psychology was in some ways an extension of Helmholtz's constructivist ideas, and the greatest contributions of this intellectual movement were in the area of GESTALT PERCEPTION. Where the behaviorists insisted that psychology was simply the study of how objective stimuli come to elicit objective responses, the Gestaltists pointed to simple demonstrations casting doubt on the idea that "objective" stimuli—that is, stimuli perceived in a way that can be described strictly in terms of the sensory input—even exist. Figure 1 illustrates a famous Gestalt example of the constructive nature of perception, the ambiguous Necker cube. Although this figure is simply a flat line drawing, we immediately perceive it as a three-dimensional cube. Moreover, if you look carefully, you will see that the figure can actually be seen as either of two different three-dimensional cubes. The same objective stimulus—the two-dimensional line drawing—gives rise to two distinct three-dimensional perceptions.

Although many of the major contributions by key Gestalt figures such as Max Wertheimer were in the area of perception, their central ideas were extended to memory and problem solving as well, through the work of people such as Wolfgang Köhler and Karl Duncker. Indeed, one of the central tenets of Gestalt psychology was that highlevel thinking is based on principles similar to those that govern basic perception. As we do in everyday language, Gestalt psychologists spoke of suddenly "seeing" the solution to a problem, often after "looking at it" in a different way and achieving a new "insight." In all the areas in which they worked, the Gestalt idea of "a whole different from the sum of parts" was based on the fundamental fact that organized configurations are based not simply on individual elements, but also on the relations between those elements. Just as H2O is not simply two hydrogen atoms and one oxygen atom, but also a particular spatial organization of these elements into a configuration that makes a molecule of water, so too "squareness" is more than four lines: it crucially depends on the way the lines are related to one another to make four right angles. Furthermore, relations can take on a "life of their own," separable from any particular set of elements. For example, we can take a tune, move it to a different key so that all the notes are changed, and still immediately recognize it as the "same" tune as long as the relations among the notes are preserved. A focus on relations calls attention to the centrality of the BINDING PROBLEM, which involves the issue of how elements are systematically organized to fill relational roles. Modern work on such topics as ANALOGY and SIMILARITY emphasizes the crucial role of relations in cognition.

Modern cognitive psychology emerged in the second half of this century. The "cognitive revolution" of the 1950s and 1960s involved not only psychology but also the allied disciplines that now contribute to cognitive science. In the 1940s the Canadian psychologist DONALD HEBB began to draw connections between cognitive processes and neural mechanisms, anticipating modern cognitive neuroscience. During World War II, many experimental psychologists (including JAMES GIBSON) were confronted with such pressing military problems as finding ways to select good pilots and train radar operators, and it turned out that the then-dominant stimulus-response theories simply had little to offer in the way of solutions. More detailed process models of human information processing were needed. After the war, DONALD BROADBENT in England developed the first such detailed model of attention. Even more importantly, Broadbent helped develop and popularize a wide range of experimental tasks in which an observer's attention is carefully controlled by having him or her perform some task, such as listening

to a taped message for a particular word, and then precisely measuring how quickly responses can be made and what can be remembered. In the United States, William K. Estes added to the mathematical tools available for theory building and data analysis, and Saul Sternberg developed a method for decomposing reaction times into component processes using a simple recognition task.

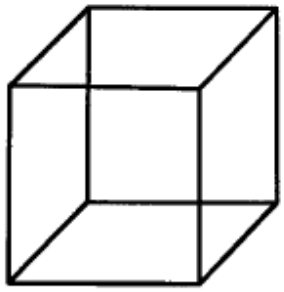

Figure 1.

Meanwhile, the birth of computer science provided further conceptual tools. Strict behaviorists had denounced models of internal mental processes as unscientific. However, the modern digital computer provided a clear example of a device that took inputs, fed them through a complex series of internal procedures, and then produced outputs. As well as providing concrete examples of what an information-processing device could be, computers made possible the beginnings of artificial intelligence—the construction of computer programs designed to perform tasks that require intelligence, such as playing chess, understanding stories, or diagnosing diseases. Herbert Simon (1978 Nobel Laureate in Economics) and Allan Newell were leaders in building close ties between artificial intelligence and the new cognitive psychology. It was also recognized that actual computers represent only a small class of a much larger set of theoretically possible computing devices, which had been described back in the 1940s by the brilliant mathematician ALAN TURING. Indeed, it was now possible to view the brain itself as a biological computer, and to use various real and possible computing devices as models of human cognition. Another key influence on modern cognitive psychology came from the field of linguistics. In the late 1950s work by the young linguist Noam Chomsky radically changed conceptions of the nature of human language by demonstrating that language could not be learned or understood by merely associating adjacent words, but rather required computations on abstract structures that existed in the minds of the speaker and listener.

The collective impact of this work in the mid-twentieth century was to provide a seminal idea that became the foundation of cognitive psychology and also cognitive science in general: the COMPUTATIONAL THEORY OF MIND, according to which human cognition is based on mental procedures that operate on abstract mental representations. The nature of the COGNITIVE ARCHITECTURE has been controversial, including proposals such as PRODUCTION SYSTEMS and NEURAL NETWORKS. In particular, there has been disagreement as to whether procedures and representations are inherently separable or whether procedures actually embody representations, and whether some

mental representations are abstract and amodal, rather than tied to specific perceptual systems. Nonetheless, the basic conception of biological information processing as some form of computation continues to guide psychological theories of the representation and processing of information.

3 The Science of Information Processing

In broad strokes, an intelligent organism operates in a perception-action cycle (Neisser 1967), taking in sensory information from the environment, performing internal computations on it, and using the results of the computation to guide the selection and execution of goal-directed actions. The initial sensory input is provided by separate sensory systems, including smell, taste, haptic perception, and audition. The most sophisticated sensory system in primates is vision (see MID-LEVEL VISION; HIGH-LEVEL VISION), which includes complex specialized subsystems for DEPTH PERCEPTION, SHAPE PERCEPTION, LIGHTNESS PERCEPTION, and COLOR VISION.

The interpretation of sensory inputs begins with FEATURE DETECTORS that respond selectively to relatively elementary aspects of the stimulus (e.g., lines at specific orientations in the visual field, or phonetic cues in an acoustic speech signal). Some basic properties of the visual system result in systematic misperceptions, or ILLUSIONS. TOP-DOWN PROCESSING IN VISION serves to integrate the local visual input with the broader context in which it occurs, including prior knowledge stored in memory. Theorists working in the tradition of Gibson emphasize that a great deal of visual information may be provided by higher-order features that become available to a perceiver moving freely in a natural environment, rather than passively viewing a static image (see ECOLOGICAL PSYCHOLOGY).

In their natural context, both perception and action are guided by the AFFORDANCES of the environment: properties of objects that enable certain uses (e.g., the elongated shape of a stick may afford striking an object otherwise out of reach). Across all the sensory systems, psychophysics methods are used to investigate the quantitative functions relating physical inputs received by sensory systems to subjective experience (e.g., the relation between luminance and perceived brightness, or between physical and subjective weight). SIGNAL DETECTION THEORY provides a statistical method for measuring how accurately observers can distinguish a signal from noise under conditions of uncertainty (i.e., with limited viewing time or highly similar alternatives) in a way that separates the signal strength received from possible response bias. In addition to perceiving sensory information about objects at locations in space, animals perceive and record information about time (see TIME IN THE MIND).

Knowledge about both space and time must be integrated to provide the capability for animal and HUMAN NAVIGATION in the environment. Humans and other animals are capable of forming sophisticated representations of spatial relations integrated as COGNITIVE MAPS. Some more central mental representations appear to be closely tied to perceptual systems. Humans use various forms of imagery based on visual, auditory and other perceptual systems to perform internal mental processes such as MENTAL

ROTATION. The close connection between PICTORIAL ART AND VISION also reflects the links between perceptual systems and more abstract cognition.

A fundamental property of biological information processing is that it is capacitylimited and therefore necessarily selective. Beginning with the seminal work of Broadbent, a great deal of work in cognitive psychology has focused on the role of attention in guiding information processing. Attention operates selectively to determine what information is received by the senses, as in the case of EYE MOVEMENTS AND VISUAL ATTENTION, and also operates to direct more central information processing, including the operation of memory. The degree to which information requires active attention or memory resources varies, decreasing with the AUTOMATICITY of the required processing.

Modern conceptions of memory maintain some version of William James's basic distinction between primary and secondary memory. Primary memory is now usually called WORKING MEMORY, which is itself subdivided into multiple stores involving specific forms of representation, especially phonological and visuospatial codes. Working memory also includes a central executive, which provides attentional resources for strategic management of the cognitive processes involved in problem solving and other varieties of deliberative thought. Secondary or long-term memory is also viewed as involving distinct subsystems, particularly EPISODIC VS. SEMANTIC MEMORY. Each of these subsystems appears to be specialized to perform one of the two basic functions of long-term memory. One function is to store individuated representations of "what happened when" in specific contexts (episodic memory); a second function is to extract and store generalized representations of "the usual kind of thing" (semantic memory). Another key distinction, related to different types of memory measures, is between IMPLICIT VS. EXPLICIT MEMORY. In explicit tests (typically recall or recognition tests), the person is aware of the requirement to access memory. In contrast, implicit tests (such as completing a word stem, or generating instances of a category) make no reference to any particular memory episode. Nonetheless, the influence of prior experiences may be revealed by the priming of particular responses (e.g., if the word "crocus" has recently been studied, the person is more likely to generate "crocus" when asked to list flowers, even if they do not explicitly remember having studied the word).

There is evidence that implicit and explicit knowledge are based on separable neural systems. In particular, forms of amnesia caused by damage to the hippocampus and related structures typically impair explicit memory for episodes, but not implicit memory as revealed by priming measures. A striking part of human cognition is the ability to speak and comprehend language. The psychological study of language, or psycholinguistics, has a close relationship to work in linguistics and on LANGUAGE ACQUISITION. The complex formal properties of language, together with its apparent ease of acquisition by very young children, have made it the focus of debates about the extent and nature of NATIVISM in cognition. COMPUTATIONAL PSYCHOLINGUISTICS is concerned with modeling the complex processes involved in language use. In modern cultures that have achieved LITERACY with the introduction of written forms of language, the process of READING lies at the interface of psycholinguistics, perception, and memory retrieval. The intimate relationship between language and thought, and between language and human concepts, is widely recognized but still poorly understood. The use of METAPHOR in language is related to other

symbolic processes in human cognition, particularly ANALOGY and CATEGORIZATION.

One of the most fundamental aspects of biological intelligence is the capacity to adaptively alter behavior. It has been clear at least from the time of William James that the adaptiveness of human behavior and the ability to achieve EXPERTISE in diverse domains is not generally the direct product of innate predispositions, but rather the result of adaptive problem solving and LEARNING SYSTEMS that operate over the lifespan. Both production systems and neural networks provide computational models of some aspects of learning, although no model has captured anything like the full range of human learning capacities. Humans as well as some other animals are able to learn by IMITATION, for example, translating visual information about the behavior of others into motor routines that allow the observer/imitator to produce comparable behavior. Many animal species are able to acquire expectancies about the environment and the consequences of the individual's actions on the basis of CONDITIONING, which enables learning of contingencies among events and actions. Conditioning appears to be a primitive form of causal induction, the process by which humans and other animals learn about the cause-effect structure of the world. Both causal knowledge and similarity relations contribute to the process of categorization, which leads to the development of categories and concepts that serve to organize knowledge. People act as if they assume the external appearances of category members are caused by hidden (and often unknown) internal properties (e.g., the appearance of an individual dog may be attributed to its internal biology), an assumption sometimes termed psychological ESSENTIALISM.

There are important developmental influences that lead to CONCEPTUAL CHANGE over childhood. These developmental aspects of cognition are particularly important in understanding SCIENTIFIC THINKING AND ITS DEVELOPMENT. Without formal schooling, children and adults arrive at systematic beliefs that comprise NAIVE MATHEMATICS and NAIVE PHYSICS. Some of these beliefs provide the foundations for learning mathematics and physics in formal EDUCATION, but some are misconceptions that can impede learning these topics in school (see also AI AND EDUCATION). Young children are prone to ANIMISM, attributing properties of people and other animals to plants and nonliving things. Rather than being an aberrant form of early thought, animism may be an early manifestation of the use of ANALOGY to make inferences and learn new cognitive structures. Analogy is the process used to find systematic structural correspondences between a familiar, well-understood situation and an unfamiliar, poorly understood one, and then using the correspondences to draw plausible inferences about the less familiar case. Analogy, along with hypothesis testing and evaluation of competing explanations, plays a role in the discovery of new regularities and theories in science.

In its more complex forms, learning is intimately connected to thinking and reasoning. Humans are not only able to think, but also to think *about* their own cognitive processes, resulting in METACOGNITION. They can also form higher-level representations, termed METAREPRESENTATION. There are major individual differences in intelligence as assessed by tasks that require abstract thinking. Similarly, people differ in their CREATIVITY in finding solutions to problems. Various neural disorders, such as forms of MENTAL RETARDATION and AUTISM, can impair or radically alter normal

thinking abilities. Some aspects of thinking are vulnerable to disruption in later life due to the links between AGING AND COGNITION.

Until the last few decades, the psychology of DEDUCTIVE REASONING was dominated by the view that human thinking is governed by formal rules akin to those used in LOGIC. Although some theorists continue to argue for a role for formal, content-free rules in reasoning, others have focused on the importance of content-specific rules. For example, people appear to have specialized procedures for reasoning about broad classes of pragmatically important tasks, such as understanding social relations or causal relations among events. Such pragmatic reasoning schemas (Cheng and Holyoak 1985) enable people to derive useful inferences in contexts related to important types of recurring goals. In addition, both deductive and inductive inferences may sometimes be made using various types of MENTAL MODELS, in which specific possible cases are represented and manipulated (see also CASE-BASED REASONING AND ANALOGY).

Much of human inference depends not on deduction, but on inductive PROBABILISTIC REASONING under conditions of UNCERTAINTY. Work by researchers such as AMOS TVERSKY and Daniel Kahneman has shown that everyday inductive reasoning and decision making is often based on simple JUDGMENT HEURISTICS related to ease of memory retrieval (the *availability* heuristic) and degree of similarity (the *representativeness* heuristic). Although judgment heuristics are often able to produce fast and accurate responses, they can sometimes lead to errors of prediction (e.g., conflating the subjective ease of remembering instances of a class of events with their objective frequency in the world).

More generally, the impressive power of human information processing has apparent limits. People all too often take actions that will not achieve their intended ends, and pursue short-term goals that defeat their own long-term interests. Some of these mistakes arise from motivational biases, and others from computational limitations that constrain human attention, memory, and reasoning processes. Although human cognition is fundamentally adaptive, we have no reason to suppose that “all’s for the best in this best of all possible minds.”

References

Cheng, P. W., and K. J. Holyoak. (1985). Pragmatic reasoning schemas. *Cognitive Psychology* 17: 391–394.

James, W. (1890). *The Principles of Psychology.* New York: Dover.

Neisser, U. (1967). *Cognitive Psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Further Readings

Anderson, J. R. (1995). *Cognitive Psychology and Its Implications.* 4th ed. San Francisco: W. H. Freeman.

Baddeley, A. D. (1997). *Human Memory: Theory and Practice.* 2nd ed. Hove, Sussex: Psychology Press.

Evans, J., S. E. Newstead, and R. M. J. Byrne. (1993). *Human Reasoning.* Mahwah, NJ: Erlbaum.

Gallistel, C. R. (1990). *The Organization of Learning.* Cambridge, MA: MIT Press.

Gazzaniga, M. S. (1995). *The Cognitive Neurosciences.* Cambridge, MA: MIT Press.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton-Mifflin.

Gregory, R. L. (1997). *Eye and Brain: The Psychology of Seeing.* 5th ed. Princeton, NJ: Princeton University Press.

Holyoak, K. J., and P. Thagard. (1995). *Mental Leaps: Analogy in Creative Thought.* Cambridge, MA: MIT Press.

James, W. (1890). *Principles of Psychology.* New York: Dover.

Kahneman, D., P. Slovic, and A. Tversky. (1982). *Judgments Under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press.

Keil, F. C. (1989). *Concepts, Kinds, and Cognitive Development.* Cambridge, MA: MIT Press.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate.* Cambridge, MA: MIT Press.

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

Pashler, H. (1997). *The Psychology of Attention.* Cambridge, MA: MIT Press.

Pinker, S. (1994). *The Language Instinct.* New York: William Morrow.

Reisberg, D. (1997). *Cognition: Exploring the Science of the Mind.* New York: Norton.

Rumelhart, D. E., J. L. McClelland, and PDP Research Group. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* 2 vols. Cambridge, MA: MIT Press.

Smith, E. E., and D. L. Medin. (1981). *Categories and Concepts.* Cambridge, MA: Harvard University Press.

Sperber, D., D. Premack, and A. J. Premack. (1995). *Causal Cognition: A Multidisciplinary Debate.* Oxford: Clarendon Press.

Tarpy, R. M. (1997). *Contemporary Learning Theory and Research.* New York: McGraw Hill.

Tomasello, M., and J. Call. (1997). *Primate Cognition.* New York: Oxford University Press.

Neurosciences

Thomas D. Albright and Helen J. Neville

1 Cognitive Neuroscience

The term alone suggests a field of study that is pregnant and full of promise. It is a large field of study, uniting concepts and techniques from many disciplines, and its boundaries are rangy and often loosely defined. At the heart of cognitive neuroscience, however, lies the fundamental question of knowledge and its representation by the brain—a relationship characterized not inappropriately by WILLIAM JAMES (1842–1910) as "the most mysterious thing in the world" (James 1890 vol. 1, 216). Cognitive neuroscience is thus a science of information processing. Viewed as such, one can identify key experimental questions and classical areas of study: How is information acquired (sensation), interpreted to confer meaning (perception and recognition), stored or modified (learning and memory), used to ruminate (thinking and consciousness), to predict the future state of the environment and the consequences of action (decision making), to guide behavior (motor control), and to communicate (language)?

These questions are, of course, foundational in cognitive science generally, and it is instructive to consider what distinguishes cognitive neuroscience from cognitive science and psychology, on the one hand, and the larger field of neuroscience, on the other.

The former distinction is perhaps the fuzzier, depending heavily as it does upon how one defines cognitive science. A neurobiologist might adopt the progressive (or naive) view that the workings of the brain are the subject matter of both, and the distinction is therefore moot. But this view evidently has not prevailed (witness the fact that neuroscience is but one of the subdivisions of this volume); indeed the field of cognitive science was founded upon and continues to press the distinction between software (the content of cognition) and hardware (the physical stuff, for example, the brain) upon which cognitive processes are implemented.

Much has been written on this topic, and one who pokes at the distinction too hard is likely to unshelve as much dusty political discourse as true science. In any case, for present purposes, we will consider both the biological hardware and the extent to which it constrains the software, and in doing so we will discuss answers to the questions of cognitive science that are rooted in the elements of biological systems.

The relationship between cognitive neuroscience and the umbrella of modern neuroscience is more straightforward and less embattled. While the former is clearly a subdivision of the latter, the questions of cognitive neuroscience lie at the root of much of neuroscience's turf. Where distinctions are often made, they arise from the fact that cognitive neuroscience is a functional neuroscience—particular structures and signals of the nervous system are of interest inasmuch as they can be used to explain cognitive functions.

There being many levels of explanation in biological systems—ranging from cellular and molecular events to complex behavior—a key challenge of the field of cognitive neuroscience has been to identify the relationships between different levels and the train of causality. In certain limited domains, this challenge has met with spectacular success; in others, it is clear that the relevant concepts have only begun to take shape and the necessary experimental tools are far behind. Using examples drawn from well-developed areas of research, such as vision, memory, and language, we illustrate concepts, experimental approaches, and general principles that have emerged—and, more specifically, how the work has answered many of the information processing questions identified above. Our contemporary view of cognitive neuroscience owes much to the heights attained by our predecessors; to appreciate the state of this field fully, it is useful to begin with a consideration of how we reached this vantage point.

2 Origins of Cognitive Neuroscience

Legend has it that the term "cognitive neuroscience" was coined by George A. Miller—the father of modern cognitive psychology—in the late 1970s over cocktails with Michael Gazzaniga at the Rockefeller University Faculty Club. That engaging tidbit of folklore nevertheless belies the ancient history of this pursuit. Indeed, identification of the biological structures and events that account for our ability to acquire, store, and utilize knowledge of the world was one of the earliest goals of empirical science. The emergence of the interdisciplinary field of cognitive neuroscience that we know today, which lies squarely at the heart of twentieth-century neuroscience, can thus be traced from a common stream in antiquity, with many tributaries converging in time as new concepts and techniques have evolved (Boring 1950).

Localization of Function

The focal point of the earliest debates on the subject—and a topic that has remained a centerpiece of cognitive neuroscience to the present day—is localization of the material source of psychological functions. With Aristotle as a notable exception (he thought the heart more important), scholars of antiquity rightly identified the brain as the seat of intellect. Relatively little effort was made to localize specific mental functions to particular brain regions until the latter part of the eighteenth century, when the anatomist Franz Josef Gall (1758–1828) unleashed the science of phrenology. Although flawed in its premises, and touted by charlatans, phrenology focused attention on the CEREBRAL CORTEX and brought the topic of localization of function to the forefront of an emerging nineteenth century physiology and psychology of mind (Zola-Morgan 1995). The subsequent HISTORY OF CORTICAL LOCALIZATION of function (Gross 1994a) is filled with colorful figures and weighty confrontations between localizationists and functional holists (antilocalizationists).

Among the longest shadows is that cast by PAUL BROCA (1824–1880), who in 1861 reported that damage to a "speech center" in the left frontal lobe resulted in loss of speech function, and was thus responsible for the

first widely cited evidence for localization of function in the cerebral cortex. An important development of a quite different nature came in the form of the Bell-Magendie law, discovered independently in the early nineteenth century by the physiologists Sir Charles Bell (1774–1842) and François Magendie (1783–1855). This law identified the fact that sensory and motor nerve fibers course through different roots (dorsal and ventral, respectively) of the spinal cord. Although far from the heavily contested turf of the cerebral cortex, the concept of nerve specificity paved the way for the publication in 1838 by Johannes Muller (1801–1858) of the law of specific nerve energies, which included among its principles the proposal that nerves carrying different types of sensory information terminate in distinct brain loci, perhaps in the cerebral cortex.

Persuasive though the accumulated evidence seemed at the dawn of the twentieth century, the debate between localizationists and antilocalizationists raged on for another three decades. By this time the chief experimental tool had become the "lesion method," through which the functions of specific brain regions are inferred from the behavioral or psychological consequences of loss of the tissue in question (either by clinical causes or deliberate experimental intervention). A central player during this period was the psychologist KARL SPENCER LASHLEY (1890–1958)—often inaccurately characterized as professing strong antilocalizationist beliefs, but best known for the concept of equipotentiality and the law of mass action of brain function.

Lashley's descendants include several generations of flag bearers for the localizationist front—Carlyle Jacobsen, John Fulton, Karl Pribram, Mortimer Mishkin, Lawrence Weiskrantz, and Charles Gross, among others—who established footholds for our present understanding of the cognitive functions of the frontal and temporal lobes. These later efforts to localize cognitive functions using the lesion method were complemented by studies of the effects of electrical stimulation of the human brain on psychological states. The use of stimulation as a probe for cognitive function followed its more pragmatic application as a functional brain mapping procedure executed in preparation for surgical treatment of intractable epilepsy. The neurosurgeon WILDER PENFIELD (1891–1976) pioneered this approach in the 1930s at the legendary Montreal Neurological Institute and, with colleagues Herbert Jasper and Brenda Milner, subsequently began to identify specific cortical substrates of language, memory, emotion, and perception.

The years of the mid-twentieth century were quarrelsome times for the expanding field of psychology, which up until that time had provided a home for much of the work on localization of brain function. It was from this fractious environment, with inspiration from the many successful experimental applications of the lesion method and a growing link to wartime clinical populations, that the field of neuropsychology emerged—and with it the wagons were drawn up around the first science explicitly devoted to the relationship between brain and cognitive function. Early practitioners included the great Russian neuropsychologist ALEKSANDR ROMANOVICH LURIA (1902–1977) and the American behavioral neurologist NORMAN GESCHWIND (1926–1984), both of whom promoted the localizationist cause with human case studies and focused

attention on the role of connections between functionally specific brain regions. Also among the legendary figures of the early days of neuropsychology was HANS-LUKAS TEUBER (1916–1977). Renowned scientifically for his systematization of clinical neuropsychology, Teuber is perhaps best remembered for having laid the cradle of modern cognitive neuroscience in the 1960s MIT Psychology Department, through his inspired recruitment of an interdisciplinary faculty with a common interest in brain structure and function, and its relationship to complex behavior (Gross 1994b).

Neuron Doctrine

Although the earliest antecedents of modern cognitive neuroscience focused by necessity on the macroscopic relationship between brain and psychological function, the last 50 years have seen a shift of focus, with major emphasis placed upon local neuronal circuits and the causal link between the activity of individual cells and behavior. The payoff has been astonishing, but one often takes for granted the resolution of much hotly debated turf. The debates in question focused on the elemental units of nervous system structure and function. We accept these matter-of-factly to be specialized cells known as NEURONS, but prior to the development of techniques to visualize cellular processes, their existence was mere conjecture.

Thus the two opposing views of the nineteenth century were reticular theory, which held that the tissue of the brain was composed of a vast anastomosing reticulum, and neuron theory, which postulated neurons as differentiated cell types and the fundamental unit of nervous system function. The ideological chasm between these camps ran deep and wide, reinforced by ties to functional holism in the case of reticular theory, and localizationism in the case of neuron theory. The deadlock broke in 1873 when CAMILLO GOLGI (1843–1926) introduced a method for selective staining of individual neurons using silver nitrate, which permitted their visualization for the first time. (Though this event followed the discovery of the microscope by approximately two centuries, it was the Golgi method's complete staining of a minority of neurons that enabled them to be distinguished from one another.) In consequence, the neuron doctrine was cast, and a grand stage was set for studies of differential cellular morphology, patterns of connectivity between different brain regions, biochemical analysis, and, ultimately, electrophysiological characterization of the behavior of individual neurons, their synaptic interactions, and relationship to cognition.

Undisputedly, the most creative and prolific applicant of the Golgi technique was the Spanish anatomist SANTIAGO RAMÓN Y CAJAL (1852–1934), who used this new method to characterize the fine structure of the nervous system in exquisite detail. Cajal's efforts yielded a wealth of data pointing to the existence of discrete neuronal elements. He soon emerged as a leading proponent of the neuron doctrine and subsequently shared the 1906 Nobel Prize in physiology and medicine with Camillo Golgi. (Ironically, Golgi held vociferously to the reticular theory throughout his career.) Discovery of the existence of independent neurons led naturally

to investigations of their means of communication. The fine-scale stereotyped contacts between neurons were evident to Ramón y Cajal, but it was Sir Charles Scott Sherrington (1857–1952) who, at the turn of the century, applied the term "synapses" to label them. The transmission of information across synapses by chemical means was demonstrated experimentally by Otto Loewi (1873–1961) in 1921. The next several decades saw an explosion of research on the nature of chemical synaptic transmission, including the discovery of countless putative NEUROTRANSMITTERS and their mechanisms of action through receptor activation, as well as a host of revelations regarding the molecular events that are responsible for and consequences of neurotransmitter release. These findings have provided a rich foundation for our present understanding of how neurons compute and store information about the world (see COMPUTING IN SINGLE NEURONS).

The ability to label neurons facilitated two other noteworthy developments bearing on the functional organization of the brain: (1) cytoarchitectonics, which is the use of coherent regional patterns of cellular morphology in the cerebral cortex to identify candidates for functional specificity; and (2) neuroanatomical tract tracing, by which the patterns of connections between and within different brain regions are established.

The practice of cytoarchitectonics began at the turn of the century and its utility was espoused most effectively by the anatomists Oscar Vogt (1870–1950), Cecile Vogt (1875–1962), and Korbinian Brodmann (1868–1918). Cytoarchitectonics never fully achieved the functional parcellation that it promised, but clear histological differences across the cerebral cortex, such as those distinguishing primary visual and motor cortices from surrounding tissues, added considerable reinforcement to the localizationist camp. By contrast, the tracing of neuronal connections between different regions of the brain, which became possible in the late nineteenth century with the development of a variety of specialized histological staining techniques, has been an indispensable source of knowledge regarding the flow of information through the brain and the hierarchy of processing stages. Recent years have seen the emergence of some remarkable new methods for tracing individual neuronal processes and for identifying the physiological efficacy of specific anatomical connections (Callaway 1998), the value of which is evidenced most beautifully by studies of the CELL TYPES AND CONNECTIONS IN THE VISUAL CORTEX.

The neuron doctrine also paved the way for an understanding of the information represented by neurons via their electrical properties, which has become a cornerstone of cognitive neuroscience in the latter half of the twentieth century. The electrical nature of nervous tissue was well known (yet highly debated) by the beginning of the nineteenth century, following advancement of the theory of "animal electricity" by Luigi Galvani (1737–1798) in 1791. Subsequent work by Emil du Bois-Reymond (1818–1896), Carlo Matteucci (1811–1862), and HERMANN LUDWIG FERDINAND VON HELMHOLTZ (1821–1894) established the spreading nature of electrical potentials in nervous tissue (nerve conduction), the role of the nerve membrane in maintaining and propagating an electrical charge ("wave of negativity"), and the velocity of nervous conduction. It was in the 1920s that Lord Edgar Douglas Adrian (1889–1977), using new cathode ray tube

and amplification technology, developed the means to record "action potentials" from single neurons. Through this means, Adrian discovered the "all-or-nothing property" of nerve conduction via action potentials and demonstrated that action potential frequency is the currency of information transfer by neurons.

Because of the fundamental importance of these discoveries, Adrian shared the 1932 Nobel Prize in physiology and medicine with Sherrington. Not long afterward, the Finnish physiologist Ragnar Granit developed techniques for recording neuronal activity using electrodes placed on the surface of the skin (Granit discovered the electroretinogram, or ERG, which reflects large-scale neuronal activity in the RETINA). These techniques became the foundation for non-invasive measurements of brain activity (see ELECTROPHYSIOLOGY, ELECTRIC AND MAGNETIC EVOKED FIELDS), which have played a central role in human cognitive neuroscience over the past 50 years.

With technology for SINGLE-NEURON RECORDING and large-scale electrophysiology safely in hand, the mid-twentieth century saw a rapid proliferation of studies of physiological response properties in the central nervous system. Sensory processing and motor control emerged as natural targets for investigation, and major emphasis was placed on understanding (1) the topographic mapping of the sensory or motor field onto central target zones (such as the retinotopic mapping in primary visual cortex), and (2) the specific sensory or motor events associated with changes in frequency of action potentials. Although some of the earliest and most elegant research was directed at the peripheral auditory system—culminating with Georg von Bekesy's (1889–1972) physical model of cochlear function and an understanding of its influence on AUDITORY PHYSIOLOGY—it is the visual system that has become the model for physiological investigations of information processing by neurons.

The great era of single-neuron studies of visual processing began in the 1930s with the work of Haldan Keffer Hartline (1903–1983), whose recordings from the eye of the horseshoe crab (Limulus) led to the discovery of neurons that respond when stimulated by light and detect differences in the patterns of illumination (i.e., contrast; Hartline, Wagner, and MacNichol 1952). It was for this revolutionary advance that Hartline became a corecipient of the 1967 Nobel Prize in physiology and medicine (together with Ragnar Granit and George Wald). Single-neuron studies of the mammalian visual system followed in the 1950s, with the work of Steven Kuffler (1913– 1980) and Horace Barlow, who recorded from retinal ganglion cells. This research led to the development of the concept of the center-surround receptive field and highlighted the key role of spatial contrast detection in early vision (Kuffler 1953). Subsequent experiments by Barlow and Jerome Lettvin, among others, led to the discovery of neuronal FEATURE DETECTORS for behaviorally significant sensory inputs. This set the stage for the seminal work of David Hubel and Torsten Wiesel, whose physiological investigations of visual cortex, beginning in the late 1950s, profoundly shaped our understanding of the relationship between neuronal and sensory events (Hubel and Wiesel 1977).

Sensation, Association, Perception, and Meaning

The rise of neuroscience from its fledgling origins in the nineteenth century was paralleled by the growth of experimental psychology and its embracement of sensation and perception as primary subject matter. The origins of experimental psychology as a scientific discipline coincided, in turn, with the convergence and refinement of views on the nature of the difference between sensation and perception. These views, which began to take their modern shape with the concept of "associationism" in the empiricist philosophy of John Locke (1632–1704), served to focus attention on the extraction of meaning from sensory events and, not surprisingly, lie at the core of much twentieth century cognitive neuroscience.

The proposition that things perceived cannot reflect directly the material of the external world, but rather depend upon the states of the sense organs and the intermediary nerves, is as old as rational empiricism itself. Locke's contribution to this topic was simply that meaning—knowledge of the world, functional relations between sensations, nee perception—is born from an association of "ideas," of which sensation was the primary source. The concept was developed further by George Berkeley (1685–1753) in his "theory of objects," according to which a sensation has meaning—that is, a reference to an external material source—only via the context of its relationship to other sensations. This associationism was a principal undercurrent of Scottish and English philosophy for the next two centuries, the concepts refined and the debate further fueled by the writings of James Mill and, most particularly, John Stuart Mill. It was the latter who defined the "laws of association" between elemental sensations, and offered the useful dictum that perception is the belief in the "permanent possibilities of sensation." By so doing, Mill bridged the gulf between the ephemeral quality of sensations and the permanence of objects and our experience of them: it is the link between present sensations and those known to be possible (from past experience) that allows us to perceive the enduring structural and relational qualities of the external world.

In the mid-nineteenth century the banner of associationism was passed from philosophy of mind to the emerging German school of experimental psychology, which numbered among its masters Gustav Fechner (1801–1887), Helmholtz, WILHELM WUNDT (1832–1920), and the English-American disciple of that tradition Edward Titchener (1867–1927). Fechner's principal contribution in this domain was the introduction of a systematic scientific methodology to a topic that had before that been solely the province of philosophers and a target of introspection. Fechner's Elements of Psychophysics, published in 1860, founded an "exact science of the functional relationship . . . between body and mind," based on the assumption that the relationship between brain and perception could be measured experimentally as the relationship between a stimulus and the sensation it gives rise to. PSYCHOPHYSICS thus provided the new nineteenth-century psychology with tools of a rigorous science and has subsequently become a mainstay of modern cognitive neuroscience. It was during this move toward quantification and systematization that Helmholtz upheld the prevailing associationist view of objects as sensations bound together through experience and memory, and he advanced the concept of unconscious inference to account for the attribution of perceptions to

specific environmental causes. Wundt pressed further with the objectification and deconstruction of psychological reality by spelling out the concept—implicit in the manifestoes of his associationist predecessors—of elementism.

Although Wundt surely believed that the meaning of sensory events lay in the relationship between them, elementism held that any complex association of sensations—any perception—was reducible to the sensory elements themselves. Titchener echoed the Wundtian view and elaborated upon the critical role of context in the associative extraction of meaning from sensation. It was largely in response to this doctrine of elementism, its spreading influence, and its corrupt reductionistic account of perceptual experience that GESTALT PSYCHOLOGY was born in the late nineteenth century. In simplest terms, the Gestalt theorists, led by the venerable trio of Max Wertheimer (1880–1943), Wolfgang Kohler (1887–1967), and Kurt Koffka (1886–1941), insisted—and backed up their insistence with innumerable compelling demonstrations—that our phenomenal experience of objects, which includes an appreciation of their meanings and functions, is not generally reducible to a set of elemental sensations and the relationships between them.

Moreover, rather than accepting the received wisdom that perception amounts to an inference about the world drawn from the associations between sensations, the Gestalt theorists held the converse to be true: perception is native experience and efforts to identify the underlying sensory elements are necessarily inferential (Koffka 1935). In spite of other flaws and peculiarities of the broad-ranging Gestalt psychology, this holistic view of perception, its distinction from sensation, and the nature of meaning, has become a central theme of modern cognitive neuroscience. At the time the early associationist doctrine was being formed, there emerged a physiological counterpart in the form of Johannes Muller's (1801–1858) law of specific nerve energies, which gave rise in turn to the concept of specific fiber energies, and, ultimately, our twentieth-century receptive fields and feature detectors. Muller's law followed, intellectually as well as temporally, the Bell-Magendie law of distinct sensory and motor spinal roots, which set a precedent for the concept of specificity of nerve action. Muller's law was published in his 1838 Handbook of Physiology and consisted of several principles, those most familiar being the specificity of the sensory information (Muller identified five kinds) carried by different nerves and the specificity of the site of termination in the brain (a principle warmly embraced by functional localizationists of the era). For present discussion, the essential principle is that "the immediate objects of the perception of our senses are merely particular states induced in the nerves, and felt as sensations either by the nerves themselves or by the sensorium" (Boring 1950). Muller thus sidestepped the ancient problem of the mind's access to the external world by observing that all it can hope to access is the state of its sensory nerves. Accordingly, perception of the external world is a consequence of the stable relationship between external stimuli and nerve activation, and—tailing the associationist philosophers—meaning is granted by the associative interactions between nerves carrying different types of information. The concept was elaborated further by Helmholtz and others to address the different submodalities (e.g., color vs. visual distance) and

qualities (e.g., red vs. green) of information carried by different fibers, and is a tenet of contemporary sensory neurobiology and cognitive neuroscience. The further implications of associationism for an understanding of the neuronal basis of perception—or, more precisely, of functional knowledge of the world—are profound and, as we shall see, many of the nineteenth-century debates on the topic are being replayed in the courts of modern single-neuron physiology.

3 Cognitive Neuroscience Today

And so it was from these ancient but rapidly converging lines of inquiry, with the blush still on the cheek of a young cognitive science, that the modern era of cognitive neuroscience began. The field continues to ride a groundswell of optimism borne by new experimental tools and concepts—particularly single-cell electrophysiology, functional brain imaging, molecular genetic manipulations, and neuronal computation—and the access they have offered to neuronal operations underlying cognition.

The current state of the field and its promise of riches untapped can be summarized through a survey of the processes involved in the acquisition, storage, and use of information by the nervous system: sensation, perception, decision formation, motor control, memory, language, emotions, and consciousness.

Sensation

We acquire knowledge of the world through our senses. Not surprisingly, sensory processes are among the most thoroughly studied in cognitive neuroscience. Systematic explorations of these processes originated in two domains. The first consisted of investigations of the physical nature of the sensory stimuli in question, such as the wave nature of light and sound. Sir Isaac Newton's (1642–1727) Optiks is an exemplar of this approach. The second involved studies of the anatomy of the peripheral sense organs, with attention given to the manner in which anatomical features prepared the physical stimulus for sensory transduction. Von Bekesy's beautiful studies of the structural features of the cochlea and the relation of those features to the neuronal frequency coding of sound is a classic example (for which he was awarded the 1961 Nobel Prize in physiology and medicine). Our present understanding of the neuronal bases of sensation was further enabled by three major developments: (1) establishment of the neuron doctrine, with attendant anatomical and physiological studies of neurons; (2) systematization of behavioral studies of sensation, made possible through the development of psychophysics; and (3) advancement of sophisticated theories of neuronal function, as embodied by the discipline of COMPUTATIONAL NEUROSCIENCE.

For a variety of reasons, vision has emerged as the model for studies of sensory processing, although many fundamental principles of sensory processing are conserved across modalities. Initial acquisition of information about the world, by all sensory modalities, begins with a process known as transduction, by which forms of physical energy (e.g., photons) alter the electrical state of a sensory neuron. In the case of vision,

phototransduction occurs in the RETINA, which is a specialized sheet-like neural network with a regular repeating structure. In addition to its role in transduction, the retina also functions in the initial detection of spatial and temporal contrast (Enroth-Cugell and Robson 1966; Kaplan and Shapley 1986) and contains specialized neurons that subserve COLOR VISION (see also COLOR, NEUROPHYSIOLOGY OF). The outputs of the retina are carried by a variety of ganglion cell types to several distinct termination sites in the central nervous system. One of the largest projections forms the "geniculostriate" pathway, which is known to be critical for normal visual function in primates. This pathway ascends to the cerebral cortex by way of the lateral geniculate nucleus of the THALAMUS. The cerebral cortex itself has been a major focus of study during the past forty years of vision research (and sensory research of all types). The entry point for ascending visual information is via primary visual cortex, otherwise known as striate cortex or area V1, which lies on the posterior pole (the occipital lobe) of the cerebral cortex in primates. The pioneering studies of V1 by Hubel and Wiesel (1977) established the form in which visual information is represented by the activity of single neurons and the spatial arrangement of these representations within the cortical mantle ("functional architecture"). With the development of increasingly sophisticated techniques, our understanding of cortical VISUAL ANATOMY AND PHYSIOLOGY, and their relationships to sensory experience, has been refined considerably. Several general principles have emerged:

Receptive Field

This is an operationally defined attribute of a sensory neuron, originally offered by the physiologist Haldan Keffer Hartline, which refers to the portion of the sensory field that, when stimulated, elicits a change in the electrical state of the cell. More generally, the receptive field is a characterization of the filter properties of a sensory neuron, which are commonly multidimensional and include selectivity for parameters such as spatial position, intensity, and frequency of the physical stimulus. Receptive field characteristics thus contribute to an understanding of the information represented by the brain, and are often cited as evidence for the role of a neuron in specific perceptual and cognitive functions.

Contrast Detection

The elemental sensory operation, that is, one carried out by all receptive fields—is detection of spatial or temporal variation in the incoming signal. It goes without saying that if there are no environmental changes over space and time, then nothing in the input is worthy of detection. Indeed, under such constant conditions sensory neurons quickly adapt. The result is a demonstrable loss of sensation—such as "snow blindness"—that occurs even though there may be energy continually impinging on the receptor surface. On the other hand, contrast along some sensory dimension indicates a change in the environment, which may in turn be a call for action. All sensory modalities have evolved mechanisms for detection of such changes.

Topographic Organization

Representation of spatial patterns of activation within a sensory field is a key feature of visual, auditory, and tactile senses, which serves the behavioral goals of locomotor navigation and object recognition. Such representations are achieved for these modalities, in part, by topographically organized neuronal maps. In the visual system, for example, the retinal projection onto the lateral geniculate nucleus of the thalamus possesses a high degree of spatial order, such that neurons with spatially adjacent receptive fields lie adjacent to one another in the brain. Similar visuotopic maps are seen in primary visual cortex and in several successively higher levels of processing (e.g., Gattass, Sousa, and Covey 1985). These maps are commonly distorted relative to the sensory field, such that, in the case of vision, the numbers of neurons representing the central portion of the visual field greatly exceed those representing the visual periphery. These variations in "magnification factor" coincide with (and presumably underlie) variations in the observer's resolving power and sensitivity.

Modular and Columnar Organization

The proposal that COLUMNS AND MODULES form the basis for functional organization in the sensory neocortex is a natural extension of the nineteenth-century concept of localization of function. The 1970s and 1980s saw a dramatic rise in the use of electrophysiological and anatomical tools to subdivide sensory cortices—particularly visual cortex—into distinct functional modules. At the present time, evidence indicates that the visual cortex of monkeys is composed of over thirty such regions, including the well-known and heavily studied areas V1, V2, V3, V4, MT, and IT, as well as some rather more obscure and equivocal designations (Felleman and Van Essen 1991). These efforts to reveal order in heterogeneity have been reinforced by the appealing computational view (e.g., Marr 1982) that larger operations (such as seeing) can be subdivided and assigned to dedicated taskspecific modules (such as ones devoted to visual motion or color processing, for example). The latter argument also dovetails nicely with the nineteenth-century concept of elementism, the coincidence of which inspired a fevered effort to identify visual areas that process specific sensory "elements." Although this view appears to be supported by physiological evidence for specialized response properties in some visual areas—such as a preponderance of motion-sensitive neurons in area MT (Albright 1993) and color-sensitive neurons in area V4 (Schein and Desimone 1990)—the truth is that very little is yet known of the unique contributions of most other cortical visual areas.

Modular organization of sensory cortex also occurs at a finer spatial scale, in the form of regional variations in neuronal response properties and anatomical connections, which are commonly referred to as columns, patches, blobs, and stripes. The existence of a column-like anatomical substructure in the cerebral cortex has been known since the early twentieth century, following the work of Ramón y Cajal, Constantin von Economo (1876–1931), and Rafael Lorente de Nó. It was the latter who first suggested that this characteristic structure may have some functional significance (Lorento de Nó 1938). The concept of modular functional organization was later expanded upon by the physiologist Vernon B. Mountcastle (1957), who obtained the first evidence for columnar function

through his investigations of the primate somatosensory system, and offered this as a general principle of cortical organization. The most well known examples of modular organization of the sort predicted by Mountcastle are the columnar systems for contour orientation and ocular dominance discovered in primary visual cortex in the 1960s by David Hubel and Torsten Wiesel (1968).

Additional evidence for functional columns and for the veracity of Mountcastle's dictum has come from studies of higher visual areas, such as area MT (Albright, Desimone, and Gross 1984) and the inferior temporal cortex (Tanaka 1997). Other investigations have demonstrated that modular representations are not limited to strict columnar forms (Born and Tootell 1993; Livingstone and Hubel 1984) and can exist as relatively large cortical zones in which there is a common feature to the neuronal representation of sensory information (such as clusters of cells that exhibit a greater degree of selectivity for color, for example). The high incidence of columnar structures leads one to wonder why they exist. One line of argument, implicit in Mountcastle's original hypothesis, is based on the need for adequate "coverage"—that is, nesting the representation of one variable (such as preferred orientation of a visual contour) across changes in another (such as the topographic representation of the visual field)—which makes good computational sense and has received considerable empirical support (Hubel and Wiesel 1977). Other arguments include those based on developmental constraints (Swindale 1980; Miller 1994; Goodhill 1997) and computational advantages afforded by representation of sensory features in a regular periodic structure (see COMPUTATIONAL NEUROANATOMY; Schwartz 1980).

Hierarchical Processing

A consistent organizational feature of sensory systems is the presence of multiple hierarchically organized processing stages, through which incoming sensory information is represented in increasingly complex or abstract forms. The existence of multiple stages has been demonstrated by anatomical studies, and the nature of the representation at each stage has commonly been revealed through electrophysiological analysis of sensory response properties. As we have seen for the visual system, the first stage of processing beyond transduction of the physical stimulus is one in which a simple abstraction of light intensity is rendered, namely a representation of luminance contrast. Likewise, the outcome of processing in primary visual cortex is, in part, a representation of image contours—formed, it is believed, by lx Neurosciences a convergence of inputs from contrast-detecting neurons at earlier stages (Hubel and Wiesel 1962). At successively higher stages of processing, information is combined to form representations of even greater complexity, such that, for example, at the pinnacle of the pathway for visual pattern processing—a visual area known as inferior temporal (IT) cortex—individual neurons encode complex, behaviorally significant objects, such as faces (see FACE RECOGNITION).

Parallel Processing

In addition to multiple serial processing stages, the visual system is known to be organized in parallel streams. Incoming information of different types is channeled through a variety of VISUAL PROCESSING STREAMS, such that the output of each serves a unique function. This type of channeling occurs on several scales, the grossest of which is manifested as multiple retinal projections (typically six) to different brain regions. As we have noted, it is the geniculostriate projection that serves pattern vision in mammals. The similarly massive retinal projection to the midbrain superior colliculus (the "tectofugal" pathway) is known to play a role in orienting responses, OCULOMOTOR CONTROL, and MULTISENSORY INTEGRATION. Other pathways include a retinal projection to the hypothalamus, which contributes to the entrainment of circadian rhythms by natural light cycles. Finer scale channeling of visual information is also known to exist, particularly in the case of the geniculostriate pathway (Shapley 1990). Both anatomical and physiological evidence (Perry, Oehler, and Cowey 1984; Kaplan and Shapley 1986) from early stages of visual processing support the existence of at least three subdivisions of this pathway, known as parvocellular, magnocellular, and the more recently identified koniocellular (Hendry and Yoshioka 1994). Each of these subdivisions is known to convey a unique spectrum of retinal image information and to maintain that information in a largely segregated form at least as far into the system as primary visual cortex (Livingstone and Hubel 1988).

Beyond V1, the ascending anatomical projections fall into two distinct streams, one of which descends ventrally into the temporal lobe, while the other courses dorsally to the parietal lobe. Analyses of the behavioral effects of lesions, as well as electrophysiological studies of neuronal response properties, have led to the hypothesis (Ungerleider and Mishkin 1982) that the ventral stream represents information about form and the properties of visual surfaces (such as their color or TEXTURE)—and is thus termed the "what" pathway—while the dorsal stream represents information regarding motion, distance, and the spatial relations between environmental surfaces—the so-called "where" pathway. The precise relationship, if any, between the early-stage channels (magno, parvo, and konio) and these higher cortical streams has been a rich source of debate and controversy over the past decade, and the answers remain far from clear (Livingstone and Hubel 1988; Merigan and Maunsell 1993).

Perception

Perception reflects the ability to derive meaning from sensory experience, in the form of information about structure and causality in the perceiver's environment, and of the sort necessary to guide behavior. Operationally, we can distinguish sensation from perception by the nature of the internal representations: the former encode the physical properties of the proximal sensory stimulus (the retinal image, in the case of vision), and the latter reflect the world that likely gave rise to the sensory stimulus (the visual scene). Because the mapping between sensory and perceptual events is never unique—multiple scenes can cause the same retinal image—

perception is necessarily an inference about the probable causes of sensation. As we have seen, the standard approach to understanding the information represented by sensory neurons, which has evolved over the past fifty years, is to measure the correlation between a feature of the neuronal response (typically magnitude) and some physical parameter of a sensory stimulus (such as the wavelength of light or the orientation of a contour). Because the perceptual interpretation of a sensory event is necessarily context-dependent, this approach alone is capable of revealing little, if anything, about the relationship between neuronal events and perceptual state. There are, however, some basic variations on this approach that have led to increased understanding of the neuronal bases of perception.

Experimental Approaches to the Neuronal Bases of Perception

Origins of a Neuron Doctrine for Perceptual Psychology

The first strategy involves evaluation of neuronal responses to visual stimuli that consist of complex objects of behavioral significance. The logic behind this approach is that if neurons are found to be selective for such stimuli, they may be best viewed as representing something of perceptual meaning rather than merely coincidentally selective for the collection of sensory features. The early studies of "bug detectors" in the frog visual system by Lettvin and colleagues (Lettvin, Maturana, MCCULLOCH, and PITTS 1959) exemplify this approach and have led to fully articulated views on the subject, including the concept of the "gnostic unit" advanced by Jerzy Konorski (1967) and the "cardinal cell" hypothesis from Barlow's (1972) classic "Neuron Doctrine for Perceptual Psychology." Additional evidence in support of this concept came from the work of Charles Gross in the 1960s and 1970s, in the extraordinary form of cortical cells selective for faces and hands (Gross, Bender, and Rocha-Miranda 1969; Desimone et al. 1984). Although the suggestion that perceptual experience may be rooted in the activity of single neurons or small neuronal ensembles has been decried, in part, on the grounds that the number of possible percepts greatly exceeds the number of available neurons, and is often ridiculed as the "grandmother-cell" hypothesis, the evidence supporting neuronal representations for visual patterns of paramount behavioral significance, such as faces, is now considerable (Desimone 1991; Rolls 1992).

Although a step in the right direction, the problem with this general approach is that it relies heavily upon assumptions about how the represented information is used. If a cell is activated by a face, and only a face, then it seems likely that the cell contributes directly to the perceptually meaningful experience of face recognition rather than simply representing a collection of sensory features (Desimone et al. 1984). To some, that distinction is unsatisfactorily vague, and it is, in any case, impossible to prove that a cell only responds to a face. An alternative approach that has proved quite successful in recent years is one in which an effort is made to directly relate neuronal and perceptual events.

Neuronal Discriminability

Predicts Perceptual Discriminability In the last quarter of the twentieth century, the marriage of single-neuron recording with visual psychophysics has yielded one of the dominant experimental paradigms of cognitive neuroscience, through which it has become possible to explain behavioral performance on a perceptual task in terms of the discriminative capacity of sensory neurons. The earliest effort of this type was a study of tactile discrimination conducted by Vernon Mountcastle in the 1960s (Mountcastle et al. 1967). In this study, thresholds for behavioral discrimination performance were directly compared to neuronal thresholds for the same stimulus set. A later study by Tolhurst, Movshon, and Dean (1983) introduced techniques from SIGNAL DETECTION THEORY that allowed more rigorous quantification of the discriminative capacity of neurons and thus facilitated neuronalperceptual comparisons. Several other studies over the past ten years have significantly advanced this cause (e.g., Dobkins and Albright 1995), but the most direct approach has been that adopted by William Newsome and colleagues (e.g., Newsome, Britten, and Movshon 1989). In this paradigm, behavioral and neuronal events are measured simultaneously in response to a sensory stimulus, yielding by brute force some of the strongest evidence to date for neural substrates of perceptual discriminability.

Decoupling Sensation and Perception A somewhat subtler approach has been forged by exploiting the natural ambiguity between sensory events and perceptual experience (see ILLUSIONS). This ambiguity is manifested in two general forms: (1) single sensory events that elicit multiple distinct percepts, a phenomenon commonly known as “perceptual metastability,” and (2) multiple sensory events—”sensory synonyms”— that elicit the same perceptual state. Both of these situations, which are ubiquitous in normal experience, afford opportunities to experimentally decouple sensation and perception.

The first form of sensory-perceptual ambiguity (perceptual metastability) is a natural consequence of the indeterminate mapping between a sensory signal and the physical events that gave rise to it. A classic and familiar example is the Necker Cube, in which the three-dimensional interpretation—the observer's inference about visual scene structure—periodically reverses despite the fact that the retinal image remains unchanged. Logothetis and colleagues (Logothetis and Schall 1989) have used a form of perceptual metastability known as binocular rivalry to demonstrate the existence of classes of cortical neurons that parallel changes in perceptual state in the face of constant retinal inputs.

The second type of sensory-perceptual ambiguity, in which multiple sensory images give rise to the same percept, is perhaps the more common. Such effects are termed perceptual constancies, and they reflect efforts by sensory systems to reconstruct behaviorally significant attributes of the world in the face of variation along irrelevant sensory dimensions. Size constancy—the invariance of perceived size of an object across different retinal sizes—and brightness or color constancy—the invariance of perceived reflectance or color of a surface in the presence of illumination changes—are classic examples. These perceptual constancies suggest an underlying neuronal invariance across specific image changes. Several examples of neuronal constancies have been reported, including invariant representations of

direction of motion and shape across different cues for form (Albright 1992; Sary et al. 1995).

Contextual Influences on Perception and its Neuronal Bases One of the most promising new approaches to the neuronal bases of perception is founded on the use of contextual manipulations to influence the perceptual interpretation of an image feature. As we have seen, the contextual dependence of perception is scarcely a new finding, but contextual manipulations have been explicitly avoided in traditional physiological approaches to sensory coding. As a consequence, most existing data do not reveal whether and to what extent the neuronal representation of an image feature is context dependent. Gene Stoner, Thomas Albright, and colleagues have pioneered the use of contextual manipulations in studies of the neuronal basis of the PERCEPTION OF MOTION (e.g., Stoner and Albright 1992, 1993). The results of these studies demonstrate that context can alter neuronal filter properties in a manner that predictably parallels its influence on perception.

Stages of Perceptual Representation

Several lines of evidence suggest that there may be multiple steps along the path to extracting meaning from sensory signals. These steps are best illustrated by examples drawn from studies of visual processing. Sensation itself is commonly identified with “early” or “low-level vision.” Additional steps are as follows. Mid-Level Vision This step involves a reconstruction of the spatial relationships between environmental surfaces. It is implicit in the accounts of the perceptual psychologist JAMES JEROME GIBSON (1904–1979), present in the computational approach of DAVID MARR (1945–1980), and encompassed by what has recently come to be known as MID-LEVEL VISION. Essential features of this processing stage include a dependence upon proximal sensory context to establish surface relationships (see SURFACE PERCEPTION) and a relative lack of dependence upon prior experience. By establishing environmental STRUCTURE FROM VISUAL INFORMATION SOURCES, midlevel vision thus invests sensory events with some measure of meaning. A clear example of this type of visual processing is found in the phenomenon of perceptual TRANSPARENCY (Metelli 1974) and the related topic of LIGHTNESS PERCEPTION.

Physiological studies of the response properties of neurons at mid-levels of the cortical hierarchy have yielded results consistent with a mid-level representation (e.g., Stoner and Albright 1992). High-Level Vision HIGH-LEVEL VISION is a loosely defined processing stage, but one that includes a broad leap in the assignment of meaning to sensory events—namely identification and classification on the basis of previous experience with the world. It is through this process that recognition of objects occurs (see OBJECT RECOGNITION, HUMAN NEUROPSYCHOLOGY; OBJECT RECOGNITION, ANIMAL STUDIES; and VISUAL OBJECT RECOGNITION, AI), as well as assignment of affect and semantic categorization. This stage thus constitutes a bridge between sensory processing and MEMORY. Physiological and neuropsychological studies of

the primate temporal lobe have demonstrated an essential contribution of this region to object recognition (Gross 1973;Gross et al. 1985).

Sensory-Perceptual Plasticity

The processes by which information is acquired and interpreted by the brain are modifiable throughout life and on many time scales. Although plasticity of the sort that occurs during brain development and that which underlies changes in the sensitivity of mature sensory systems may arise from similar mechanisms, it is convenient to consider them separately.

Developmental Changes

The development of the mammalian nervous system is a complex, multistaged process that extends from embryogenesis through early postnatal life. This process begins with determination of the fate of precursor cells such that a subset becomes neurons. This is followed by cell division and proliferation, and by differentiation of cells into different types of neurons. The patterned brain then begins to take shape as cells migrate to destinations appropriate for their assigned functions. Finally, neurons begin to extend processes and to make synaptic connections with one another. These connections are sculpted and pruned over a lengthy postnatal period. A central tenet of modern neuroscience is that these final stages of NEURAL DEVELOPMENT correspond to specific stages of COGNITIVE DEVELOPMENT. These stages are known as “critical periods,” and they are characterized by an extraordinary degree of plasticity in the formation of connections and cognitive functions.

Although critical periods for development are known to exist for a wide range of cognitive functions such as sensory processing, motor control, and language, they have been studied most intensively in the context of the mammalian visual system. These studies have included investigations of the timing, necessary conditions for, and mechanisms of (1) PERCEPTUAL DEVELOPMENT (e.g., Teller 1997), (2) formation of appropriate anatomical connections (e.g., Katz and Shatz 1996), and (3) neuronal representations of sensory stimuli (e.g., Hubel, Wiesel, and LeVay 1977). The general view that has emerged is that the newborn brain possesses a considerable degree of order, but that sensory experience is essential during critical periods to maintain that order and to fine-tune it to achieve optimal performance in adulthood. These principles obviously have profound implications for clinical practice and social policy. Efforts to further understand the cellular mechanisms of developmental plasticity, their relevance to other facets of cognitive function, the relative contributions of genes and experience, and routes of clinical intervention, are all among the most important topics for the future of cognitive neuroscience.

Dynamic Control of Sensitivity in the Mature Brain

Mature sensory systems have limited information processing capacities. An exciting area of research in recent years has been that addressing the conditions under which processing capacity is dynamically reallocated,

resulting in fluctuations in sensitivity to sensory stimuli. The characteristics of sensitivity changes are many and varied, but all serve to optimize acquisition of information in a world in which environmental features and behavioral goals are constantly in flux. The form of these changes may be broad in scope or highly stimulus-specific and task-dependent. Changes may be nearly instantaneous, or they may come about gradually through exposure to specific environmental features. Finally, sensitivity changes differ greatly in the degree to which they are influenced by stored information about the environment and the degree to which they are under voluntary control.

Studies of the visual system reveal at least three types of sensitivity changes represented by the phenomena of (1) contrast gain control, (2) attention, and (3) perceptual learning. All can be viewed as recalibration of incoming signals to compensate for changes in the environment, the fidelity of signal detection (such as that associated with normal aging or trauma to the sensory periphery), and behavioral goals. Generally speaking, neuronal gain control is the process by which the sensitivity of a neuron (or neural system) to its inputs is dynamically controlled. In that sense, all of the forms of adult plasticity discussed below are examples of gain control, although they have different dynamics and serve different functions.

Contrast Gain Control A well-studied example of gain control is the invariance of perceptual sensitivity to the features of the visual world over an enormous range of lighting conditions. Evidence indicates that the limited dynamic range of responsivity of individual neurons in visual cortex is adjusted in an illumination-dependent manner (Shapley and Victor 1979), the consequence of which is a neuronal invariance that can account for the sensory invariance. It has been suggested that this scaling of neuronal sensitivity as a function of lighting conditions may be achieved by response "normalization," in which the output of a cortical neuron is effectively divided by the pooled activity of a large number of other cells of the same type (Carandini, Heeger, and Movshon 1997).

Attention Visual ATTENTION is, by definition, a rapidly occurring change in visual sensitivity that is selective for a specific location in space or specific stimulus features. The stimulus and mnemonic factors that influence attentional allocation have been studied for over a century (James 1890), and the underlying brain structures and events are beginning to be understood (Desimone and Duncan 1995). Much of our understanding comes from analysis of ATTENTION IN THE HUMAN BRAIN—particularly the effects of cortical lesions, which can selectively interfere with attentional allocation (VISUAL NEGLECT), and through electrical and magnetic recording (ERP, MEG) and imaging studies—POSITRON EMISSION TOMOGRAPHY (PET) and functional MAGNETIC RESONANCE IMAGING (fMRI). In addition, studies of ATTENTION IN THE ANIMAL BRAIN have revealed that attentional shifts are correlated with changes in the sensitivity of single neurons to sensory stimuli (Moran and Desimone 1985; Bushnell, Goldberg, and Robinson 1981; see also AUDITORY ATTENTION). Although attentional phenomena differ from contrast gain control in that they can be influenced by feedback WORKING MEMORY as well as feedforward (sensory) signals, attentional effects can

also be characterized as an expansion of the dynamic range of sensitivity, but in a manner that is selective for the attended stimuli.

Perceptual Learning

Both contrast gain control and visual attention are rapidly occurring and short-lived sensitivity changes. Other experiments have targeted neuronal events that parallel visual sensitivity changes occurring over a longer time scale, such as those associated with the phenomenon of perceptual learning. Perceptual learning refers to improvements in discriminability along any of a variety of sensory dimensions that come with practice. Although it has long been known that the sensitivity of the visual system is refined in this manner during critical periods of neuronal development, recent experiments have provided tantalizing evidence of improvements in the sensitivity of neurons at early stages of processing, which parallel perceptual learning in adults (Recanzone, Schreiner, and Merzenich 1993; Gilbert 1996).

Forming a Decision to Act

The meaning of many sensations can be found solely in their symbolic and experiencedependent mapping onto actions (e.g., green = go, red = stop). These mappings are commonly many-to-one or one-to-many (a whistle and a green light can both be signals to “go”; conversely, a whistle may be either a signal to “go” or a call to attention, depending upon the context). The selection of a particular action from those possible at any point in time is thus a context-dependent transition between sensory processing and motor control. This transition is commonly termed the decision stage, and it has become a focus of recent electrophysiological studies of the cerebral cortex (e.g., Shadlen and Newsome 1996). Because of the nonunique mappings, neurons involved in making such decisions should be distinguishable from those representing sensory events by a tendency to generalize across specific features of the sensory signal. Similarly, the representation of the neuronal decision should be distinguishable from a motor control signal by generalization across specific motor actions. In addition, the strength of the neuronal decision signal should increase with duration of exposure to the sensory stimulus (integration time), in parallel with increasing decision confidence on the part of the observer. New data in support of some of these predictions suggests that this may be a valuable new paradigm for accessing the neuronal substrates of internal cognitive states, and for bridging studies of sensory or perceptual processing, memory, and motor control.

Motor Control

Incoming sensory information ultimately leads to action, and actions, in turn, are often initiated in order to acquire additional sensory information. Although MOTOR CONTROL systems have often been studied in relative isolation from sensory processes, this sensory-motor loop suggests that they are best viewed as different phases of a processing continuum. This integrated view, which seeks to understand how the nature of sensory representations influences movements, and vice-versa, is rapidly gaining

acceptance. The oculomotor control system has become the model for the study of motor processes at behavioral and neuronal levels. Important research topics that have emerged from consideration of the transition from sensory processing to motor control include (1) the process by which representations of space (see SPATIAL PERCEPTION) are transformed from the coordinate system of the sensory field (e.g., retinal space) to a coordinate system for action (e.g., Graziano and Gross 1998) and (2) the processes by which the neuronal links between sensation and action are modifiable (Raymond, Lisberger, and Mauk 1996), as needed to permit MOTOR LEARNING and to compensate for degenerative sensory changes or structural changes in the motor apparatus.

The brain structures involved in motor control include portions of the cerebral cortex, which are thought to contribute to fine voluntary motor control, as well as the BASAL GANGLIA and CEREBELLUM, which play important roles in motor learning; the superior colliculus, which is involved in sensorimotor integration, orienting responses, and oculomotor control; and a variety of brainstem motor nuclei, which convey motor signals to the appropriate effectors.

Learning and Memory

Studies of the neuronal mechanisms that enable information about the world to be stored and retrieved for later use have a long and rich history—being, as they were, a central part of the agenda of the early functional localizationists—and now lie at the core of our modern cognitive neuroscience. Indeed, memory serves as the linchpin that binds and shapes nearly every aspect of information processing by brains, including perception, decision making, motor control, emotion, and consciousness. Memory also exists in various forms, which have been classified on the basis of their relation to other cognitive functions, the degree to which they are explicitly encoded and available for use in a broad range of contexts, and their longevity. (We have already considered some forms of nonexplicit memory, such as those associated with perceptual and motor learning.) Taxonomies based upon these criteria have been reviewed in detail elsewhere (e.g., Squire, Knowlton, and Musen 1993). The phenomenological and functional differences among different forms of memory suggest the existence of a variety of different brain substrates. Localization of these substrates is a major goal of modern cognitive neuroscience. Research is also clarifying the mechanisms underlying the oft-noted role of affective or emotional responses in memory consolidation (see MEMORY STORAGE, MODULATION OF; AMYGDALA, PRIMATE), and the loss of memory that occurs with aging (see AGING, MEMORY, AND THE BRAIN). Three current approaches (broadly defined and overlapping) to memory are among the most promising for the future of cognitive neuroscience: (1) neuropsychological and neurophysiological studies of the neuronal substrates of explicit memory in primates, (2) studies of the relationship between phenomena of synaptic facilitation or depression and behavioral manifestations of learning and memory, and (3) molecular genetic studies that enable highly selective disruption of cellular structures and events thought to be involved in learning and memory.

Brain Substrates of Explicit Memory in Primates

The current approach to this topic has its origins in the early studies of Karl Lashley and colleagues, in which the lesion method was used to infer the contributions of specific brain regions to a variety of cognitive functions, including memory. The field took a giant step forward in the 1950s with the discovery by Brenda Milner and colleagues of the devastating effects of damage to the human temporal lobe—particularly the HIPPOCAMPUS—on human memory formation (see MEMORY, HUMAN NEUROPSYCHOLOGY).

Following that discovery, Mortimer Mishkin and colleagues began to use the lesion technique to develop an animal model of amnesia. More recently, using a similar approach, Stuart Zola, Larry Squire, and colleagues have further localized the neuronal substrates of memory consolidation in the primate temporal lobe (see MEMORY, ANIMAL STUDIES). Electrophysiological studies of the contributions of individual cortical neurons to memory began in the 1970s with the work of Charles Gross and Joaquin Fuster. The logic behind this approach is that by examining neuronal responses of an animal engaged in a standard memory task (e.g., match-to-sample: determine whether a sample stimulus corresponds to a previously viewed cue stimulus), one can distinguish the components of the response that reflect memory from those that are sensory in nature.Subsequent electrophysiological studies by Robert Desimone and Patricia Goldman-Rakic, among others, have provided some of the strongest evidence for single-cell substrates of working memory in the primate temporal and frontal lobes. These traditional approaches to explicit memory formation in primates are now being complemented by brain imaging studies in humans.

Do Synaptic Changes Mediate Memory Formation?

The phenomenon of LONG-TERM POTENTIATION (LTP), originally discovered in the 1970s—and the related phenomenon of long-term depression—consists of physiologically measurable changes in the strength of synaptic connections between neurons. LTP is commonly produced in the laboratory by coincident activation of pre- and post-synaptic neurons, in a manner consistent with the predictions of DONALD O. HEBB (1904–1985), and it is often dependent upon activation of the postsynaptic NMDA glutamate receptor. Because a change in synaptic efficacy could, in principle, underlie behavioral manifestations of learning and memory, and because LTP is commonly seen in brain structures that have been implicated in memory formation (such as the hippocampus, cerebellum, and cerebral cortex) by other evidence, it is considered a likely mechanism for memory formation. Attempts to test that hypothesis have led to one of the most exciting new approaches to memory.

From Genes to Behavior: A Molecular Genetic Approach to Memory

The knowledge that the NMDA receptor is responsible for many forms of LTP, in conjunction with the hypothesis that LTP underlies memory formation, led to the prediction that memory formation should be disrupted

by elimination of NMDA receptors. The latter can be accomplished in mice by engineering genetic mutations that selectively knock out the NMDA receptor, although this technique has been problematic because it has been difficult to constrain the effects to specific brain regions and over specific periods of time. Matthew Wilson and Susumu Tonegawa have recently overcome these obstacles by production of a knockout in which NMDA receptors are disrupted only in a subregion of the hippocampus (the CA1 layer), and only after the brain has matured. In accordance with the NMDA-mediated synaptic plasticity hypothesis, these animals were deficient on both behavioral and physiological assays of memory formation (Tonegawa et al. 1996). Further developments along these lines will surely involve the ability to selectively disrupt action potential generation in specific cell populations, as well as genetic manipulations in other animals (such as monkeys).

Language

One of the first cognitive functions to be characterized from a biological perspective was language. Nineteenth-century physicians, including Broca, observed the effects of damage to different brain regions and described the asymmetrical roles of the left and right hemispheres in language production and comprehension (see HEMISPHERIC SPECIALIZATION; APHASIA; LANGUAGE, NEURAL BASIS OF). Investigators since then have discovered that different aspects of language, including the PHONOLOGY, SYNTAX, and LEXICON, each rely on different and specific neural structures (see PHONOLOGY, NEURAL BASIS OF; GRAMMAR, NEURAL BASIS OF; LEXICON, NEURAL BASIS OF).

Modern neuroimaging techniques, including ERPs, PET, and fMRI, have confirmed the role of the classically defined language areas and point to the contribution of several other areas as well. Such studies have also identified "modality neutral" areas that are active when language is processed through any modality: auditory, written, and even sign language (see SIGN LANGUAGE AND THE BRAIN). Studies describing the effects of lesions on language can identify neural tissue that is necessary and sufficient for processing. An important additional perspective can be obtained from neuroimaging studies of healthy neural tissue, which can reveal all the activity associated with language production and comprehension. Taken together the currently available evidence reveals a strong bias for areas within the left hemisphere to mediate language if learned early in childhood, independently of its form or modality. However, the nature of the language learned and the age of acquisition have effects on the configuration of the language systems of the brain (see BILINGUALISM AND THE BRAIN).

Developmental disorders of language (see LANGUAGE IMPAIRMENT, DEVELOPMENTAL; DYSLEXIA) can occur in isolation or in association with other disorders and can result from deficits within any of the several different skills that are central to the perception and modulation of language.

 See also APHASIA; BILINGUALISM AND THE BRAIN; DYSLEXIA; GRAMMAR, NEURAL BASIS OF; HEMISPHERIC SPECIALIZATION; LANGUAGE, NEURAL BASIS OF; LANGUAGE IMPAIRMENT, DEVELOPMENTAL; LEXICON;

LEXICON, NEURAL BASIS OF; PHONOLOGY; PHONOLOGY, NEURAL BASIS OF; SIGN LANGUAGE AND THE BRAIN; SYNTAX

Consciousness

Rediscovery of the phenomena of perception and memory without awareness has renewed research and debate on issues concerning the neural basis of CONSCIOUSNESS (see CONSCIOUSNESS, NEUROBIOLOGY OF). Some patients with cortical lesions that have rendered them blind can nonetheless indicate (by nonverbal methods) accurate perception of stimuli presented to the blind portion of the visual field (see BLINDSIGHT). Similarly, some patients who report no memory for specific training events nonetheless demonstrate normal learning of those skills. Systematic study of visual consciousness employing several neuroimaging tools within human and nonhuman primates is being conducted to determine whether consciousness emerges as a property of a large collection of interacting neurons or whether it arises as a function of unique neuronal characteristics possessed by some neurons or by an activity pattern temporarily occurring within a subset of neurons (see BINDING BY NEURAL SYNCHRONY). Powerful insights into systems and cellular and molecular events critical in cognition and awareness, judgment and action have come from human and animal studies of SLEEP and DREAMING. Distinct neuromodulatory effects of cholenergic and aminergic systems permit the panoply of conscious cognitive processing, evaluation, and planning during waking states and decouple cognition, emotional, and mnemonic functions during sleep. Detailed knowledge of the neurobiology of sleep and dreaming presents an important opportunity for future studies of cognition and consciousness.

Emotions

Closely related to questions about consciousness are issues of EMOTIONS and feelings that have, until very recently, been ignored in cognitive science. Emotions sit at the interface between incoming events and preparation to respond, however, and recent studies have placed the study of emotion more centrally in the field.

Animal models

have provided detailed anatomical and physiological descriptions of fear responses (Armony and LeDoux 1997) and highlight the role of the amygdala and LIMBIC SYSTEM as well as different inputs to this system (see EMOTION AND THE ANIMAL BRAIN). Studies of human patients suggest specific roles for different neural systems in the perception of potentially emotional stimuli (Adolphs et al. 1994; Hamann et al. 1996), in their appraisal, and in organizing appropriate responses to them (see EMOTION AND THE HUMAN BRAIN; PAIN). An important area for future research is to characterize the neurochemistry of emotions. The multiple physiological responses to real or imagined threats (i.e., STRESS) have been elucidated in both animal and human studies. Several of the systems most affected by stress play central roles in emotional and cognitive

functions (see NEUROENDOCRINOLOGY). Early pre- and postnatal experiences play a significant role in shaping the activity of these systems and in their rate of aging. The profound role of the stress-related hormones on memory-related brain structures, including the hippocampus, and their role in regulating neural damage following strokes and seizures and in aging, make them a central object for future research in cognitive neuroscience (see AGING AND COGNITION).

4 Cognitive Neuroscience: A Promise for the Future

A glance at the neuroscience entries for this volume reveals that we are amassing detailed knowledge of the highly specialized neural systems that mediate different and specific cognitive functions. Many questions remain unanswered, however, and the applications of new experimental techniques have often raised more questions than they have answered. But such are the expansion pains of a thriving science. Among the major research goals of the next century will be to elucidate how these highly differentiated cognitive systems arise in ontogeny, the degree to which they are maturationally constrained, and the nature and the timing of the role of input from the environment in NEURAL DEVELOPMENT. This is an area where research has just begun. It is evident that there exist strong genetic constraints on the overall patterning of different domains within the developing nervous system. Moreover, the same class of genes specify the rough segmentation of the nervous systems of both vertebrates and invertebrates. However, the information required to specify the fine differentiation and connectivity within the cortex exceeds that available in the genome. Instead, a process of selective stabilization of transiently redundant connections permits individual differences in activity and experience to organize developing cortical systems. Some brain circuits display redundant connectivity and pruning under experience only during a limited time period in development (“critical period”). These time periods are different for different species and for different functional brain systems within a species.

Other brain circuits retain the ability to change under external stimulation throughout life, and this capability, which now appears more ubiquitous and long lasting than initially imagined, is surely a substrate for adult learning, recovery of function after brain damage, and PHANTOM LIMB phenomena (see also AUDITORY PLASTICITY; NEURAL PLASTICITY). A major challenge for future generations of cognitive neuroscientists will be to characterize and account for the markedly different extents and timecourses of biological constraints and experience-dependent modifiability of the developing human brain. Though the pursuit may be ancient, consider these the halcyon days of cognitive neuroscience. As we cross the threshold of the millenium, look closely as the last veil begins to fall. And bear in mind that if cognitive neuroscience fulfills its grand promise, later editions of this volume may contain a section on history, into which all of the nonneuro cognitive science discussion will be swept.

References

Adolphs, R., D. Tranel, H. Damasio, and A. Damasio. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature 372: 669–672.

Albright, T. D. (1992). Form-cue invariant motion processing in primate visual cortex. Science 255: 1141–1143.

Albright, T. D. (1993). Cortical processing of visual motion. In J. Wallman and F. A. Miles, Eds.,Visual Motion and its Use in the Stabilization of Gaze. Amsterdam: Elsevier.

Albright, T. D., R. Desimone, and C. G. Gross. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. Journal of Neurophysiology 51: 16–31.

Armony, J. L., and J. E. LeDoux. (1997). How the brain processes emotional information. Annals of the New York Academy of Sciences 821: 259–270.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? Perception 1: 371–394.

Boring, E. G. (1950). A History of Experimental Psychology, 2nd ed. R. M. Elliott, Ed. New Jersey: Prentice-Hall.

Born, R. T., and R. B. Tootell. (1993). Segregation of global and local motion processing in primate middle temporal visual area. Nature 357: 497–499.

Bushnell, M. C., M. E. Goldberg, and D. L. Robinson. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. 1. Modulation in posterior parietal cortex related to selective visual attention. Journal of Neurophysiology 46(4): 755–772.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. Annual Review of Neuroscience 21: 47–74.

Carandini, M., D. J. Heeger, and J. A. Movshon. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. Journal of Neuroscience 17(21): 8621–8644.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. Journal of Cognitive Neuroscience 3: 1–7.

Desimone, R., T. D. Albright, C. G. Gross, and C. J. Bruce. (1984). Stimulus selective properties of inferior temporal neurons in the macaque. Journal of Neuroscience 8: 2051–2062.

Desimone, R., and J. Duncan. (1995). Neural mechanisms of selective visual attention. Annual Review of Neuroscience 18: 193–222.

Dobkins, K. R., and T. D. Albright. (1995). Behavioral and neural effects of chromatic isoluminance in the primate visual motion system. Visual Neuroscience 12: 321–332.

Enroth-Cugell, C., and J. G. Robson. (1966). The contrast sensitivity of retinal ganglion cells of the cat. Journal of Physiology (London) 187: 517–552.

Felleman, D. J., and D. C. Van Essen. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex 1: 1–47.

Gattass, R., A. P. B. Sousa, and E. Cowey. (1985). Cortical visual areas of the macaque: Possible substrates for pattern recognition mechanisms. In C. Chagas, R. Gattass, and C. Gross, Eds., Pattern Recognition Mechanisms. Vatican City: Pontifica Academia Scientiarum, pp. 1–17.

Gilbert, C. D. (1996). Learning and receptive field plasticity. Proceedings of the National Academy of Sciences USA 93: 10546–10547.

Goodhill, G. J. (1997). Stimulating issues in cortical map development. Trends in Neurosciences 20: 375–376.

Graziano, M. S., and C. G. Gross. (1998). Spatial maps for the control of movement. Current Opinion in Neurobiology 8: 195–201.

Gross, C. G. (1973). Visual functions of inferotemporal cortex. In H. Autrum, R. Jung, W. Lowenstein, D. McKay, and H.-L. Teuber, Eds., Handbook of Sensory Physiology, vol. 7, 3B. Berlin: Springer.

Gross, C. G. (1994a). How inferior temporal cortex became a visual area. Cerebral Cortex 5: 455–469.

Gross, C. G. (1994b). Hans-Lukas Teuber: A tribute. Cerebral Cortex 4: 451–454.

Gross, C. G. (1998). Brain, Vision, Memory: Tales in the History of Neuroscience. Cambridge, MA: MIT Press.

Gross, C. G., D. B. Bender, and C. E. Rocha-Miranda. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science 166: 1303–1306.

Gross, C. G., R. Desimone, T. D. Albright, and E. L. Schwartz. (1985). Inferior temporal cortex and pattern recognition. In C. Chagas, Ed., Study Group on Pattern Recognition Mechanisms. Vatican City: Pontifica Academia Scientiarum, pp. 179–200.

Hamann, S. B., L. Stefanacci, L. R. Squire, R. Adolphs, D. Tranel, H. Damasio, and A. Damasio. (1996). Recognizing facial emotion [letter]. Nature 379(6565): 497.

Hartline, H. K., H. G. Wagner, and E. F. MacNichol, Jr. (1952). The peripheral origin of nervous activity in the visual system. Cold Spring Harbor Symposium on Quantitative Biology 17: 125– 141.

Hendry, S. H., and T. Yoshioka. (1994). A neurochemically distinct third channel in the macaque dorsal lateral geniculate nucleus. Science 264(5158): 575–577.

Hubel, D. H., and T. N. Wiesel. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. Journal of Physiology 160: 106–154.

Hubel, D. H., and T. N. Wiesel. (1968). Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology 195: 215–243.

Hubel, D. H., and T. N. Wiesel. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. Proceedings of the Royal Society of London, Series B, Biological Sciences 198(1130): 1–59.

Hubel, D. H., T. N. Wiesel, and S. LeVay. (1977). Plasticity of ocular dominance columns in monkey striate cortex. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences 278: 377–409.

James, W. (1890). The Principles of Psychology, vol. 1. New York: Dover.

Kaplan, E., and R. M. Shapley. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. Proceedings of the National Academy of Sciences of the USA 83(8): 2755–2757.

Katz, L. C., and C. J. Shatz. (1996). Synaptic activity and the construction of cortical circuits. Science 274: 1133–1138.

Koffka, K. (1935). Principles of Gestalt Psychology. New York: Harcourt, Brace.

Konorski, J. (1967). Integrative Activity of the Brain. Chicago: University of Chicago Press.

Kuffler, S. W. (1953). Discharge patterns and functional organization of the mammalian retina. Journal of Neurophysiology 16: 37–68.

Lettvin, J. Y., H. R. Maturana, W. S. McCulloch, and W. H. Pitts. (1959). What the frog's eye tells the frog's brain. Proceedings of the Institute of Radio Engineers 47: 1940–1951.

Livingstone, M. S., and D. H. Hubel. (1984). Anatomy and physiology of a color system in the primate visual cortex. Journal of Neuroscience 4: 309–356.

Livingstone, M. S., and D. H. Hubel. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. Science 240: 740–749.

Logothetis, N. K., and J. D. Schall. (1989). Neuronal correlates of subjective visual perception. Science 245: 761–763.

Lorento de Nó, R. (1938). Cerebral cortex: Architecture, intracortical connections, motor projections.

In J. F. Fulton, Ed., Physiology of the Nervous System. New York: Oxford University Press, pp. 291–339.

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco: W. H. Freeman.

Merigan, W. H., and J. H. Maunsell. (1993). How parallel are the primate visual pathways? Annual Review of Neuroscience 16: 369–402.

Metelli, F. (1974). The perception of transparency. Scientific American 230(4): 90–98.

Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. Journal of Neuroscience 14: 409–441.

Moran, J., and R. Desimone. (1985). Selective attention gates visual processing in the extrastriate cortex. Science 229(4715): 782–784.

Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. Journal of Neurophysiology 20: 408–434.

Mountcastle, V. B., W. H. Talbot, I. Darian-Smith, and H. H. Kornhuber. (1967). Neural basis of the sense of flutter-vibration. Science 155(762): 597–600.

Newsome, W. T., K. H. Britten, and J. A. Movshon. (1989). Neuronal correlates of a perceptual decision. Nature 341: 52–54.

Perry, V. H., R. Oehler, and A. Cowey. (1984). Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in the macaque monkey. Neuroscience 12(4): 1101–1123.

Raymond, J. L., S. G. Lisberger, and M. D. Mauk. (1996). The cerebellum: A neuronal learning machine? Science 272: 1126–1131.

Recanzone, G. H., C. E. Schreiner, and M. M. Merzenich. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. Journal of Neuroscience 13: 87–103.

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. In V. Bruce, A. Cowey, W. W. Ellis, and D. I. Perrett, Eds., Processing the Facial Image. Oxford: Clarendon Press, pp. 11–21.

Sary, G., R. Vogels, G. Kovacs, and G. A. Orban. (1995). Responses of monkey inferior temporal neurons to luminance-, motion-, and texture-defined gratings. Journal of Neurophysiology 73: 1341–1354.

Schein, S. J., and R. Desimone. (1990). Spectral properties of V4 neurons in the macaque. Journal of Neuroscience 10: 3369–3389.

Schwartz, E. L. (1980). Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. Vision Research 20: 645–669.

Shadlen, M. N., and W. T. Newsome. (1996). Motion perception: Seeing and deciding. Proceedings of the National Academy of Sciences 93: 628–633.

Shapley, R. (1990). Visual sensitivity and parallel retinocortical channels. Annual Review of Psychology 41: 635–658.

Shapley, R. M., and J. D. Victor. (1979). Nonlinear spatial summation and the contrast gain control of cat retinal ganglion cells. Journal of Physiology (London) 290: 141–161.

Squire, L. R., B. Knowlton, and G. Musen. (1993). The structure and organization of memory. Annual Review of Psychology 44: 453–495.

Stoner, G. R., and T. D. Albright. (1992). Neural correlates of perceptual motion coherence. Nature 358: 412–414.

Stoner, G. R., and T. D. Albright. (1993). Image segmentation cues in motion processing: Implications for modularity in vision. Journal of Cognitive Neuroscience 5: 129–149.

Swindale, N. V. (1980). A model for the formation of ocular dominance stripes. Proceedings of the Royal Society of London Series B, Biological Sciences 208(1171): 243–264.

Tanaka, K. (1997). Columnar organization in the inferotemporal cortex. Cerebral Cortex 12: 469–498.

Teller, D. Y. (1997). First glances: The vision of infants. The Friedenwald lecture. Investigative Ophthalmology and Visual Science 38: 2183–2203.

Tolhurst, D. J., J. A. Movshon, and A. F. Dean. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Research 23(8): 775–785.

Tonegawa, S., J. Z. Tsien, T. J. McHugh, P. Huerta, K. I. Blum, and M. A. Wilson. (1996). Hippocampal

CA1-region-restricted knockout of NMDAR1 gene disrupts synaptic plasticity, place fields, and spatial learning. Cold Spring Harbor Symposium on Quantitative Biology 61: 225–238.

Ungerleider, L. G., and M. Mishkin. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, Eds., Analysis of Visual Behavior. Cambridge, MA: MIT Press, pp. 549–586.

Zola-Morgan, S. (1995). Localization of brain function: The legacy of Franz Joseph Gall. Annual Review of Neuroscience 18: 359–383.

Further Readings

Computational Intelligence

Michael I. Jordan and Stuart Russell

There are two complementary views of artificial intelligence (AI): one as an engineering discipline concerned with the creation of intelligent machines, the other as an empirical science concerned with the computational modeling of human intelligence. When the field was young, these two views were seldom distinguished. Since then, a substantial divide has opened up, with the former view dominating modern AI and the latter view characterizing much of modern cognitive science. For this reason, we have adopted the more neutral term "computational intelligence" as the title of this article—both communities are attacking the problem of understanding intelligence in computational terms.

It is our belief that the differences between the engineering models and the cognitively inspired models are small compared to the vast gulf in competence between these models and human levels of intelligence. For humans are, to a first approximation, *intelligent*; they can perceive, act, learn, reason, and communicate *successfully* despite the enormous difficulty of these tasks. Indeed, we expect that as further progress is made in trying to emulate this success, the engineering and cognitive models will become more similar. Already, the traditionally antagonistic "connectionist" and "symbolic" camps are finding common ground, particularly in their understanding of reasoning under uncertainty and learning. This sort of cross-fertilization was a central aspect of the early vision of cognitive science as an interdisciplinary enterprise.

1 Machines and Cognition

The conceptual precursors of AI can be traced back many centuries. LOGIC, the formal theory of deductive reasoning, was studied in ancient Greece, as were ALGORITHMS for mathematical computations. In the late seventeenth century, Wilhelm Leibniz actually constructed simple "conceptual calculators," but their representational and combinatorial powers were far too limited. In the nineteenth century, Charles Babbage designed (but did not build) a device capable of universal computation, and his collaborator Ada Lovelace speculated that the machine might one day be programmed to play chess or compose music. Fundamental work by ALAN TURING in the 1930s formalized the notion of universal computation; the famous CHURCH-TURING THESIS proposed that all sufficiently powerful computing devices were essentially identical in the sense that any one device could emulate the operations of any other. From here it was a small step to the bold hypothesis that human cognition was a form of COMPUTATION in exactly this sense, and could therefore be emulated by computers.

By this time, neurophysiology had already established that the brain consisted largely of a vast interconnected network of NEURONS that used some form of electrical signalling mechanism. The first mathematical model relating computation and the brain appeared in a seminal paper entitled "A logical calculus of the ideas immanent in nervous activity," by WARREN MCCULLOCH and WALTER PITTS (1943). The paper proposed an

abstract model of neurons as linear threshold units—logical "gates" that output a signal if the weighted sum of their inputs exceeds a threshold value (see COMPUTING IN SINGLE NEURONS). It was shown that a network of such gates could represent any logical function, and, with suitable delay components to implement memory, would be capable of universal computation. Together with HEBB's model of learning in networks of neurons, this work can be seen as a precursor of modern and connectionist cognitive modeling. Its stress on the representation of logical concepts by neurons also provided impetus to the "logicist" view of AI.

The emergence of AI proper as a recognizable field required the availability of usable computers; this resulted from the wartime efforts led by Turing in Britain and by JOHN VON NEUMANN in the United States. It also required a banner to be raised; this was done with relish by Turing's (1950) paper "Computing Machinery and Intelligence," wherein an operational definition for intelligence was proposed (the Turing test) and many future developments were sketched out.

One should not underestimate the level of controversy surrounding AI's initial phase. The popular press was only too ready to ascribe intelligence to the new "electronic super-brains," but many academics refused to contemplate the idea of intelligent computers. In his 1950 paper, Turing went to great lengths to catalogue and refute many of their objections. Ironically, one objection already voiced by Kurt Gödel, and repeated up to the present day in various forms, rested on the ideas of incompleteness and undecidability in formal systems to which Turing himself had contributed (see GÖDEL'S THEOREMS and FORMAL SYSTEMS, PROPERTIES OF). Other objectors denied the possibility of CONSCIOUSNESS in computers, and with it the possibility of intelligence. Turing explicitly sought to separate the two, focusing on the objective question of intelligent *behavior* while admitting that consciousness might remain a mystery—as indeed it has.

The next step in the emergence of AI was the formation of a research community; this was achieved at the 1956 Dartmouth meeting convened by John McCarthy. Perhaps the most advanced work presented at this meeting was that of ALLEN NEWELL and Herb Simon, whose program of research in symbolic cognitive modeling was one of the principal influences on cognitive psychology and information-processing psychology.

Newell and Simon's IPL languages were the first symbolic programming languages and among the first high-level languages of any kind. McCarthy's LISP language, developed slightly later, soon became the standard programming language of the AI community and in many ways remains unsurpassed even today. Contemporaneous developments in other fields also led to a dramatic increase in the precision and complexity of the models that could be proposed and analyzed. In linguistics, for example, work by Chomsky (1957) on formal grammars opened up new avenues for the mathematical modeling of mental structures. NORBERT WIENER developed the field of cybernetics (see CONTROL THEORY and MOTOR CONTROL) to provide mathematical tools for the analysis and synthesis of physical control systems.The theory of optimal control in particular has many parallels with the theory of rational agents (see below), but within this tradition no model of internal representation was ever developed.

As might be expected from so young a field with so broad a mandate that draws on so many traditions, the history of AI has been marked by substantial changes in fashion and opinion. Its early days might be described as the "Look, Ma, no hands!" era, when the emphasis was on showing a doubting world that computers *could* play chess, learn, see,

and do all the other things thought to be impossible. A wide variety of methods was tried, ranging from general-purpose symbolic problem solvers to simple neural networks. By the late 1960s, a number of practical and theoretical setbacks had convinced most AI researchers that there would be no simple "magic bullet." The general- purpose methods that had initially seemed so promising came to be called *weak methods* because their reliance on extensive combinatorial search and first-principles knowledge could not overcome the complexity barriers that were, by that time, seen as unavoidable. The 1970s saw the rise of an alternative approach based on the application of large amounts of domain-specific knowledge, expressed in forms that were close enough to the explicit solution as to require little additional computation. Ed Feigenbaum's gnomic dictum, "Knowledge is power," was the watchword of the boom in industrial and commercial application of *expert systems* in the early 1980s.

When the first generation of expert system technology turned out to be too fragile for widespread use, a so-called AI Winter set in—government funding of AI and public perception of its promise both withered in the late 1980s. At the same time, a revival of interest in neural network approaches led to the same kind of optimism as had characterized "traditional" AI in the early 1980s. Since that time, substantial progress has been made in a number of areas within AI, leading to renewed commercial interest in fields such as *data mining* (applied machine learning) and a new wave of expert system technology based on probabilistic inference. The 1990s may in fact come to be seen as the decade of probability. Besides expert systems, the so-called Computational *Bayesian* approach (named after the Reverend Thomas Bayes, eighteenth-century author of the fundamental rule for probabilistic reasoning) has led to new methods in planning, natural language understanding, and learning. Indeed, it seems likely that work on the latter topic will lead to a reconciliation of symbolic and connectionist views of intelligence.

2 Artificial Intelligence: What's the Problem?

The consensus apparent in modern textbooks (Russell and Norvig 1995; Poole, Mackworth, and Goebel 1997; Nilsson 1998) is that AI is about the design of intelligent *agents*. An agent is an entity that can be understood as perceiving and acting on its environment. An agent is *rational* to the extent that its actions can be expected to achieve its goals, given the information available from its perceptual processes. Whereas the Turing test defined only an informal notion of intelligence as emulation of humans, the theory of RATIONAL AGENCY (see also RATIONAL CHOICE THEORY) provides a first pass at a *formal specification* for intelligent agents, with the possibility of a constructive theory to satisfy this specification. Although the last section of this introduction argues that this specification needs a radical rethinking, the idea of RATIONAL DECISION MAKING has nonetheless been the foundation for most of the current research trends in AI.

The focus on AI as the design of intelligent agents is a fairly recent preoccupation. Until the mid-1980s, most research in "core AI" (that is, AI excluding the areas of robotics and computer vision) concentrated on isolated reasoning tasks, the inputs of which were provided by humans and the outputs of which were interpreted by humans. Mathematical theorem-proving systems, English question-answering systems, and medical expert

systems all had this flavor—none of them took actions in any meaningful sense. The so-called situated movement in AI (see SITUATEDNESS/ EMBEDDEDNESS) stressed the point that reasoning is not an end in itself, but serves the purpose of enabling the selection of actions that will affect the reasoner's environment in desirable ways. Thus, reasoning always occurs in a specific context for specific goals. By removing context and taking responsibility for action selection, AI researchers were in danger of defining a subtask that, although useful, actually had no role in the design of a complete intelligent system. For example, some early medical expert systems were constructed in such a way as to accept as input a complete list of symptoms and to output the most likely diagnosis. This might seem like a useful tool, but it ignores several key aspects of medicine: the crucial role of hypothesis-directed *gathering* of information, the very complex task of interpreting sensory data to obtain suggestive and uncertain indicators of symptoms, and the overriding goal of *curing* the patient, which may involve treatments aimed at less likely but potentially dangerous conditions rather than more likely but harmless ones. A second example occurred in robotics. Much research was done on motion planning under the assumption that the locations and shapes of all objects in the environment were known exactly; yet no feasible vision system can, or should, be designed to obtain this information.

When one thinks about building intelligent agents, it quickly becomes obvious that the task environment in which the agent will operate is a primary determiner of the appropriate design. For example, if all relevant aspects of the environment are immediately available to the agent's perceptual apparatus—as, for example, when playing backgammon—then the environment is said to be *fully observable* and the agent need maintain no internal model of the world at all. Backgammon is also *discrete* as opposed to *continuous*—that is, there is a finite set of distinct backgammon board states, whereas tennis, say, requires real-valued variables and changes continuously over time. Backgammon is *stochastic* as opposed to *deterministic*, because it includes dice rolls and unpredictable opponents; hence an agent may need to make contingency plans for many possible outcomes. Backgammon, unlike tennis, is also *static* rather than *dynamic*, in that nothing much happens while the agent is deciding what move to make. Finally, the "physical laws" of the backgammon universe—what the legal moves are and what effect they have—are known rather than unknown. These distinctions alone (and there are many more) define thirty-two substantially different kinds of task environment. This variety of tasks, rather than any true conceptual differences, may be responsible for the variety of computational approaches to intelligence that, on the surface, seem so philosophically incompatible.

3 Architectures of Cognition

Any computational theory of intelligence must propose, at least implicitly, an INTELLIGENT AGENT ARCHITECTURE. Such an architecture defines the underlying organization of the cognitive processes comprising intelligence, and forms the computational substrate upon which domain-specific capabilities are built. For example, an architecture may provide a generic capability for learning the "physical laws" of the environment, for combining inputs from multiple sensors, or for deliberating about actions by envisioning and evaluating their effects. There is, as yet, no satisfactory theory

that defines the range of possible architectures for intelligent systems, or identifies the optimal architecture for a given task environment, or provides a reasonable specification of what is required for an architecture to support "general-purpose" intelligence, either in machines or humans.

Some researchers see the observed variety of intelligent behaviors as a consequence of the operation of a unified, general-purpose problem-solving architecture (Newell 1990). Others propose a functional division of the architecture with modules for perception, learning, reasoning, communication, locomotion, and so on (see MODULARITY OF MIND). Evidence from neuroscience (for example, lesion studies) is often interpreted as showing that the brain is divided into areas, each of which performs some function in this sense; yet the functional descriptions (e.g., "language," "face recognition," etc.) are often subjective and informal and the nature of the connections among the components remains obscure. In the absence of deeper theory, such generalizations from scanty evidence must remain highly suspect. That is, the basic organizational principles of intelligence are still up for grabs.

Proposed architectures vary along a number of dimensions. Perhaps the most commonly cited distinction is between "symbolic" and "connectionist" approaches. These approaches are often thought to be based on fundamentally irreconcilable philosophical foundations. We will argue that, to a large extent, they are complementary; where comparable, they form a continuum. Roughly speaking, a *symbol* is an object, part of the internal state of an agent, that has two properties: it can be compared to other symbols to test for equality, and it can be combined with other symbols to form *symbol structures*. The symbolic approach to AI, in its purest form, is embodied in the physical symbol system (PSS) hypothesis (Newell and Simon 1972), which proposes that algorithmic manipulation of symbol structures is necessary and sufficient for general intelligence (see also COMPUTATIONAL THEORY OF MIND.)

The PSS hypothesis, if taken to its extreme, is identical to the view that cognition can be understood as COMPUTATION. Symbol systems can emulate any Turing machine; in particular, they can carry out finite-precision numerical operations and thereby implement neural networks. Most AI researchers interpret the PSS hypothesis more narrowly, ruling out primitive numerical quantities that are manipulated as *magnitudes* rather than simply tested for (in)equality. The Soar architecture (Newell 1990), which uses PROBLEM SOLVING as its underlying formalism, is the most well developed instantiation of the pure symbolic approach to cognition (see COGNITIVE MODELING, SYMBOLIC).

The symbolic tradition also encompasses approaches to AI that are based on logic. The symbols in the logical languages are used to represent objects and relations among objects, and symbol structures called *sentences* are used to represent facts that the agent knows. Sentences are manipulated according to certain rules to generate new sentences that follow logically from the original sentences. The details of logical agent design are given in the section on knowledge-based systems; what is relevant here is the use of symbol structures as direct representations of the world. For example, if the agent sees John sitting on the fence, it might construct an internal representation from symbols that represent John, the fence, and the sitting-on relation. If Mary is on the fence instead, the symbol structure would be the same except for the use of a symbol for Mary instead of John.

This kind of compositionality of representations is characteristic of symbolic approaches. A more restricted kind of compositionality can occur even in much simpler systems. For example, in the network of logical gates proposed by McCulloch and Pitts, we might have a neuron *J* that is "on" whenever the agent sees John on the fence; and another neuron *M* that is "on" when Mary is on the fence. Then the proposition "either John or Mary is on the fence" can be represented by a neuron that is connected to J and M with the appropriate connection strengths. We call this kind of representation *propositional*, because the fundamental elements are propositions rather than symbols, denoting objects and relations. In the words of McCulloch and Pitts (1943), the state of a neuron was conceived of as "factually equivalent to a proposition which proposed its adequate stimulus." We will also extend the standard sense of "propositional" to cover neural networks comprised of neurons with continuous real-valued activations, rather than the 1/0 activations in the original McCulloch-Pitts threshold neurons.

It is clear that, in this sense, the raw sensory data available to an agent are propositional. For example, the elements of visual perception are "pixels" whose propositional content is, for example, "this area of my retina is receiving bright red light." This observation leads to the first difficulty for the symbolic approach: how to move from sensory data to symbolic representations. This so-called symbol grounding problem has been deemed insoluble by some philosophers (see CONCEPTS), thereby dooming the symbolic approach to oblivion. On the other hand, existence proofs of its solubility abound. For example, Shakey, the first substantial robotics project in AI, used symbolic (logical) reasoning for its deliberations, but interacted with the world quite happily (albeit slowly) through video cameras and wheels (see Raphael 1976). A related problem for purely symbolic approaches is that sensory information about the physical world is usually thought of as numerical—light intensities, forces, strains, frequencies, and so on. Thus, there must at least be a layer of nonsymbolic computation between the real world and the realm of pure symbols. Neither the theory nor the practice of symbolic AI argues against the existence of such a layer, but its existence does open up the possibility that some substantial part of cognition occurs therein without ever reaching the symbolic level. A deeper problem for the narrow PSS hypothesis is UNCERTAINTY—the unavoidable fact that unreliable and partial sensory information, combined with unreliable and partial theories of how the world works, must leave an agent with some doubt as to the truth of virtually all propositions of interest. For example, the stock market may soon recover this week's losses, or it may not. Whether to buy, sell, or hold depends on one's assessment of the prospects. Similarly, a person spotted across a crowded, smoky night club may or may not be an old friend. Whether to wave in greeting depends on how certain one is (and on one's sensitivity to embarrassment due to waving at complete strangers). Although many decisions under uncertainty can be made without reference to numerical degrees of belief (Wellman 1990), one has a lingering sense that degrees of belief in propositions may be a fundamental component of our mental representations. Accounts of such phenomena based on probability theory are now widely accepted within AI as an *augmentation* of the purely symbolic view; in particular, probabilistic models are a natural generalization of the logical approach.

Recent work has also shown that some connectionist representations (e.g., Boltzmann machines) are essentially identical to probabilistic network models developed in AI (see NEURAL NETWORKS). The three issues raised in the preceding paragraphs—

sensorimotor connections to the external world, handling real-valued inputs and outputs, and robust handling of noisy and uncertain information—are primary motivations for the connectionist approach to cognition. (The existence of networks of neurons in the brain is obviously another.) Neural network models show promise for many low-level tasks such as visual pattern recognition and speech recognition. The most obvious drawback of the connectionist approach is the difficulty of envisaging a means to model higher levels of cognition (see BINDING PROBLEM and COGNITIVE MODELING, CONNECTIONIST), particularly when compared to the ability of symbol systems to generate an unbounded variety of structures from a finite set of symbols (see COMPOSITIONALITY). Some solutions have been proposed (see, for example, BINDING BY NEURAL SYNCHRONY); these solutions provide a plausible neural *implementation* of symbolic models of cognition, rather than an *alternative*.

Another problem for connectionist and other propositional approaches is the modeling of *temporally extended* behavior. Unless the external environment is completely observable by the agent's sensors, such behavior requires the agent to maintain some internal state information that reflects properties of the external world that are not directly observable. In the symbolic or logical approach, sentences such as "My car is parked at the corner of Columbus and Union" can be stored in "working memory" or in a "temporal knowledge base" and updated as appropriate. In connectionist models, internal states require the use of RECURRENT NETWORKS, which are as yet poorly understood.

In summary, the symbolic and connectionist approaches seem not antithetical but complementary—connectionist models may handle low-level cognition and may (or rather *must*, in some form) provide a substrate for higher-level symbolic processes. Probabilistic approaches to representation and reasoning may unify the symbolic and connectionist traditions. It seems that the more relevant distinction is between propositional and more expressive forms of representation. Related to the symbolic-connectionist debate is the distinction between *deliberative* and *reactive* models of cognition. Most AI researchers view intelligent behavior as resulting, at least in part, from deliberation over possible courses of action based on the agent's knowledge of the world and of the expected results of its actions. This seems self-evident to the average person in the street, but it has always been a controversial hypothesis—according to BEHAVIORISM, it is meaningless. With the development of KNOWLEDGE-BASED SYSTEMS, starting from the famous "Advice Taker" paper by McCarthy (1958), the deliberative model could be put to the test. The core of a knowledge-based agent is the knowledge base and its associated reasoning procedures; the rest of the design follows straightforwardly. First, we need some way of acquiring the necessary knowledge. This could be from experience through MACHINE LEARNING methods, from humans and books through NATURAL LANGUAGE PROCESSING, by direct programming, or through perceptual processes such as MACHINE VISION. Given knowledge of its environment and of its objectives, an agent can reason that certain actions will achieve those objectives and should be executed. At this point, if we are dealing with a physical environment, robotics takes over, handling the mechanical and geometric aspects of motion and manipulation. The following sections deal with each of these areas in turn. It should be noted, however, that the story in the preceding paragraph is a gross idealization. It is, in fact, close to the view caricatured as good old-fashioned AI (GOFAI) by John Haugeland (1985) and Hubert Dreyfus (1992). In the five decades

since Turing's paper, AI researchers have discovered that attaining real competence is not so simple—the principle barrier being COMPUTATIONAL COMPLEXITY. The idea of *reactive systems* (see also AUTOMATA) is to implement direct mappings from perception to action that avoid the expensive intermediate steps of representation and reasoning. This observation was made within the first month of the Shakey project (Raphael 1976) and given new life in the field of BEHAVIOR-BASED ROBOTICS (Brooks 1991). Direct mappings of this kind can be learned from experience or can be compiled from the results of deliberation within a knowledge-based architecture (see EXPLANATION-BASED LEARNING). Most current models propose a *hybrid* agent design incorporating a variety of decision-making mechanisms, perhaps with capabilities for METAREASONING to control and integrate these mechanisms. Some have even proposed that intelligent systems should be constructed from large numbers of separate agents, each with percepts, actions, and goals of its own (Minsky 1986)—much as a nation's economy is made up of lots of separate humans. The theory of MULTIAGENT SYSTEMS explains how, in some cases, the goals of the whole agent can be achieved even when each sub-agent pursues its own ends.

4 Knowledge-Based Systems

The *procedural-declarative controversy*, which raged in AI through most of the 1970s, was about which way to build AI systems (see, for example, Boden 1977). The procedural view held that systems could be constructed by encoding expertise in domain-specific algorithms—for example, a procedure for diagnosing migraines by asking specific sequences of questions. The declarative view, on the other hand, held that systems should be *knowledge-based*, that is, composed from domain-specific *knowledge*—for example, the symptoms typically associated with various ailments—combined with a general-purpose reasoning system. The procedural view stressed efficiency, whereas the declarative view stressed the fact that the overall internal representation can be decomposed into separate *sentences*, each of which has an identifiable meaning. Advocates of knowledge-based systems often cited the following advantages:

Ease of construction: knowledge-based systems can be constructed simply by encoding domain knowledge extracted from an expert; the system builder need not construct and encode a *solution* to the problems in the domain.

Flexibility: the same knowledge can be used to answer a variety of questions and as a component in a variety of systems; the same reasoning mechanism can be used for all domains.

Modularity: each piece of knowledge can be identified, encoded, and debugged independently of the other pieces.

Learnability: various learning methods exist that can be used to extract the required knowledge from data, whereas it is very hard to construct programs by automatic means.

Explainability: a knowledge-based system can *explain* its decisions by reference to the explicit knowledge it contains.

With arguments such as these, the declarative view prevailed and led to the boom in expert systems in the late 1970s and early 1980s. Unfortunately for the field, the early knowledge-based systems were seldom equal to the challenges of the real world, and

since then there has been a great deal of research to remedy these failings. The area of KNOWLEDGE REPRESENTATION deals with methods for encoding knowledge in a form that can be processed by a computer to derive consequences. Formal logic is used in various forms to represent definite knowledge. To handle areas where definite knowledge is not available (for example, medical diagnosis), methods have been developed for representation and reasoning under uncertainty, including the extension of logic to so-called NONMONOTONIC LOGICS.

All knowledge representation systems need some process for KNOWLEDGE ACQUISITION, and much has been done to automate this process through better interface tools, machine learning methods, and, most recently, extraction from natural language texts. Finally, substantial progress has been made on the question of the computational complexity of reasoning.

5 Logical Representation and Reasoning

Logical reasoning is appropriate when the available knowledge is definite. McCarthy's (1958) "Advice Taker" paper proposed first-order logic (FOL) as a formal language for the representation of commonsense knowledge in AI systems. FOL has sufficient expressive power for most purposes, including the representation of objects, relations among objects, and universally quantified statements about sets of objects. Thanks to work by a long line of philosophers and mathematicians, who were also interested in a formal language for representing general (as well as mathematical) knowledge, FOL came with a well-defined syntax and semantics, as well as the powerful guarantee of *completeness:* there exists a computational procedure such that, if the answer to a question is entailed by the available knowledge, then the procedure will find that answer (see GÖDEL'S THEOREMS). More expressive languages than FOL generally do not allow completeness—roughly put, there exist theorems in these languages that cannot be proved.

The first complete logical reasoning system for FOL, the resolution method, was devised by Robinson (1965). An intense period of activity followed in which LOGICAL REASONING SYSTEMS were applied to mathematics, automatic programming, planning, and general-knowledge question answering. Theorem-proving systems for full FOL have proved new theorems in mathematics and have found widespread application in areas such as program verification, which spun off from mainstream AI in the early 1970s. Despite these early successes, AI researchers soon realized that the computational complexity of general-purpose reasoning with full FOL is prohibitive; such systems could not scale up to handle large knowledge bases. A great deal of attention has therefore been given to more restricted languages. *Database systems*, which have long been distinct from AI, are essentially logical question-answering systems the knowledge bases of which are restricted to very simple sentences about specific objects.

Propositional languages avoid objects altogether, representing the world by the discrete values of a fixed set of propositional variables and by logical combinations thereof. (Most neural network models fall into this category also.) Propositional reasoning methods based on CONSTRAINT SATISFACTION and GREEDY LOCAL SEARCH have been very successful in real-world applications, but the restricted expressive power of propositional languages severely limits their scope. Much closer to

the expressive power of FOL are the languages used in LOGIC PROGRAMMING. Although still allowing most kinds of knowledge to be expressed very naturally, logic programming systems such as Prolog provide much more efficient reasoning and can work with extremely large knowledge bases.

Reasoning systems must have content with which to reason. Researchers in knowledge representation study methods for codifying and reasoning with particular kinds of knowledge. For example, McCarthy (1963) proposed the SITUATION CALCULUS as a way to represent states of the world and the effects of actions within first-order logic. Early versions of the situation calculus suffered from the infamous FRAME PROBLEM—the apparent need to specify sentences in the knowledge base for all the *noneffects* of actions. Some philosophers see the frame problem as evidence of the impossibility of the formal, knowledge-based approach to AI, but simple technical advances have resolved the original issues. Situation calculus is perhaps the simplest form of TEMPORAL REASONING; other formalisms have been developed that provide substantially more general frameworks for handling time and extended events. Reasoning about knowledge itself is important particularly when dealing with other agents, and is usually handled by MODAL LOGIC, an extension of FOL. Other topics studied include reasoning about ownership and transactions, reasoning about substances (as distinct from objects), and reasoning about physical representations of information. A general *ontology*—literally, a description of existence—ties all these areas together into a unified taxonomic hierarchy of categories. FRAME-BASED SYSTEMS are often used to represent such hierarchies, and use specialized reasoning methods based on *inheritance* of properties in the hierarchy.

6 Logical Decision Making

An agent's job is to make *decisions*, that is, to commit to particular actions. The connection between logical reasoning and decision making is simple: the agent must conclude, based on its knowledge, that a certain action is best. In philosophy, this is known as *practical reasoning*. There are many routes to such conclusions. The simplest leads to a reactive system using *condition-action rules* of the form "If P then do A." Somewhat more complex reasoning is required when the agent has explicitly represented *goals*. A goal "G" is a description of a desired state of affairs—for example, one might have the goal "On vacation in the Seychelles." The *practical syllogism*, first expounded by Aristotle, says that if G is a goal, and A achieves G, then A should be done. Obviously, this rule is open to many objections: it does not specify which of many eligible As should be done, nor does it account for possibly disastrous sideeffects of A. Nonetheless, it underlies most forms of decision making in the logical context.

Often, there will be no single action A that achieves the goal G, but a solution may exist in the form of a *sequence* of actions. Finding such a sequence is called PROBLEM SOLVING, where the word "problem" refers to a task defined by a set of actions, an initial state, a goal, and a set of reachable states. Much of the early cognitive modeling work of Newell and Simon (1972) focused on problem solving, which was seen as a quintessentially intelligent activity. A great deal of research has been done on efficient algorithms for problem solving in the areas of HEURISTIC SEARCH and GAME-PLAYING SYSTEMS. The "cognitive structure" of such systems is very simple, and

problem-solving competence is often achieved by means of searching through huge numbers of possibilities. For example, the Deep Blue chess program, which defeated human world champion Gary Kasparov, often examined over a billion positions prior to each move. Human competence is not thought to involve such computations (see CHESS, PSYCHOLOGY OF).

Most problem-solving algorithms treat the states of the world as atomic—that is, the internal structure of the state representation is not accessible to the algorithm as it considers the possible sequences of actions. This fails to take advantage of two very important sources of power for intelligent systems: the ability to *decompose* complex problems into subproblems and the ability to identify relevant actions from explicit goal descriptions. For example, an intelligent system should be able decompose the goal "have groceries and a clean car" into the subgoals "have groceries" and "have a clean car." Furthermore, it should immediately consider buying groceries and washing the car. Most search algorithms, on the other hand, may consider a variety of action sequences—sitting down, standing up, going to sleep, and so on—before happening on some actions that are relevant.

In principle, a logical reasoning system using McCarthy's situation calculus can generate the kinds of reasoning behaviors necessary for decomposing complex goals and selecting relevant actions. For reasons of computational efficiency, however, special-purpose PLANNING systems have been developed, originating with the STRIPS planner used by Shakey the Robot (Fikes and Nilsson 1971). Modern planners have been applied to logistical problems that are, in some cases, too complex for humans to handle effectively.

7 Representation and Reasoning under Uncertainty

In many areas to which one might wish to apply knowledge-based systems, the available knowledge is far from definite. For example, a person who experiences recurrent headaches may suffer from migraines or a brain tumor. A logical reasoning system can represent this sort of disjunctive information, but cannot represent or reason with the belief that migraine is a *more likely* explanation. Such reasoning is obviously essential for diagnosis, and has turned out to be central for expert systems in almost all areas. The theory of *probability* (see PROBABILITY, FOUNDATIONS OF) is now widely accepted as the basic calculus for reasoning under uncertainty (but see FUZZY LOGIC for a complementary view). Questions remain as to whether it is a good model for human reasoning (see TVERSKY and PROBABILISTIC REASONING), but within AI many of the computational and representational problems that deterred early researchers have been resolved. The adoption of a probabilistic approach has also created rich connections with statistics and control theory.

Standard probability theory views the world as comprised of a set of interrelated random variables the values of which are initially unknown. Knowledge comes in the form of *prior* probability distributions over the possible assignments of values to subsets of the random variables. Then, when evidence is obtained about the values of some of the variables, inference algorithms can infer *posterior* probabilities for the remaining unknown variables. Early attempts to use probabilistic reasoning in AI came up against complexity barriers very soon, because the number of probabilities that make up the prior probability distribution can grow exponentially in the number of variables considered.

Starting in the early 1980s, researchers in AI, decision analysis, and statistics developed what are now known as BAYESIAN NETWORKS (Pearl 1988). These networks give structure to probabilistic knowledge bases by expressing *conditional independence*relationships among the variables. For example, given the actual temperature, the temperature measurements of two thermometers are independent. In this way, Bayesian networks capture our intuitive notions of the causal structure of the domain of application. In most cases, the number of probabilities that must be specified in a Bayesian network grows only linearly with the number of variables. Such systems can therefore handle quite large problems, and applications are very widespread. Moreover, methods exist for *learning* Bayesian networks from raw data (see BAYESIAN LEARNING), making them a natural bridge between the symbolic and neural-network approaches to AI.

In earlier sections, we have stressed the importance of the distinction between propositional and first-order languages. So far, probability theory has been limited to essentially propositional representations; this prevents its application to the more complex forms of cognition addressed by first-order methods. The attempt to unify probability theory and first-order logic, two of the most fundamental developments in the history of mathematics and philosophy, is among the more important topics in current AI research.

8 Decision Making under Uncertainty

Just as logical reasoning is connected to action through goals, probabilistic reasoning is connected to action through *utilities*, which describe an agent's preferences for some states over others. It is a fundamental result of UTILITY THEORY (see also RATIONAL CHOICE THEORY) that an agent whose preferences obey certain rationality constraints, such as transitivity, can be modeled as possessing a *utility function* that assigns a numerical value to each possible state. Furthermore, RATIONAL DECISION MAKING consists of selecting an action to maximize the expected utility of outcome states. An agent that makes rational decisions will, on average, do better than an agent that does not—at least as far as satisfying its own preferences is concerned.

In addition to their fundamental contributions to utility theory, von Neumann and Morgenstern (1944) also developed GAME THEORY to handle the case where the environment contains other agents, which must be modeled as independent utility maximizers. In some game-theoretic situations, it can be shown that optimal behavior must be *randomized*. Additional complexities arise when dealing with so-called *sequential* decision problems, which are analogous to planning problems in the logical case.

DYNAMIC PROGRAMMING algorithms, developed in the field of operations research, can generate optimal behavior for such problems. (See also the discussion of REINFORCEMENT LEARNING in segment 12—Learning.) In a sense, the theory of rational decision making provides a zeroth-order theory of intelligence, because it provides an operational definition of what an agent *ought* to do in any situation. Virtually every problem an agent faces, including such problems as how to gather information and how to update its beliefs given that information, can be formulated within the theory and,

in principle, solved. What the theory ignores is the question of complexity, which we discuss in the final section of this introduction.

9 Learning

LEARNING has been a central aspect of AI from its earliest days. It is immediately apparent that learning is a vital characteristic of any intelligent system that has to deal with changing environments. Learning may also be the only way in which complex and competent systems can be constructed—a proposal stated clearly by Turing (1950), who devoted a quarter of his paper to the topic. Perhaps the first major public success for AI was Arthur Samuel's (1959) checker-playing system, which learned to play checkers to a level far superior to its creator's abilities and attracted substantial television coverage. State-of-the-art systems in almost all areas of AI now use learning to avoid the need for the system designer to have to anticipate and provide knowledge to handle every possible contingency. In some cases, for example speech recognition, humans are simply incapable of providing the necessary knowledge accurately.

The discipline of machine learning has become perhaps the largest subfield of AI as well as a meeting point between AI and various other engineering disciplines concerned with the design of autonomous, robust systems. An enormous variety of learning systems has been studied in the AI literature, but once superficial differences are stripped away, there seem to be a few core principles at work. To reveal these principles it helps to classify a given learning system along a number of dimensions: (1) the type of feedback available, (2) the component of the agent to be improved, (3) how that component is represented, and (4) the role of prior knowledge. It is also important to be aware that there is a tradeoff between learning and inference and different systems rely more on one than on the other.

The type of feedback available is perhaps the most useful categorizer of learning algorithms. Broadly speaking, learning algorithms fall into the categories of *supervised learning*, *unsupervised learning*, and *reinforcement learning*. Supervised learning algorithms (see, e.g., DECISION TREES and SUPERVISED LEARNING IN MULTILAYER NEURAL NETWORKS) require that a target output is available for every input, an assumption that is natural in some situations (e.g., categorization problems with labeled data, imitation problems, and prediction problems, in which the present can be used as a target for a prediction based on the past). UNSUPERVISED LEARNING algorithms simply find structure in an ensemble of data, whether or not this structure is useful for a particular classification or prediction (examples include clustering algorithms, dimensionality-reducing algorithms, and algorithms that find independent components). REINFORCEMENT LEARNING algorithms require an evaluation signal that gives some measure of progress without necessarily providing an example of correct behavior. Reinforcement learning research has had a particular focus on temporal learning problems, in which the evaluation arrives after a sequence of responses. The different components of an agent generally have different kinds of representational and inferential requirements. Sensory and motor systems must interface with the physical world and therefore generally require continuous representations and smooth input-output behavior. In such situations, neural networks have provided a useful class of architectures, as have probabilistic systems such as HIDDEN MARKOV MODELS and Bayesian networks. The latter models also are generally characterized by a clear

propositional semantics, and as such have been exploited for elementary cognitive processing. Decision trees are also propositional systems that are appropriate for simple cognitive tasks. There are variants of decision trees that utilize continuous representations, and these have close links with neural networks, as well as variants of decision trees that utilize relational machinery, making a connection with INDUCTIVE LOGIC PROGRAMMING. The latter class of architecture provides the full power of firstorder logic and the capability of learning complex symbolic theories.

Prior knowledge is an important component of essentially all modern learning architectures, particularly so in architectures that involve expressive representations. Indeed, the spirit of inductive logic programming is to use the power of logical inference to bootstrap background knowledge and to interpret new data in the light of that knowledge. This approach is carried to what is perhaps its (logical) extreme in the case of EXPLANATION-BASED LEARNING (EBL), in which the system uses its current theory to *explain* a new observation, and extracts from that explanation a useful rule for future use. EBL can be viewed as a form of generalized caching, also called *speedup learning*. CASE-BASED REASONING AND ANALOGY provides an alternate route to the same end through the solution of problems by reference to previous experience instead of first principles.

Underlying all research on learning is a version of the general problem of INDUCTION; in particular, on what basis can we expect that a system that performs well on past "training" data should also perform well on future "test" data? The theory of learning (see COMPUTATIONAL LEARNING THEORY and STATISTICAL LEARNING THEORY) attacks this problem by assuming that the data provided to a learner is obtained from a fixed but unknown probability distribution. The theory yields a notion of *sample complexity*, which quantifies the amount of data that a learner must see in order to expect—with high probability—to perform (nearly) as well in the future as in the past. The theory also provides support for the intuitive notion of Ockham's razor—the idea that if a simple hypothesis performs as well as a complex hypothesis, one should prefer the simple hypothesis (see PARSIMONY AND SIMPLICITY). General ideas from probability theory in the form of Bayesian learning, as well as related ideas from INFORMATION THEORY in the form of the MINIMUM DESCRIPTION LENGTH approach provide a link between learning theory and learning practice. In particular, Bayesian learning, which views learning as the updating of probabilistic beliefs in hypotheses given evidence, naturally embodies a form of Ockham's razor. Bayesian methods have been applied to neural networks, Bayesian networks, decision trees, and many other learning architectures.

We have seen that learning has strong relationships to knowledge representation and to the study of uncertainty. There are also important connections between learning and search. In particular, most learning algorithms involve some form of search through the hypothesis space to find hypotheses that are consistent (or nearly so) with the data and with prior expectations. Standard heuristic search algorithms are often invoked—either explicitly or implicitly—to perform this search. EVOLUTIONARY COMPUTATION also treats learning as a search process, in which the "hypothesis" is an entire agent, and learning takes place by "mutation" and "natural selection" of agents that perform well (see also ARTIFICIAL LIFE). There are also interesting links between learning and planning; in particular, it is possible to view reinforcement learning as a form of "on-

line" planning. Finally, it is worth noting that learning has been a particularly successful branch of AI research in terms of its applications to real-world problems in specific fields;

10 Language

NATURAL LANGUAGE PROCESSING, or NLP—the ability to perceive, understand, and generate language—is an essential part of HUMAN-COMPUTER INTERACTION as well as the most obvious task to be solved in passing the Turing test. As with logical reasoning, AI researchers have benefited from a pre-existing intellectual tradition. The field of linguistics (see also LINGUISTICS, PHILOSOPHICAL ISSUES) has produced formal notions of SYNTAX and SEMANTICS, the view of utterances as *speech acts*, and very careful philosophical analyses of the meanings of various constructs in natural language. The field of COMPUTATIONAL LINGUISTICS has grown up since the 1960s as a fertile union of ideas from AI, cognitive science, and linguistics.

As soon as programs were written to process natural language, it became obvious that the problem was much harder than had been anticipated. In the United States substantial effort was devoted to Russian-English translation from 1957 onward, but in 1966 a government report concluded that "there has been no machine translation of general scientific text, and none is in immediate prospect." Successful MACHINE TRANSLATION appeared to require an *understanding* of the content of the text; the barriers included massive ambiguity (both syntactic and semantic), a huge variety of word senses, and the vast numbers of idiosyncratic ways of using words to convey meanings. Overcoming these barriers seems to require the use of large amounts of commonsense knowledge and the ability to reason with it—in other words, solving a large fraction of the AI problem. For this reason, Robert Wilensky has described natural language processing as an "AI-complete" problem (see also MODULARITY AND LANGUAGE).

Research in NLP has uncovered a great deal of new information about language. There is a better appreciation of the *actual* syntax of natural language—as opposed to the vastly oversimplified models that held sway before computational investigation was possible. Several new families of FORMAL GRAMMARS have been proposed as a result. In the area of semantics, dozens of interesting phenomena have surfaced—for example, the surprising range of semantic relationships in noun-noun pairs such as "alligator shoes" and "baby shoes." In the area of DISCOURSE understanding, researchers have found that grammaticality is sometimes thrown out of the window, leading some to propose that grammar itself is not a useful construct for NLP.

One consequence of the richness of natural language is that it is very difficult to build by hand a system capable of handling anything close to the full range of phenomena. Most systems constructed prior to the 1990s functioned only in predefined and highly circumscribed domains. Stimulated in part by the availability of large online text corpora, the use of STATISTICAL TECHNIQUES IN NATURAL LANGUAGE PROCESSINGhas created something of a revolution. Instead of building complex grammars by hand, these techniques train very large but very simple probabilistic grammars and semantic models from millions of words of text. These techniques have

reached the point where they can be usefully applied to extract information from general newspaper articles.

Few researchers expect simple probability models to yield human-level understanding. On the other hand, the view of language entailed by this approach—that the text is a form of *evidence* from which higher-level facts can be inferred by a process of probabilistic inference—may prove crucial for further progress in NLP. A probabilistic framework allows the smooth integration of the multiple "cues" required for NLP, such as syntax, semantics, discourse conventions, and prior expectations.

In contrast to the general problem of natural language understanding, the problem of SPEECH RECOGNITION IN MACHINES may be feasible without recourse to general knowledge and reasoning capabilities. The statistical approach was taken much earlier in the speech field, beginning in the mid-1970s. Together with improvements in the signal processing methods used to extract acoustic features, this has led to steady improvements in performance, to the point where commercial systems can handle dictated speech with over 95 percent accuracy. The combination of speech recognition and SPEECH SYNTHESIS (see also NATURAL LANGUAGE GENERATION) promises to make interaction with computers much more natural for humans. Unfortunately, accuracy rates for natural dialogue seldom exceed 75 percent; possibly, speech systems will have to rely on knowledge-based expectations and real understanding to make further progress.

11 Vision

The study of vision presents a number of advantages—visual processing systems are present across a wide variety of species, they are reasonably accessible experimentally (psychophysically, neuropsychologically, and neurophysiologically), and a wide variety of artificial imaging systems are available that are sufficiently similar to their natural counterparts so as to make research in machine vision highly relevant to research in natural vision. An integrated view of the problem has emerged, linking research in COMPUTATIONAL VISION, which is concerned with the development of explicit theories of human and animal vision, with MACHINE VISION, which is concerned with the development of an engineering science of vision.

Computational approaches to vision, including the influential theoretical framework of MARR, generally involve a succession of processes that begin with localized numeric operations on images (so-called early vision) and proceed toward the highlevel abstractions thought to be involved in OBJECT RECOGNITION. The current view is that the interpretation of complex scenes involves inference in both the bottom-up and top-down directions (see also TOP-DOWN PROCESSING IN VISION). High-level object recognition is not the only purpose of vision. Representations at intermediate levels can also be an end unto themselves, directly subserving control processes of orienting, locomotion, reaching, and grasping. Visual analysis at all levels can be viewed as a process of recovering aspects of the visual scene from its projection onto a 2-D image. Visual properties such as shape and TEXTURE behave in lawful ways under the geometry of perspective projection, and understanding this geometry has been a focus of research. Related geometrical issues have been studied in STEREO AND MOTION PERCEPTION, where the issue of finding correspondences between multiple images also

arises. In all of these cases, localized spatial and temporal cues are generally highly ambiguous with respect to the aspects of the scene from which they arise, and algorithms that recover such aspects generally involve some form of spatial or temporal integration. It is also important to prevent integrative processes from wrongly smoothing across discontinuities that correspond to visually meaningful boundaries. Thus, visual processing also requires segmentation. Various algorithms have been studied for the segmentation of image data. Again, an understanding of projective geometry has been a guide for the development of such algorithms. Integration and segmentation are also required at higher levels of visual processing, where more abstract principles (such as those studied by GESTALT PSYCHOLOGY; see GESTALT PERCEPTION) are needed to group visual elements.

Finally, in many cases the goal of visual processing is to detect or recognize objects in the visual scene. A number of difficult issues arise in VISUAL OBJECT RECOGNITION, including the issue of what kinds of features should be used (2-D or 3-D, edge-based or filter-based), how to deal with missing features (e.g., due to occlusion or shadows), how to represent flexible objects (such as humans), and how to deal with variations in pose and lighting. Methods based on learning (cf. VISION AND LEARNING) have played an increasingly important role in addressing some of these issues.

12 Robotics

Robotics is the control of physical effectors to achieve physical tasks such as navigation and assembly of complex objects. Effectors include grippers and arms to perform MANIPULATION AND GRASPING and wheels and legs for MOBILE ROBOTS and WALKING AND RUNNING MACHINES. The need to interact directly with a physical environment, which is generally only partially known and partially controllable, brings certain issues to the fore in robotics that are often skirted in other areas in AI. One important set of issues arises from the fact that environments are generally dynamical systems, characterizable by a large (perhaps infinite) collection of real-valued state variables, whose values are not generally directly observable by the robot (i.e., they are “hidden”). The presence of the robot control algorithm itself as a feedback loop in the environment introduces additional dynamics. The robot designer must be concerned with the issue of *stability* in such a situation. Achieving stability not only prevents disasters but it also simplifies the dynamics, providing a degree of predictability that is essential for the success of planning algorithms.

Stability is a key issue in manipulation and grasping, where the robot must impart a distributed pattern of forces and torques to an object so as to maintain a desired position and orientation in the presence of external disturbances (such as gravity). Research has tended to focus on static stability (ignoring the dynamics of the grasped object). Static stability is also of concern in the design of walking and running robots, although rather more pertinent is the problem of dynamic stability, in which a moving robot is stabilized by taking advantage of its inertial dynamics.

Another important set of issues in robotics has to do with uncertainty. Robots are generally equipped with a limited set of sensors and these sensors are generally noisy and

inherently ambiguous. To a certain extent the issue is the same as that treated in the preceding discussion of vision, and the solutions, involving algorithms for integration and smoothing, are often essentially the same. In robotics, however, the sensory analysis is generally used to subserve a control law and the exigencies of feedback control introduce new problems (cf. CONTROL THEORY). Processing time must be held to a minimum and the system must focus on obtaining only that information needed for control. These objectives can be difficult to meet, and recent research in robotics has focused on minimizing the need for feedback, designing sequences of control actions that are guaranteed to bring objects into desired positions and orientations regardless of the initial conditions.

Uncertainty is due not only to noisy sensors and hidden states, but also to ignorance about the structure of the environment. Many robot systems actively model the environment, using system identification techniques from control theory, as well as more general supervised and unsupervised methods from machine learning. Specialized representations are often used to represent obstacles ("configuration space") and location in space (graphs and grids). Probabilistic approaches are often used to explicitly represent and manipulate uncertainty within these formalisms.

In classical robotic control methodology, the system attempts to recover as much of the state of the environment as possible, operates on the internal representation of the state using general planning and reasoning algorithms, and chooses a sequence of control actions to implement the selected plan. The sheer complexity of designing this kind of architecture has led researchers to investigate simpler architectures that make do with minimal internal state. BEHAVIOR-BASED ROBOTICS approaches the problem via an interacting set of elemental processes called "behaviors," each of which is a simplified control law relating sensations and actions. REINFORCEMENT LEARNING has provided algorithms that utilize simplified evaluation signals to guide a search for improved laws; over time these algorithms approach the optimal plans that are derived (with more computational effort) from explicit planning algorithms (see ROBOTICS AND LEARNING).

13 Complexity, Rationality, and Intelligence

We have observed at several points in this introduction that COMPUTATIONAL COMPLEXITY is a major problem for intelligent agents. To the extent that they can be analyzed, most of the problems of perceiving, learning, reasoning, and decision making are believed to have a worst-case complexity that is at least exponential in the size of the problem description. Exponential complexity means that, for example, a problem of size 100 would take 10 billion years to solve on the fastest available computers. Given that humans face much larger problems than this all the time—we receive as input several billion bytes of information every second—one wonders how we manage at all. Of course, there are a number of mitigating factors: an intelligent agent must deal largely with the typical case, not the worst case, and accumulated experience with similar problems can greatly reduce the difficulty of new problems. The fact remains, however, that humans cannot even come close to achieving perfectly rational behavior—most of us do fairly poorly even on problems such as chess, which is an *infinitesimal* subset of the

real world. What, then, is the right thing for an agent to do, if it cannot possibly compute the right thing to do?

In practical applications of AI, one possibility is to restrict the allowable set of problems to those that are efficiently soluble. For example, deductive database systems use restricted subsets of logic that allow for polynomial-time inference. Such research has given us a much deeper understanding of the sources of complexity in reasoning, but does not seem directly applicable to the problem of general intelligence. Somehow, we must face up to the inevitable compromises that must be made in the quality of decisions that an intelligent agent can make. Descriptive theories of such compromises—for example, Herbert Simon's work on *satisficing*—appeared soon after the development of formal theories of rationality. Normative theories of BOUNDED RATIONALITY address the question at the end of the preceding paragraph by examining what is achievable with fixed computational resources. One promising approach is to devote some of those resources to METAREASONING (see also METACOGNITION), that is, reasoning about what reasoning to do. The technique of EXPLANATION-BASED LEARNING (a formalization of the common psychological concept of *chunking* or *knowledge compilation*) helps an agent cope with complexity by caching efficient solutions to common problems. Reinforcement learning methods enable an agent to learn effective (if not perfect) behaviors in complex environments without the need for extended problem-solving computations.

What is interesting about all these aspects of intelligence is that without the need for effective use of limited computational resources, they make no sense. That is, computational complexity may be responsible for many, perhaps most, of the aspects of cognition that make intelligence an interesting subject of study. In contrast, the cognitive structure of an infinitely powerful computational device could be very straightforward indeed.

14 Additional Sources

Early AI work is covered in Feigenbaum and Feldman's (1963) *Computers and Thought*, Minsky's (1968) *Semantic Information Processing*, and the *Machine Intelligence* series edited by Donald Michie. A large number of influential papers are collected in *Readings in Artificial Intelligence* (Webber and Nilsson 1981). Early papers on neural networks are collected in *Neurocomputing* (Anderson and Rosenfeld 1988). The *Encyclopedia of AI* (Shapiro 1992) contains survey articles on almost every topic in AI. The four-volume *Handbook of Artificial Intelligence* (Barr and Feigenbaum 1981) contains descriptions of almost every major AI system published before 1981.

Standard texts on AI include *Artificial Intelligence: A Modern Approach* (Russell and Norvig 1995) and *Artificial Intelligence: A New Synthesis* (Nilsson 1998). Historical surveys include Kurzweil (1990) and Crevier (1993).

The most recent work appears in the proceedings of the major AI conferences: the biennial International Joint Conference on AI (IJCAI); the annual National Conference on AI, more often known as AAAI after its sponsoring organization; and the European Conference on AI (ECAI). The major journals for general AI are *Artificial Intelligence, Computational Intelligence,* the IEEE *Transactions on Pattern Analysis and Machine*

Intelligence, and the electronic *Journal of Artificial IntelligenceResearch*. There are also many journals devoted to specific areas, some of which are listed in the relevant articles. The main professional societies for AI are the American Association for Artificial Intelligence (AAAI), the ACM Special Interest Group in Artificial Intelligence (SIGART), and the Society for Artificial Intelligence and Simulation of Behaviour (AISB). AAAI's *AI Magazine* and the *SIGART Bulletin* contain many topical and tutorial articles as well as announcements of conferences and workshops

References

Anderson, J. A., and E. Rosenfeld, Eds. (1988). *Neurocomputing: Foundations of Research*. Cambridge, MA: MIT Press.

Barr, A., P. R. Cohen, and E. A. Feigenbaum, Eds. (1989). *The Handbook of Artificial Intelligence* vol. 4. Reading, MA: Addison-Wesley.

Barr, A., and E. A. Feigenbaum, Eds. (1981). *The Handbook of Artificial Intelligence*, vol. 1. Stanford and Los Altos, CA: HeurisTech Press and Kaufmann.

Barr, A., and E. A. Feigenbaum, Eds. (1982). *The Handbook of Artificial Intelligence*, vol. 2. Stanford and Los Altos, CA: HeurisTech Press and Kaufmann.

Boden, M. A. (1977). *Artificial Intelligence and Natural Man.* New York: Basic Books.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence* 47(1–3): 139–159.

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

Cohen, P. R., and E. A. Feigenbaum, Eds. (1982). *The Handbook of Artificial Intelligence,* vol. 3. Stanford and Los Altos, CA: HeurisTech Press and Kaufmann.

Crevier, D. (1993). *AI: The Tumultuous History of the Search for Artificial Intelligence.* New York: Basic Books.

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.

Feigenbaum, E. A., and J. Feldman, Eds. (1963). *Computers and Thought*. New York: McGraw-Hill.

Fikes, R. E., and N. J. Nilsson. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2(3–4): 189–208.

Haugeland, J., Ed. (1985). *Artificial Intelligence: The Very Idea.* Cambridge, MA: MIT Press.

Kurzweil, R. (1990). *The Age of Intelligent Machines*. Cambridge, MA: MIT Press. McCarthy, J. (1958). Programs with common sense. *Proceedings of the Symposium on Mechanisation of Thought Processes,* vol. 1. London: Her Majesty's Stationery Office, pp. 77–84.

McCarthy, J. (1963). Situations, actions, and causal laws. Memo 2. Stanford, CA: Stanford University Artificial Intelligence Project.

McCulloch, W. S., and W. Pitts. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5: 115–137.

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

Minsky, M. L. (1986). *The Society of Mind.* New York: Simon & Schuster.

Newell, A. (1990). *Unified Theories of Cognition.* Cambridge, MA: Harvard University Press.

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

Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis.* San Mateo, CA: Morgan Kaufmann.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA: Morgan Kaufmann.

Poole, D., A. Mackworth, and R. Goebel. (1997). *Computational Intelligence: A Logical Approach.* Oxford: Oxford University Press.

Raphael, B. (1976). *The Thinking Computer: Mind Inside Matter*. New York: W. H. Freeman.

Robinson, J. A. (1965). A machine-oriented logic based on the resolution principle. *Journal of the Association for Computing Machinery* 12: 23–41.

Russell, S. J., and P. Norvig. (1995). *Artificial Intelligence: A Modern Approach.* Englewood Cliffs, NJ: Prentice-Hall.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development* 3(3): 210–229.

Shapiro, S. C., Ed. (1992). *Encyclopedia of Artificial Intelligence.* 2nd ed. New York: Wiley.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind* 59: 433–460.

Von Neumann, J., and O. Morgenstern. (1944). *Theory of Games and Economic Behavior*. 1st ed., Princeton, NJ: Princeton University Press.

Webber, B. L., and N. J. Nilsson, Eds. (1981). *Readings in Artificial Intelligence.* San Mateo, CA: Morgan Kaufmann.

Wellman, M. P. (1990). Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence* 44(3): 257–303.