ONE

Cognitive Science and Human Experience

COGNITIVE SCIENCE—that part of the science of the mind traditionally concerned with cognitive processes—has been described as having "a very long past but a relatively short history" (Gardner 1985, p. 9). Scientific concern with the mind can be traced all the way back to Plato and Aristotle, but the term *cognitive science* did not arise until the late twentieth century, as a name for the new, modern, scientific research progam that integrated psychology, neuroscience, linguistics, computer science, artificial intelligence (AI), and philosophy. What united these disciplines, and set cognitive science apart from earlier approaches in psychology and philosophy, was the goal of making explicit the principles and mechanisms of cognition. Cognitive science, in providing a whole new array of concepts, models, and experimental techniques, claimed to be able to provide rigorous scientific knowledge of the mind beyond what earlier forms of psychology and philosophy had offered.

In recent years, however, it has become increasingly clear to many researchers that cognitive science is incomplete. Cognitive science has focused on cognition while neglecting emotion, affect, and motivation (LeDoux 2002, p. 24). In addition, a complete science of the mind needs to account for subjectivity and consciousness.

With hindsight it has also become evident that, in the passage from traditional philosophy and psychology to modern-day cognitive science, something was lost that only now is beginning to be reclaimed. What was lost, in a nutshell, was scientific concern with subjective ex-

perience. In 1892 William James quoted with approval George Trumball Ladd's definition of psychology "as the description and explanation of states of consciousness as such" (James 1985, p. xxv; emphasis omitted). Consciousness was supposed to be the subject matter of psychology, yet cognitive science has had virtually nothing to say about it until recent years. To understand this neglect we need to consider the development of cognitive science since the 1950s.

Three major approaches to the study of the mind can be distinguished within cognitive science—cognitivism, connectionism, and embodied dynamicism. Each approach has its preferred theoretical metaphor for understanding the mind. For cognitivism, the metaphor is the mind as digital computer; for connectionism, it is the mind as neural network; for embodied dynamicism, it is the mind as an embodied dynamic system. Cognitivism dominated the field from the 1950s to the 1970s. In the 1980s, connectionism began to challenge the cognitivist orthodoxy, followed in the 1990s by embodied dynamicism. In contemporary research, all three approaches coexist, both separately and in various hybrid forms.[1]

Cognitivism

Cognitive science came into being in the 1950s with the "cognitive revolution" against behaviorist psychology. At the center of this revolution was the computer model of mind, now known as the classical conception of cognitive processes. According to this classical model, cognition is information processing after the fashion of the digital computer. Behaviorism had allowed no reference to internal states of the organism; explanations of behavior had to be formulated in terms of sensory stimuli and behavioral conditioning (on the input side), and overt behavioral response (on the output side). The computer model of the mind not only made reference to internal states legitimate, but also showed it to be necessary in accounting for the behavior of complex information processing systems. Even more important, the computer model was taken to show how content or meaning could be attributed to states inside the system. A computer is supposed to be a symbol-manipulating machine.[2] A symbol is an item that has a physical shape or form, and that stands for or represents something. According to the computer model of the mind, the brain, too, is a computer, a

"physical symbol system," and mental processes are carried out by the manipulation of symbolic representations in the brain (Newell and Simon 1976; Pylyshyn 1984). A typical cognitivist model takes the form of a program for solving a problem in some domain. Nonsymbolic sensory inputs are transduced and mapped onto symbolic representations of the task domain. These representations are then manipulated in a purely formal or syntactic fashion in order to arrive at a solution to the problem. Cognitivist explanations focus on the abstract problem-solving characterization of cognitive tasks, the structure and content of symbolic representations, and the nature of the algorithms for manipulating the representations in order to solve a given problem. Cognitivism goes hand in hand with functionalism in the philosophy of mind, which in its extreme computational form holds that the embodiment of the organism is essentially irrelevant to the nature of the mind. It is the software, not the hardware, that matters most for mentality.

Cognitivism made meaning, in the sense of representational semantics, scientifically acceptable, but at the price of banishing consciousness from the science of the mind. (In fact, cognitivism inherited its consciousness taboo directly from behaviorism.) Mental processes, understood to be computations made by the brain using an inner symbolic language, were taken to be entirely nonconscious. Thus the connection between mind and meaning, on the one hand, and subjectivity and consciousness, on the other, was completely severed.

Long before cognitivism, Freud had already undermined any simplistic identification of mind and consciousness. According to his early model, the psyche is composed of three systems, which he called the conscious, the preconscious, and the unconscious (Freud 1915, pp. 159–222). The conscious corresponds to the field of awareness, and the preconscious to what we can recall but are not aware of now. The unconscious, in contrast, Freud considered to be part of our phylogenetic heritage. It is thoroughly somatic and affective, and its contents have been radically separated from consciousness by repression and cannot enter the conscious–preconscious system without distortion. (Later, Freud introduced a new structural model composed of the ego, id, and superego; see Freud 1923, pp. 339–407.)

The cognitivist separation of cognition and consciousness, however, was different from Freud's model. Mental processes, according to cog-

nitivism, are "subpersonal routines," which by nature are completely inaccessible to personal awareness under any conditions. The mind was divided into two radically different regions, with an unbridgeable chasm between them—the subjective mental states of the person and the subpersonal cognitive routines implemented in the brain. The radically nonconscious, subpersonal region, the so-called cognitive unconscious, is where the action of thought really happens; personal awareness has access merely to a few results or epiphenomenal manifestations of subpersonal processing (Jackendoff 1987). Thought corresponds to nonconscious, skull-bound, symbol manipulation. It takes place in a central cognitive module of the brain separate from the systems for perception, emotion, and motor action. The cognitive unconscious is neither somatic nor affective, and it is lodged firmly within the head.

This radical separation of cognitive processes from consciousness created a peculiar "explanatory gap" in scientific theorizing about the mind.[3] Cartesian dualism had long ago created an explanatory gap between mind and matter, consciousness and nature. Cognitivism, far from closing this gap, perpetuated it in a materialist form by opening a new gap between subpersonal, computational cognition and subjective mental phenomena. Simply put, cognitivism offered no account whatsoever of mentality in the sense of subjective experience. Some theorists even went so far as to claim that subjectivity and consciousness do not fall within the province of cognitive science (Pylyshyn 1984). Not all theorists shared this view, however. A notable exception was Ray Jackendoff, who clearly formulated the problem facing cognitivism in his 1987 book *Consciousness and the Computational Mind.* According to Jackendoff, cognitivism, in radically differentiating computational cognition from subjective experience, produced a new "mind-mind" problem, in addition to the classical mind-body problem. The mind-mind problem is the problem of the relation between the computational mind and the phenomenological mind, between subpersonal, computational, cognitive processes and conscious experience (Jackendoff 1987, p. 20). Thanks to cognitivism, a new set of mind-body problems had to be faced:

1. The phenomenological mind-body problem: How can a brain have experiences?

2. The computational mind-body problem: How can a brain accomplish reasoning?
3. The mind-mind problem: What is the relation between computational states and experience?

Each problem is a variant of the explanatory gap. The cognitivist metaphor of the mind as computer, which was meant to solve the computational mind-body problem, thus came at the cost of creating a new problem, the mind-mind problem. This problem is a version of what is now known as the "hard problem of consciousness" (Chalmers 1996; Nagel 1974).

During the heyday of cognitivism in the 1970s and early 1980s, cognitivists liked to proclaim that their view was "the only game in town" (Fodor 1975, 1981), and they insisted that the computer model of the mind is not a metaphor but a scientific theory (Pylyshyn 1984), unlike earlier mechanistic models, such as the brain as a telephone switchboard. The cognitive anthropologist Edwin Hutchins (1995), however, has argued that a confused metaphorical transference from culture to individual psychology lies at the very origin of the cognitivist view. Cognitivism derives from taking what is in fact a sociocultural activity—human computation—and projecting it onto something that goes on inside the individual's head. The cognitive properties of computation do not belong to the individual person but to the sociocultural system of individual-plus-environment.

The original model of a computational system was a person—a mathematician or logician manipulating symbols with hands and eyes, and pen and paper. (The word "computer" originally meant "one who computes.") This kind of physical symbol system is a sophisticated and culturally specific form of human activity. It is embodied, requiring perception and motor action, and embedded in a sociocultural environment of symbolic cognition and technology. It is not bounded by the skull or skin but extends into the environment. The environment, for its part, plays a necessary and active role in the cognitive processes themselves; it is not a mere contingent, external setting (Clark and Chalmers 1998; Wilson 1994). Nevertheless, the human mind is able to idealize and conceptualize computation in the abstract as the mechanical application of formal rules to symbol strings, as Alan Turing did in arriving at his mathematical notion of a Turing Machine. Turing suc-

cessfully abstracted away from both the world in which the mathematician computes and the psychological processes he or she uses to perform a computation. But what do such abstract formal systems reflect or correspond to in the real world? According to the cognitivist "creation myth," what Turing succeeded in capturing was the bare essentials of intelligent thought or cognition within the individual (all the rest being mere implementation details).

The problem with this myth is that real human computation—the original source domain for conceptualizing computation in the abstract—was never simply an internal psychological process; it was a sociocultural activity as well. Computation, in other words, never reflected simply the cognitive properties of the individual, but instead those of the sociocultural system in which the individual is embedded. Therefore, when one abstracts away from the situated individual what remains is precisely not the bare essentials of individual cognition, but rather the bare essentials of the sociocultural system: "The physical-symbol-system architecture is not a model of individual cognition. It is a model of the operation of a sociocultural system from which the human actor has been removed" (Hutchins 1995, p. 363; emphasis omitted). Whether abstract computation is well suited to model the structure of thought processes within the individual is therefore questionable. Nevertheless, cognitivism, instead of realizing that its computer programs reproduced (or extended) the abstract properties of the sociocultural system, projected the physical-symbol-system model onto the brain. Because cognitivism from its inception abstracted away from culture, society, and embodiment, it remained resistant to this kind of critical analysis and was wedded to a reified metaphor of the mind as a computer in the head.[4]

The connectionist challenge to cognitivism, however, did not take the form of this kind of critique. Rather, connectionist criticism focused on the neurological implausibility of the physical-symbol-system model and various perceived deficiencies of symbol processing compared with neural networks (McLelland, Rummelhart, and the PDP Research Group 1986; Smolensky 1988).

Connectionism

Connectionism arose in the early 1980s, revising and revitalizing ideas from the precognitivist era of cybernetics.[5] Connectionism is now

widespread. Its central metaphor of the mind is the neural network. Connectionist models of cognitive processes take the form of artificial neural networks, which are virtual systems run on a digital computer. An artificial neural network is composed of layers of many simple neuron-like units that are linked together by numerically weighted connections. The connection strengths change according to various learning rules and the system's history of activity.

The network is trained to convert numerical (rather than symbolic) input representations into numerical output representations. Given appropriate input and training, the network converges toward some particular cognitive performance, such as producing speech sounds from written text (as in the famous NETtalk system of Sejnowski and Rosenberg 1986), or categorizing words according to their lexical role (Elman 1991). Such cognitive performances correspond to emergent patterns of activity in the network. These patterns are not symbols in the traditional computational sense, although they are supposed to be approximately describable in symbolic terms (Smolensky 1988). Connectionist explanations focus on the architecture of the neural network (units, layers, and connections), the learning rules, and the distributed subsymbolic representations that emerge from the network's activity. According to connectionism, artificial neural networks capture the abstract cognitive properties of neural networks in the brain and provide a better model of the cognitive architecture of the mind than the physical symbol systems of cognitivism.

The connectionist movement of the 1980s emphasized perceptual pattern recognition as the paradigm of intelligence, in contrast to deductive reasoning, emphasized by cognitivism. Whereas cognitivism firmly lodged the mind within the head, connectionism offered a more dynamic conception of the relation between cognitive processes and the environment. For example, connectionists hypothesized that the structural properties of sequential reasoning and linguistic cognition arise not from manipulations of symbols in the brain, but from the dynamic interaction of neural networks with symbolic resources in the external environment, such as diagrams, numerical symbols, and natural language (Rummelhart et al. 1986).

Despite these advances, connectionist systems did not involve any sensory and motor coupling with the environment, but instead operated on the basis of artificial inputs and outputs (set initially by the designer of the system). Connectionism also inherited from cognitivism

the idea that cognition is basically the solving of predefined problems (posed to the system from outside by the observer or designer) and that the mind is essentially the skull-bound cognitive unconscious, the subpersonal domain of computational representation in the mind-brain. Connectionism's disagreement with cognitivism was over the nature of computation and representation (symbolic for cognitivists, subsymbolic for connectionists).

With regard to the problem of the explanatory gap, connectionism enlarged the scope of the computational mind but provided little, if any, new resources for addressing the gap between the computational mind and the phenomenological mind. Subjectivity still had no place in the sciences of mind, and the explanatory gap remained unaddressed.

Embodied Dynamicism

The third approach, embodied dynamicism, arose in the 1990s and involved a critical stance toward computationalism in either its cognitivist or connectionist form.[6] Cognitivism and connectionism left unquestioned the relation between cognitive processes and the real world. As a result, their models of cognition were disembodied and abstract. On the one hand, cognitive processes were said to be instantiated (or realized or implemented) in the brain, with little thought given to what such a notion could mean, given the biological facts of the brain and its relationship to the living body of the organism and to the environment. On the other hand, the relationship between the mind and the world was assumed to be one of abstract representation: symbolic or subsymbolic representations in the mind-brain stand for states of affairs in some restricted outside domain that has been specified in advance and independently of the cognitive system. The mind and the world were thus treated as separate and independent of each other, with the outside world mirrored by a representational model inside the head. Embodied dynamicism called into question all of these assumptions, in particular the conception of cognition as disembodied and abstract mental representation. Like connectionism, embodied dynamicism focuses on self-organizing dynamic systems rather than physical symbol systems (connectionist networks are examples of self-organizing dynamic systems), but maintains in addition that cognitive

processes emerge from the nonlinear and circular causality of continuous sensorimotor interactions involving the brain, body, and environment. The central metaphor for this approach is the mind as embodied dynamic system in the world, rather than the mind as neural network in the head.

As its name suggests, embodied dynamicism combines two main theoretical commitments. One commitment is to a dynamic systems approach to cognition, and the other is to an embodied approach to cognition.

The central idea of the dynamic systems approach is that cognition is an intrinsically temporal phenomenon and accordingly needs to be understood from the perspective of dynamic systems theory (Port and van Gelder 1995; van Gelder 1998). A dynamic systems model takes the form of a set of evolution equations that describe how the state of the system changes over time. The collection of all possible states of the system corresponds to the system's "state space" or "phase space," and the ways that the system changes state correspond to trajectories in this space. Dynamic-system explanations focus on the internal and external forces that shape such trajectories as they unfold in time. Inputs are described as perturbations to the system's intrinsic dynamics, rather than as instructions to be followed, and internal states are described as self-organized compensations triggered by perturbations, rather than as representations of external states of affairs.

The central idea of the embodied approach is that cognition is the exercise of skillful know-how in situated and embodied action (Varela, Thompson, and Rosch 1991). Cognitive structures and processes emerge from recurrent sensorimotor patterns that govern perception and action in autonomous and situated agents. Cognition as skillful know-how is not reducible to prespecified problem solving, because the cognitive system both poses the problems and specifies what actions need to be taken for their solution.

Strictly speaking, dynamicism and embodiment are logically independent theoretical commitments. For example, dynamical connectionism incorporates dynamicist ideas into artificial neural networks (see Port and van Gelder 1995, pp. 32–34), whereas autonomous agents research in robotics incorporates embodiment ideas without employing dynamic systems theory (Maes 1990). Nevertheless, dynamicism and embodiment go well together and are intimately related for

many theorists. As Randall Beer notes: "Although a dynamical approach can certainly stand alone, it is most powerful and distinctive when coupled with a situated, embodied perspective on cognition" (Beer 2000, p. 97).

Embodied dynamicism provides a different perspective on the cognitive unconscious from computationalism. No longer is the cognitive unconscious seen as disembodied symbol manipulation or pattern recognition separate from emotion and motor action in the world. Instead, the cognitive unconscious consists of those processes of embodied and embedded cognition and emotion that cannot be made experientially accessible to the person. This characterization of the cognitive unconscious is offered not as a hypothetical construct in an abstract functionalist model of the mind, but rather as a provisional indication of a large problem-space in our attempt to understand human cognition.

At least four points need emphasizing in this context. First, as a conceptual matter, the relations among what is nonconscious, unconscious, preconscious, and conscious (in any of the innumerable senses of these words)—or in a different, but not equivalent idiom, what is subpersonal and personal—remain far from clear. Second, as an empirical matter, the scope and limits of awareness of one's own psychological and somatic processes have yet to be clearly mapped and undoubtedly vary across subjects. Third, the key point still stands that most of what we are as psychological and biological beings is in some sense unconscious. It follows that subjectivity cannot be understood without situating it in relation to these unconscious structures and processes. Finally, these unconscious structures and processes, including those describable as cognitive and emotional, extend throughout the body and loop through the material, social, and cultural environments in which the body is embedded; they are not limited to neural processes inside the skull.

The emergence of embodied dynamicism in the 1990s coincided with a revival of scientific and philosophical interest in consciousness, together with a renewed willingness to address the explanatory gap between scientific accounts of cognitive processes and human subjectivity and experience. A number of works on embodied cognition were explicitly concerned with experience and challenged the objectivist assumptions of computationalism.[7] Some of these works were also ex-

plicitly dynamical in orientation.[8] In particular, the enactive approach of Varela, Thompson, and Rosch (1991) aimed to build bridges between embodied dynamicist accounts of the mind and phenomenological accounts of human subjectivity and experience. The present book continues this project.

The Enactive Approach

Enaction means the action of enacting a law, but it also connotes the performance or carrying out of an action more generally. Borrowing the words of the poet Antonio Machado, Varela described enaction as the laying down of a path in walking: "Wanderer the road is your footsteps, nothing else; you lay down a path in walking" (Varela 1987, p. 63).

The term *the enactive approach* and the associated concept of enaction were introduced into cognitive science by Varela, Thompson, and Rosch (1991) in their book *The Embodied Mind*. They aimed to unify under one heading several related ideas. The first idea is that living beings are autonomous agents that actively generate and maintain themselves, and thereby also enact or bring forth their own cognitive domains. The second idea is that the nervous system is an autonomous dynamic system: It actively generates and maintains its own coherent and meaningful patterns of activity, according to its operation as a circular and reentrant network of interacting neurons. The nervous system does not process information in the computationalist sense, but creates meaning. The third idea is that cognition is the exercise of skillful know-how in situated and embodied action. Cognitive structures and processes emerge from recurrent sensorimotor patterns of perception and action. Sensorimotor coupling between organism and environment modulates, but does not determine, the formation of endogenous, dynamic patterns of neural activity, which in turn inform sensorimotor coupling. The fourth idea is that a cognitive being's world is not a prespecified, external realm, represented internally by its brain, but a relational domain enacted or brought forth by that being's autonomous agency and mode of coupling with the environment. The fifth idea is that experience is not an epiphenomenal side issue, but central to any understanding of the mind, and needs to be investigated in a careful phenomenological manner. For this reason,

the enactive approach maintains that mind science and phenomenological investigations of human experience need to be pursued in a complementary and mutually informing way.[9]

The conviction motivating the present book is that the enactive approach offers important resources for making progress on the explanatory gap. One key point is that the enactive approach explicates selfhood and subjectivity from the ground up by accounting for the autonomy proper to living and cognitive beings. The burden of this book is to show that this approach to subjectivity is a fruitful one.

To make headway on this project, we need to draw from biology, neuroscience, psychology, philosophy, and phenomenology. In this book, I try to integrate investigations from all these fields.

One common thread running through the following chapters is a reliance on the philosophical tradition of phenomenology, inaugurated by Edmund Husserl and developed in various directions by numerous others, most notably for my purposes by Maurice Merleau-Ponty (Moran 2000; Sokolowski 2000; Spiegelberg 1994).[10] My aim, however, is not to repeat this tradition's analyses, as they are found in this or that author or text, but to present them anew in light of present-day concerns in the sciences of mind. Thus this book can be seen as contributing to the work of a new generation of phenomenologists who strive to "naturalize" phenomenology (Petitot et al. 1999). The project of naturalizing phenomenology can be understood in different ways, and my own way of thinking about it will emerge later in this book. The basic idea for the moment is that it is not enough for phenomenology simply to describe and philosophically analyze lived experience; phenomenology needs to be able to understand and interpret its investigations in relation to those of biology and mind science.

Yet mind science has much to learn from the analyses of lived experience accomplished by phenomenologists. Indeed, once science turns its attention to subjectivity and consciousness, to experience as it is lived, then it cannot do without phenomenology, which thus needs to be recognized and cultivated as an indispensable partner to the experimental sciences of mind and life. As we will see, this scientific turn to phenomenology leads as much to a renewed understanding of nature, life, and mind as to a naturalization of phenomenology (Zahavi 2004b).

There is also a deeper convergence of the enactive approach and

phenomenology that is worth summarizing briefly here. Both share a view of the mind as having to constitute its objects. Here constitution does not mean fabrication or creation; the mind does not fabricate the world. "To constitute," in the technical phenomenological sense, means to bring to awareness, to present, or to disclose. The mind brings things to awareness; it discloses and presents the world. Stated in a classical phenomenological way, the idea is that objects are disclosed or made available to experience in the ways they are thanks to the intentional activities of consciousness. Things show up, as it were, having the features they do, because of how they are disclosed and brought to awareness by the intentional activities of our minds. Such constitution is not apparent to us in everyday life but requires systematic analysis to disclose. Consider our experience of time (discussed in Chapter 11). Our sense of the present moment as both simultaneously opening into the immediate future and slipping away into the immediate past depends on the formal structure of our consciousness of time. The present moment manifests as a zone or span of actuality, instead of as an instantaneous flash, thanks to the way our consciousness is structured. As we will see later, the present moment also manifests this way because of the nonlinear dynamics of brain activity. Weaving together these two types of analysis, the phenomenological and neurobiological, in order to bridge the gap between subjective experience and biology, defines the aim of neurophenomenology (Varela 1996), an offshoot of the enactive approach.

The enactive approach and phenomenology also meet on the common ground of life or living being. For the enactive approach, autonomy is a fundamental characteristic of biological life, and there is a deep continuity of life and mind. For phenomenology, intentionality is a fundamental characteristic of the lived body. The enactive approach and phenomenology thus converge on the proposition that subjectivity and consciousness have to be explicated in relation to the autonomy and intentionality of life, in a full sense of "life" that encompasses, as we will see, the organism, one's subjectively lived body, and the life-world.

It will take some work before these ideas can stand clearly before us in this book. In the next chapter I introduce phenomenological philosophy in more detail, before returning to the enactive approach in Chapter 3.